Physics Education
Research Group

Department of Physics and Astronomy
Iowa State University
Ames, Iowa

Physics Education Research

Collected Papers

Table of Contents

VI. Methodological Issues in PER

VII. PER in Overview

I.

Overview of the Group’s Work

Iowa State University is New Entrant into Physics Education Research Community

David E. Meltzer

Last year the Department of Physics and Astronomy at Iowa State University inaugurated a new group devoted to physics education research. Thus ISU has joined about a dozen other physics departments around the country in which the new subfield of Physics Education joins more traditional fields as a legitimate area for scholarly research, and for training of graduate students. Department Chairman Douglas Finnemore said that "We want to put Physics Education on the same intellectual and competitive level as particle physics, nuclear physics, condensed matter physics, and astronomy."

The origins of physics education research (PER) lie in the strong desire of physics instructors to maximize the effectiveness of the teaching and learning of physics. It seems only natural that physicists are now applying their training and systematic analytical methods – so successfully used to understand the physical world – to explore the problems related to the learning of their subject. Within the past two decades, physicists in the colleges and universities have initiated intense efforts to study physics learning, particularly among undergraduate students. The efforts of PER to identify and address learning difficulties in physics should result in improved learning by both average students *and* high-performing students.

At Iowa State, in common with other PER groups, we engage in three distinct yet closely linked activities: (1) develop and assess more effective curricular materials; (2) implement and assess new instructional methods that make use of the improved curricula; (3) investigate learning difficulties, and carry out other basic research on the teaching and learning of physics. Our particular focus is on curriculum and instructional methods for large lecture classes.

Our objective is to address areas of pedagogical concern previously identified by physics education researchers. For instance, many if not most students in introductory courses develop weak *qualitative* understanding of concepts, even when they may be able to solve successfully certain types of quantitative problems. When lacking exact quantitative solutions, students often have difficulty in determining qualitative features such as comparison of magnitudes, determination of direction, and evaluation of trends.[1]

More broadly, students frequently lack a "functional" understanding of physics concepts, which would allow problem solving in a context *different* from the one in which the concept was originally learned. Students find it difficult to transfer an ability to solve standard textbook problems to situations involving actual, real-world physical objects and phenomena.[2] Moreover, there is a strong tendency to view phenomena and concepts as distinct, unrelated and highly dependent on context, rather than as comprehensible and derivable from just a few underlying universal principles.[3]

A number of factors have been identified as playing a role in these learning difficulties. For example, students enter introductory classes with their own ideas about the physical world that may strongly conflict with physicists' views.[4] Often called "misconceptions" or "alternative conceptions," these ideas are widely prevalent; there are some particular ideas that are almost universally held by beginning students. These ideas are often well-defined; they are not merely a "lack of understanding," but a very specific idea about what ***should*** be the case (but in fact is not). Examples of these ideas are that an object in motion *must* be experiencing a force, and that a given battery always produces the *same* current in *any* circuit. These ideas are often – usually – ***very*** tenacious and hard to dislodge.

Another important factor is that most students in introductory courses lack "active learning" skills, and need much guidance in scientific reasoning. Physics concepts are usually subtle, counterintuitive, and required extended chains of reasoning. Of course, ***some*** students learn efficiently. Highly successful physics students (e.g., future physics instructors) are active learners. They continuously probe their own understanding of a concept, for instance by posing their own questions and examining varied contexts. They are sensitive to areas of confusion, and have the confidence to confront those areas directly.

By contrast, the great majority of introductory students are unable to do efficient active learning on their own. They don't know "which questions they need to ask." They require considerable prodding by instructors (aided by appropriate curricular materials), and need frequent hints and confidence boosts.[3]

To address these problems, innovative pedagogical methods are being developed. To encourage active learning, students are led to engage in deeply thought-provoking activities requiring intense mental effort (so-called "Interactive Engagement"[5]). Students are frequently required to provide written or oral explanations of their reasoning process. Instruction recognizes – and deliberately elicits – students' preexisting "alternative conceptions," which are then made a focus of discussion. As much as possible, the ***process*** of science – exploration and discovery – is used as a means for ***learning*** science. Instructors avoid *telling* students that certain things are true, and instead students are guided to "figure it out for themselves," either in the instructional lab, or by step-by-step theoretical analysis.

We have been developing curricular materials along these themes for elementary topics in electricity and magnetism, and modern physics. Our "*Workbook for Introductory Physics*" (in collaboration with K. Manivannan) guides students to construct in-depth understanding through step-by-step confrontation with conceptual sticking points and counterintuitive ideas. Contexts are varied by heavy use of multiple representations – intermixing equations, word problems, pictures, diagrams, graphs and charts. In collaboration with ISU chemistry professor Tom Greenbowe – a long-time researcher in chemical education – we are developing similar materials for the thermodynamics curriculum. All materials undergo continuous testing and redesign through day-to-day class use and student assessment. Our curriculum development has been most strongly influenced by the pioneering work of Lillian McDermott and Alan Van Heuvelen.[6]

An active learning classroom is characterized by very high levels of interaction between students and instructor, and among the students themselves. There is usually collaborative group work, and students all engage in intensive learning activities far beyond passive listening and note copying. Students may be asked to make predictions of the outcome of experiments, and give written explanations of their reasoning. Instructors pose specific problems that are known to consistently trigger certain types of learning difficulties, and subsequent activities are then structured to confront these difficulties. Instructors avoid "telling" and instead provide leading questions. "Peer instruction" methods are employed in which students explain their reasoning to each other, and then critique each others' arguments.[7]

In the small-class environment, we have implemented active learning techniques in an NSF-supported elementary physics course targeted at elementary education majors.[8] For (some) large classes, we use the "Flash Card" response system to obtain instantaneous feedback on multiple-choice *Workbook* questions from all students simultaneously.[9] Students also spend a large fraction of class time working in groups on carefully structured free-response sequences in the *Workbook*. Recitations in selected courses are replaced by University-of-Washington-style "tutorials": students work in groups on *Workbook* materials while T.A.'s provide guidance through Socratic questioning.

We also carry out basic research to support curriculum development. Graduate student Jack Dostal has been investigating student understanding of gravitation, by developing and administering free-response diagnostics and conducting in-depth videotaped student interviews. He is developing and assessing curricular materials to address learning difficulties identified in his research. In other research, we are investigating the comparative effectiveness of different representational modes, i.e., the relationship between the ***form of representation*** of physics concepts, and efficiency of student learning. We are also exploring factors underlying individual differences in student learning: why do some students start (conceptually) at the same point, yet finish at different points? How can instruction most effectively target these diverse groups of students?

We view PER as a systematic, multi-faceted endeavor to expand the horizons of physics education for the new millenium. By building on past achievements and relentlessly exploring new instructional possibilities, we hope to significantly increase the impact that physics instructors worldwide will be able to have on their students' educational development.

More information about our work can be found on our website http://www.public.iastate.edu/~per or by contacting us directly.

1. Lillian C. McDermott, *Millikan Lecture 1990: What we teach and what it learned – Closing the gap*, Am. J. Phys. **59**, 301 (1991).

2. Lillian C. McDermott, *Bridging the gap between teaching and learning: The role of research*, in ***Proceedings of the International Conference on Undergraduate Physics Education***, edited by E. F. Redish and J. S. Rigden, AIP, Woodbury, NY, Part One, pp. 139-165 (1997).

3. Frederick Reif, *Millikan Lecture 1994: Understanding and teaching important scientific thought processes*, Am. J. Phys. **63**, 17 (1995).

4. Arnold Arons, *Teaching Introductory Physics*, New York, Wiley, 1997.

5. Richard R. Hake, *Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses*, Am. J. Phys. **66**, 64 (1998).

6. Alan Van Heuvelen, *Overview, Case Study Physics*, Am. J. Phys. **59**, 898 (1991).

7. Eric Mazur, ***Peer Instruction, A User's Manual***, Upper Saddle River N.J., Prentice Hall (1997).

8. David E. Meltzer, *Effectiveness of instruction on force and motion in an elementary physics course based on guided inquiry*, AAPT Announcer **28(2)**, 125 (1998).

9. David E. Meltzer and Kandiah Manivannan, *Promoting interactivity in physics lecture classes*, The Physics Teacher **34**, 72 (1996).

David E. Meltzer is Assistant Professor of Physics at Iowa State University, and is director of the ISU Physics Education Research Group.

Initiatives in Education

Physics Education at ISU

David E. Meltzer
Assistant Professor of Physics

ISU physics students respond to instructor's question using "flash cards."

In recent years, physicists at U.S. colleges and universities have significantly increased efforts to improve the teaching and learning of physics at the undergraduate level. One of the components of this effort has been the creation of what is now recognized by the American Physical Society as a new subfield of physics: "Physics Education Research," or "PER." Physicists engaged in PER attempt to treat the problems involved in physics education, as much as possible, as they would any other research problem. This involves systematic observation and data collection, and the design and execution of pedagogical experiments that may be reproduced by different instructors in diverse institutions with widely varying student populations. Based on their advanced training in physics, PER researchers are uniquely situated to identify and control many of the variables involved in physics learning, and to carry out in-depth probes and analyses of students' thinking as they engage in the process of learning physics concepts.

The rapidly expanding research literature in PER[1] includes detailed studies of student learning difficulties in a wide variety of physics topics such as mechanics, electricity and magnetism, optics, and quantum mechanics. It also includes reports of the development and rigorous testing of innovative curricular materials and instructional methods designed to address and resolve many of these learning difficulties. Numerous investigations have provided strong and consistent evidence that research-based instructional methods and materials can significantly improve the learning of physics concepts by college and university students.

There are now approximately 50 physics departments at U.S. colleges and universities in which one or more faculty members devote a majority of their research effort to PER. About a dozen research universities carry out graduate research programs in PER, including the award of Masters and Ph.D. degrees in physics for dissertations in physics education. The largest of these groups, at the University of Washington in Seattle, has awarded more than 12 Ph.D. degrees in physics education research. Beginning in August 1998, Iowa State University joined the ranks of universities offering advanced degrees in physics education research; our first Masters degree in this field was awarded in May 2001.

At ISU, our physics education research group has engaged in close collaboration with the long-standing ISU chemistry education research group led by Tom Greenbowe, Professor of Chemistry. Since 1998, three physics graduate students and three undergraduates (one from a neighboring college) have helped carry out the work in our group. Our research group engages in coordinated efforts in a number of distinct, though closely related areas. First, we carry out "basic research" in physics education by exploring in depth students' learning difficulties in diverse areas of physics. Projects currently ongoing or nearing completion include studies of student concepts in gravitation (by graduate student Jack Dostal, now at Montana State University), astronomical scale and lunar phases (by Masters graduate Tina Fanetti), and vectors (by graduate student Ngoc-Loan Nguyen). University Professor of Astronomy Lee Anne Willson has been a principal collaborator in the astronomy education research conducted by Tina Fanetti.

In 2000, with Tom Greenbowe as Co-Principal Investigator, I was awarded a $149,000 grant from the National Science Foundation to develop innovative curricular materials in thermodynamics. These materials include research-based problem sets that are carefully designed both to elicit common student difficulties regarding the subjects under study, and

Cont on pg. 14.

Physics Edu cont.
then to lead students to confront these difficulties head-on with tightly focused and strategically sequenced series of questions and exercises.

Another major project for our group has been to develop improved instructional methods and curricular materials for large-enrollment course, including lecture courses in which an instructor faces 100 to 250 students at a time. Our objective is to incorporate active-learning methods in such courses, in which students engage in diverse problem-solving activities during class time. Our current focus is on the algebra-based general physics course, populated predominantly by life-sciences majors. The level of student-student and student-instructor interaction in these classes is dramatically increased by the use of a student response system incorporating "flash cards." Every single student in the class has a pack of six large flash cards (5_" × 8"), each printed with one of the letters "A," "B," "C," "D," "E," or "F." These flash cards permit the instructor to get instantaneous responses to multiple-choice questions by *all* of the students in the class simultaneously (see accompanying photo). Curricular materials to support this instructional method – including a large collection of specially designed sequences of multiple-choice questions – have been developed in collaboration with Prof. Kandiah Manivannan of Southwest Missouri State University. These materials are incorporated in the *Workbook for Introductory Physics*, now available in a preliminary edition in CD-ROM format. Various assessments employed standardized tests have demonstrated that learning gains by ISU students enrolled in these "active-learning" course are significantly higher than those found in national surveys of students in more traditional learning environments.

Much more detailed information on the work of the ISU PER group is available on our website, http://www.public.iastate.edu/~per/. Many of our papers and conference presentations can be viewed at that site, along with details of our NSF-sponsored curriculum project. We would be happy to provide further details and samples of our group's work, including copies of the *Workbook for Introductory Physics* CD-ROM, to interested readers of this newsletter. Please contact me directly at dem@iastate.edu.

[1] *Lillian C. McDermott and Edward F. Redish, "Resource Letter: PER1: Physics Education Research," Am. J. Phys.* ***67,*** *755-767 (1999).*

II.

Teacher Preparation and Instruction for Pre-College Students

Guided Inquiry

Let Students "Discover" the Laws of Physics for Themselves

by David E. Meltzer and Amy Woodland Espinoza

This lesson on the law of reflection will help you understand how guided inquiry should be used.

There is ever-increasing interest in hands-on activities that allow middle-school science students to explore and discover physical principles on their own. The basic idea behind this "guided inquiry" is that students will gain a better grasp of scientific ideas if they perform activities that permit them to figure out these principles before the instructor actually states them explicitly in a lecture.[1] Of course, students need careful guidance from instructors if this is to succeed in practice. Here we present, as a model of such a guided inquiry activity, a lesson dealing with the law of reflection.

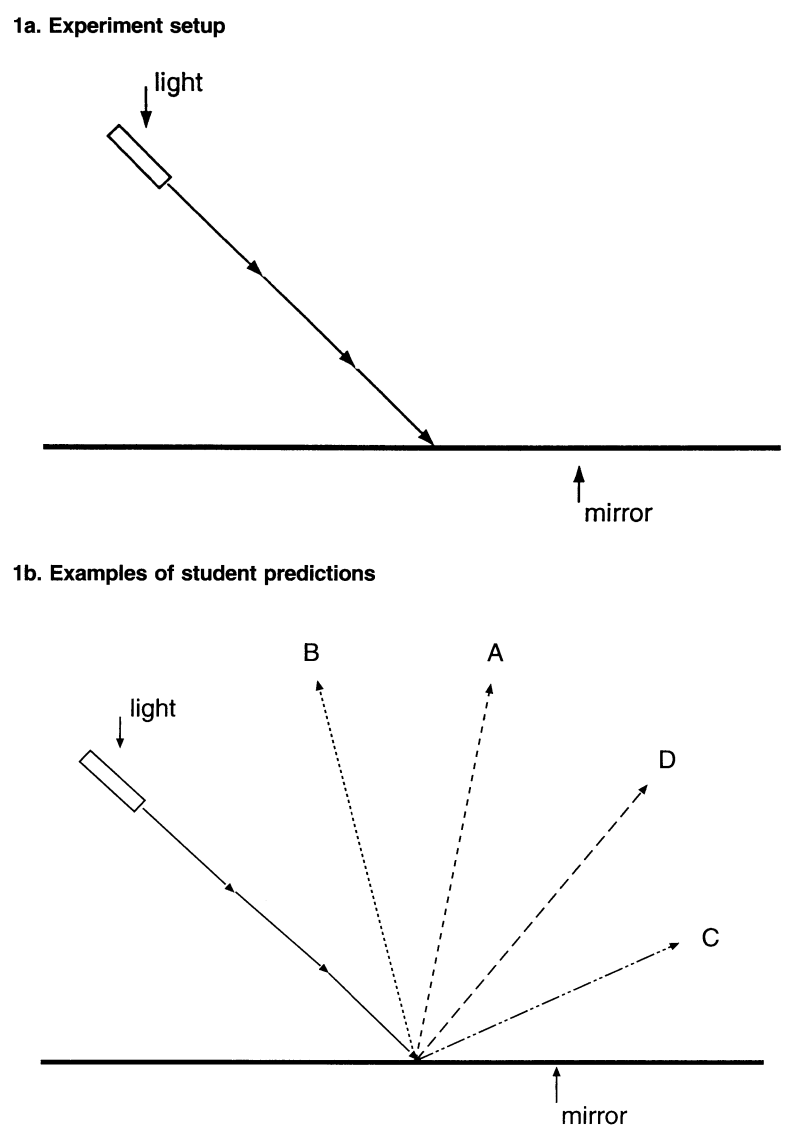
Preparing for the activity

Before beginning this lesson, students must understand the concept of angles including how to measure them using protractors. They also need to understand the definition of a right angle. In most classes, only a brief review and a bit of practice is needed, but more time may have to be spent on this phase if students are unclear on these concepts.

Students must also understand how to use ray boxes and mirrors. Ray boxes (available from many suppliers or from your local high school or university) use a light source and slotted screens to produce one or more pencil-thin rays of light. When placed on a white surface (such as a sheet of paper) in a darkened room, the rays can be easily viewed along any desired

David E. Meltzer is assistant professor of physics at Southeastern Louisiana University in Hammond, Louisiana. Amy Woodland Espinoza is a teacher at Brighton Academy in Baton Rouge.

Figure 1. Introducing angles of incidence and reflection



path. Small mirrors supplied with the boxes allow for the production and viewing of reflected rays. Before teaching the lesson, give students 10 to 15 minutes to practice with the ray boxes and mirrors.

Introducing reflected beams

Start the lesson by allowing students to predict the path of a reflected beam. First, supply groups of three or four students with a diagram of a light beam heading toward a mirror (Figure 1a). Then, project the same diagram onto a screen using an overhead projector and have a few volunteers sketch their prediction of the path of the reflected light beam, labeling each prediction A, B, C, and so on. Give the class plenty of time to consider the predictions and ask if anyone else has a prediction that differs from those already drawn. A diagram of students' typical predictions is shown in Figure 1b.

Students then test the validity of the predictions by placing their ray box on top of their copy of the diagram and examining the actual path of the reflected ray. At this point, no measurements are made; students simply observe. As a class, discuss which prediction was closest to the actual path.

Learning the terminology

In order to provide common language with which to discuss the observations students just made, as well as those to come, take a few minutes to draw and discuss angles of incidence (Figure 2a). Then, draw and discuss angles of reflection (Figure 2b). As you move from one diagram to the next, erase the "angle of incidence" diagram before moving on to the angle of reflection to avoid giving any hint of the relationship between these two angles.

After discussing the terminology, tell students that there is a simple relationship between the magnitudes of the angle of incidence and the angle of reflection that they will learn from carrying out a simple experiment. Through the activity at the end of this article, students will find that the two angles seem to be the same and will thus "discover" the law of reflection, which states that the angle of incidence is equal to the angle of reflection.

Guiding students toward discovery

While students carry out the steps of the activity, circulate around the room offering assistance as needed. When each group is finished, moderate a class discussion on their findings. Have a representative from each group report their measurements for the angles of incidence and reflection, and create a chart on the board listing the values. The chart should list both the angle of incidence and the angle of reflection found by each student group, with a separate row for each group's results to allow easy comparison of the students' observations. Even though students are measuring the same diagram, there will probably be some minor variations in their measurements; this provides students with a better understanding of experimental uncertainties. Ask students for their comments on the relationship between the two angles and let them draw their own conclusions as much as possible. They should soon reach a consensus that the angles seem to be about the same. Confirm this observation, and tell students that they have just discovered the law of reflection.

Assessment and application

Assessing students' understanding of the law of reflection is easy. Simply provide students with a diagram of a light beam (incident beam) aimed at a mirror (Figure 3), and ask the following questions:

- What is the angle of incidence?
- What would be the angle of reflection?
- Were you right? Use the ray box to check your answers.

To determine the angle measure-

Figure 2.

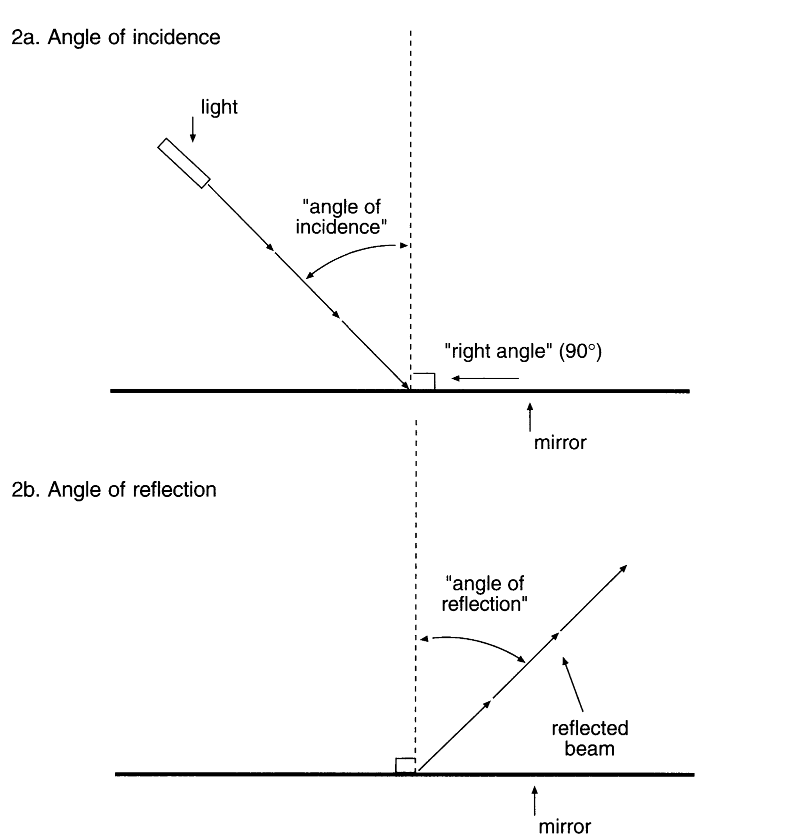
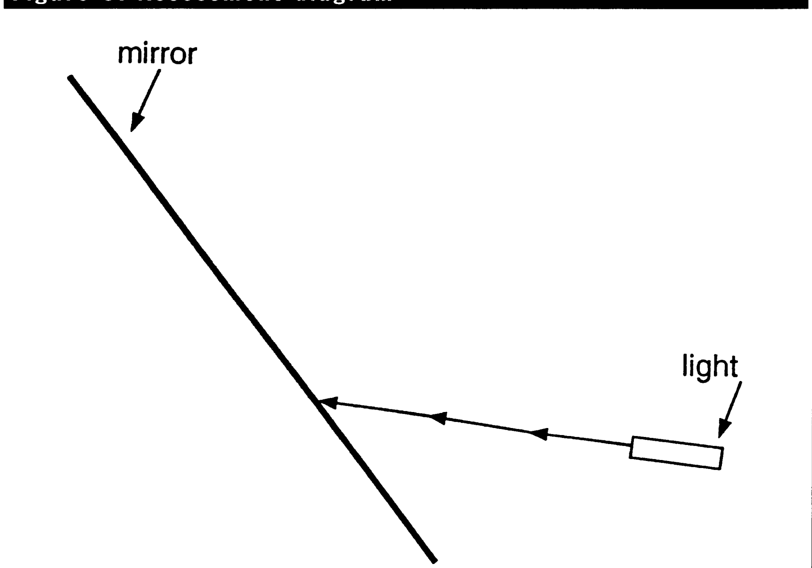


Figure 3. Assessment diagram



ments, students must first remember to draw a line perpendicular to the mirror. A useful follow-up experiment is to supply a diagram with the reflected beam indicated, and then ask students to draw the incident beam. They could then check their answer using the ray box.

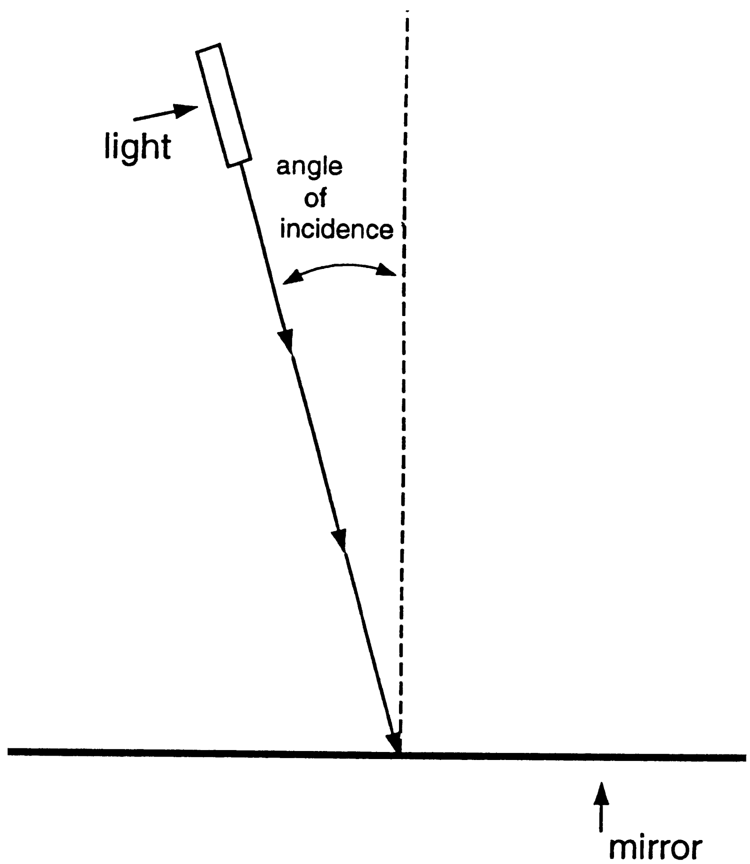
After students have successfully completed the laboratory-type experiments described previously, ask them to use the principle they have just learned. Supply each group with a mirror and a flashlight and instruct students to direct a reflected beam of light to shine on some object placed at random on their table. A more complicated activity is to give students two or three larger mirrors and ask them to direct a flashlight beam at an object located inside the (darkened) room. Through trying to carry out the tricky task of getting the beam to hit the target after two or three separate reflections, students gain a much better understanding and appreciation of the law of reflection.

Acknowledgment
The summer science class in which this lesson was taught was part of an inservice program at Louisiana State University sponsored by the Louisiana Systemic Initiatives Program and directed by Paul Lee and Sheila Pirkle. We are grateful to them for the opportunity provided us to develop and instruct this model lesson.

Reference

1. Lunetta, V.N. and S. Novick. 1982. *Inquiring and Problem-Solving in the Physical Sciences, A Sourcebook.* Dubuque, Iowa: Kendall/Hunt.

Discovering the laws of reflection



Materials
(For each group of three or four students)
- Protractor
- Ray box (if you don't have enough ray boxes for each group, students can share)
- Mirror
- Diagram of angle of incidence (above)
- Pen or marker

Procedure
1. Using a protractor, measure the angle of incidence as depicted in the diagram.
2. Record your measurement.
3. Place your ray box and mirror on the diagram and position them so that the incoming light ray follows exactly the same path as shown on the diagram and hits the mirror at exactly the same angle.
4. Use a pen or marker to trace the path of the reflected ray onto your diagram.
5. Use the protractor to measure the angle of reflection.
6. Record your measurement and be prepared to report your two measurements to the class.
7. How do your measurements of the two angles compare to each other? Is one angle obviously larger than the other, or do they both appear to be about the same? Be prepared to discuss your answer with the rest of the class.

"Micro-Document" for NSF Teacher Education Workshop, May 7-8, 1999

David E. Meltzer
Physics Education Group, Department of Physics and Astronomy
Iowa State University

(1) In order to ensure that all students master national science and mathematics standards, what qualities should effective K-12 teachers in the 21st Century possess?

In order to achieve this very ambitious goal, all students will need to have access on a regular and continuing basis to at least ***some*** teachers who combine several different essential qualities. These teachers will need a significant amount of content knowledge in the various areas of science – well beyond the level of the concepts that they are expected to teach. This is essential because, to be capable of carrying out "inquiry-based" instruction, a teacher must have considerable depth and breadth of knowledge. The teacher must be able to thoroughly comprehend typical learning difficulties encountered by students, must be able to respond to students' questions and confusion with well-thought-out, fruitful lines of questioning, and must be capable of leading students beyond their inevitable initial misunderstandings. Teachers will need a great deal of practice in carrying out guided-inquiry-based instruction; it is not something that one learns out of a text. Above all, teachers will need to have genuine enthusiasm for learning and teaching the concepts of science. Nothing will abort the educational process more rapidly than for students to be "taught" science and math by teachers who hate those subjects.

(2) What are the elements of and the barriers to an ideal program that produces and supports such a teacher? How are you or others you know overcoming these barriers?

It should go without saying (but in practice does not) that to teach science effectively, teachers-in-training will need to spend a very substantial amount of time learning science concepts in a guided-inquiry setting. ***In addition,*** they will need to practice their teaching skills under expert guidance, at least for some initial period. I believe that there is a great deal of disconnection from reality in much of the current discussion on teacher preparation for science and math teaching. Research from many groups has demonstrated one thing very convincingly: only intensive, time-consuming instruction (more than one semester in duration) has any hope of guiding most elementary-education students beyond well-known and widespread learning difficulties with basic physical science concepts. It is simply delusory to believe that significant progress toward the goals of the national science standards is possible within the current framework of teacher education, which for the most part comprises short-term exposure to many disparate subjects. The gap between what teachers at the elementary and middle-school level are "expected" to teach, and the actual knowledge that most of them possess, is vastly greater than often is imagined. A more realistic intermediate goal may well be to entrust most pre-secondary science instruction to science "specialists," who will receive substantial additional training and practice in the field.

(3) How can you or others know and document that you or they are producing a teacher that does indeed possess these qualities?

Ultimately, the only way to document this is to observe the teacher at work in a classroom with students. That should be part of any program that trains teachers to teach science. How effectively does the teacher guide student discussion and student activities? Is the teacher able to respond intelligently to student questions, by in turn asking the student the kind of question that will allow them to construct the targeted concept for themselves? Can the teacher test the student's knowledge by posing a problem in a novel, yet related context? Together, these form the ***sine qua non*** of effective science instruction. Short of field observation, those of us engaged in teacher education must intensively seek to assess student learning in depth. By posing problems in a wide variety of contexts, using multiple forms of representation (e.g., verbal, mathematical, diagrammatic, graphical, pictorial, physical, etc.), student learning may be more effectively assessed. By asking students to explain their reasoning – both in writing, and verbally – instructors will gain enormous insight into the students' actual depth of understanding. It should be considered an ***indispensable*** phase of assessment to probe and document students' thinking by analyzing their detailed written and verbal explanations of scientific concepts and principles.

(4) What more is needed to catalyze the changes outlined? Who can supply these needs? In what ways?

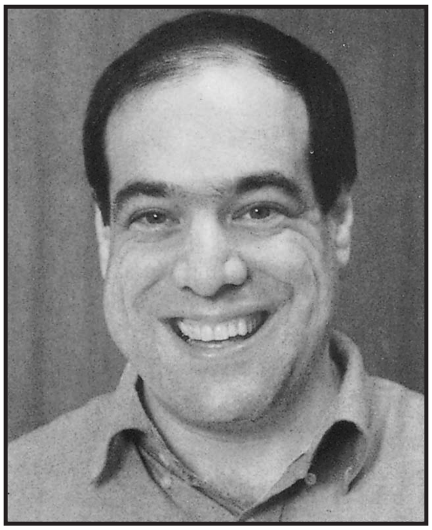
More confrontation with reality is certainly needed. Everyone involved in teacher education needs to address the assessment issue as seriously as possible. What are the goals of your instruction? How can you test whether those goals have been achieved? What means have you used to probe student understanding in depth? Have you observed your students as they attempt to explain the concepts they are expected to teach? Do you have some basis for anticipating their probable performance in the classroom? These questions should ***always*** be on the agenda.

III.

Active Learning in Large Classes

Promoting Interactivity in Physics Lecture Classes

By David E. Meltzer and Kandiah Manivannan



David E. Meltzer *is assistant professor of physics in the Department of Chemistry and Physics at Southeastern Louisiana University (Hammond, LA 70402). He received a Ph.D. in theoretical condensed matter physics from SUNY at Stony Brook in 1985 and had postdoctoral research appointments at the Oak Ridge National Laboratory, the University of Tennessee at Knoxville, and the Quantum Theory project at the University of Florida in Gainesville. He started teaching at SLU in 1991, focusing efforts on development of new curricular materials and instructional methods. He has directed an NSF-sponsored project to develop a new elementary physics course and has worked with middle-school and high-school teachers to improve precollege science instruction.*

Several innovative methods directed toward improving physics instruction in the introductory courses (both algebra- and calculus-based) have been developed recently. These include microcomputer-based laboratories,[1] integrated lab/lecture "studio" setups,[2] computerized animations and simulations employed in lectures,[3] and the use of electronic devices for linking students and instructors in the lecture hall.[4] However, the large number of students in introductory physics lecture classes makes it difficult to promote a higher level of student-faculty interaction and active student participation in the learning process during class time. At our institution, faced with limited resources and logistical constraints (e.g., no teaching assistants and little computer hardware), we have been working to develop methods that may be readily applied in the setting of lecture classes with a hundred or more students, and which are not dependent on simultaneous reorganization of the laboratory course. Our techniques are specifically aimed at converting a traditional lecture class, which may have either small or large attendance, into something that is closer in spirit to a seminar or a tutorial. We present here a number of the methods that we have been using and some of the thinking that underlies their development.

The Goal

The traditional lecture format consists of a rapid-fire presentation of ideas with little time or opportunity allowed for students to grapple with and comprehend concepts during class time. The detailed—and rather complex—thought processes that are required to master the key physical concepts tend to be glossed over or overlooked.[5,6] Instead, students become adept at recognizing certain problem types and patterns, and matching the pattern to an appropriate equation that may yield a numerical solution.[7] Studies have documented that, for instance, basic concepts in Newtonian mechanics are not learned very well even by most students who obtain good grades in traditional courses.[8,9]

We aim to *require* students to think about, discuss, work through, and solve problems *during class time* that bear directly on key conceptual issues.[10] (One consequence of this is a reduction in the sheer quantity of topics that may be presented during class.) The instructor plays more the role of a guide who promotes thinking and questioning by leading and focusing the discussion. (Quite similar methods have been pioneered during the past several years by Eric Mazur at Harvard University.[11]) We have in mind the "athletics instruction" paradigm: the "coach" doesn't just lecture and draw diagrams, but offers instantaneous critiques and feedback as the "player" attempts to perform the desired skill.

Methods Used

We utilize techniques for acquiring immediate feedback from all of the students in the class. Through these methods, the instructor is transformed from a "provider of information" into a tutorial leader who is constantly interacting with students, asking questions, hinting at answers, and helping students to move forward in their understanding. There are several interconnected phases in the instructional process, not all of which necessarily take place on the same day. The majority of class time is occupied by students working through conceptual questions and numerical problems, either with each other or in a constant back-and-forth dialogue with the instructor. The central elements of the process are as follows:

I. De-emphasis of Formal Lecture

In our large lecture classes we do not generally deliver a formal lecture in the traditional manner. Instead, we introduce concepts and solve sample problems for several minutes, at which point we pause and present either a question or a problem for the students to work on and discuss with each other. Although we might present an overview lecture in which the major ideas in a chapter are introduced and their interconnections sketched out, we would then return to these concepts one by one for approximately five to 15 minutes each.

II. Group Problem Solving

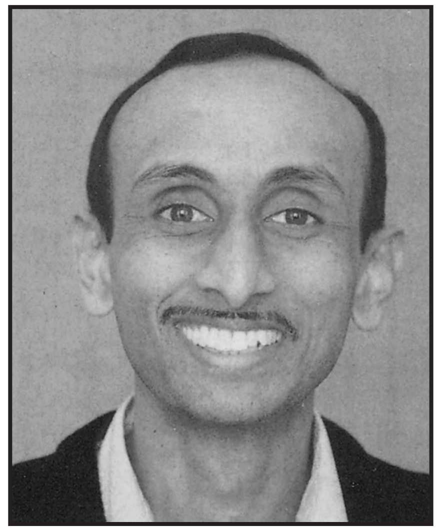
We give students time to work together on problems, typically in groups of two, three, or four neighboring students, and these groups are often encouraged to confer with each other. As the students discuss and work through these problems, the instructor frequently circulates throughout the room examining students' work when they indicate that they have a result and offering assistance to those who request it. Periodically, the instructor may go to the board and offer hints and partial solutions to the whole class as they continue to work. Then, when it appears that the majority of the class is well on the way to solving the problem, the instructor will often go to the board and sketch the solution, addressing aspects of the problem that proved particularly troublesome.

III. Use of "Flash Cards"

Each of our students has a set of six cards ($8\frac{1}{2} \times 5\frac{1}{2}$ in) labeled A, B, C, D, E, and F that are used to signal the instructor their answers to questions. Multiple-choice questions related to a particular concept are presented, either by overhead projection or written on the board. These questions usually precipitate lively class discussion regarding the different choices. Students within a group will debate with each other; sometimes one group challenges another group's decision. After a time of thought and discussion, students are asked to give a response by holding up one of their flash cards. (The final multiple-choice option may be "Don't Know" or "Not Sure" to encourage all students to participate.)

We have used the cards in three different ways: (1) all students hold up their flash cards simultaneously (this method best preserves the anonymity of the individual responses); (2) students hold up their cards as soon as they think they have the answer; (3) all "A" responses are solicited, then all "B's," and so on (omitting the "Don't Know" option). The instructor surveys the flash cards and reports the breakdown of responses. If there is substantial support for two or more choices, students are encouraged to give arguments in favor of their response; this frequently leads to further discussion and debate. We try to use flash-card questions very frequently, sometimes as many as ten times in a single class period.

Flash-Card Questions. Questions employed with the flash cards emphasize qualitative and proportional reasoning, solution strategies for problems (such as free-body diagrams), order-of-magnitude estimates, and vector concepts of magnitude and direction. (Many such examples are in the Workbook by Reif.[12]) Specific quantitative responses are de-emphasized, but are still solicited to culminate the analysis of a particular problem. We stress questions such as: "Is quantity A greater than, less than, or equal to zero? Greater than, less than, or equal to quantity B?" "If A is doubled, would B be doubled, quadrupled, or unchanged?" "Does vector C point north, south, east, or west? Is its magnitude closer to 10, 100, 1000, or 10^6?" The challenge for the student thus becomes one of determining which parameter or relationship is applicable to a particular question, and understanding its meaning, in contrast to simple numerical substitution or algebraic manipulation. (We sometimes have students practice with straightforward "plug-in" exercises as preparation for the more challenging qualitative questions.)



Kandiah Manivannan *is assistant professor of physics in the Department of Chemistry and Physics at Southeastern Louisiana University (Hammond, LA 70402). He received a Ph.D. in theoretical physics from SUNY at Stony Brook in 1986 and then began a research career in experimental and theoretical biophysics. His current research centers on the cellular mechanisms of biological clocks. He was an NSF Fellow at the Center for Biological Timing at the University of Virginia, where he was also visiting assistant professor of physics. He has a special interest in physics instructional demonstrations and has presented dozens of demonstrations to audiences of all ages and educational levels. He is involved in physics education reform aimed at increasing the level of active student participation in the classroom learning environment.*

In addition to preparing multiple-choice responses in advance, we have also allowed them to develop in tandem with class discussions. Students are asked to propose various answer options, and then the class "votes" on the options using the flash cards.

Flash-Card Feedback. Flash-card responses provide feedback to the instructor on two key parameters: (1) student misconceptions regarding the topic under discussion, and (2) pace of student understanding in the class as a whole. The instructor gets some feel for the degree of student comprehension by how quickly and confidently they are able to show their cards. Flash-card responses also offer students a means of testing the level of their understanding of the topic under discussion. Moreover, students see that others hold the same misconceptions. If the number of incorrect responses is high—for example, 30% or more—the instructor takes additional time to discuss that particular question before moving on.

For instance, after introducing the definition of acceleration, and discussing examples, the following question (taken from a widely used test bank[13]) was asked: *A ball is thrown vertically upward from the surface of the Earth. Consider the following quantities: (1) the speed of the ball; (2) the velocity of the ball; (3) the acceleration of the ball. Which of these is (are) zero when the ball has reached the maximum height? (A) 1 only; (B) 2 only; (C) 1 and 2; (D) 1 and 3; (E) 1, 2, and 3.* There were 60 students in the class; the numbers of students supporting each response were 0, 0, 15, 20, and 25, respectively. A spirited and intense discussion among the students followed (with guidance from the instructor), and continued for over 20 minutes. (Flash cards may also be used to gauge improvement in student understanding that results from class discussion.)

Sample Problem. It is possible to take a fairly complicated problem, involving several different concepts, and break it down into conceptual elements. We work through the problem piece by piece, with constant interaction and feedback from the students through the use of the flash cards.

In the Sample Problem given here, the essential steps leading to the solution are dealt with in questions 1 through 8. (Each successive question is presented only after the preceding one has been answered and discussed.) After the class completes these successfully, they proceed to the quantitative phase in questions 9 and 10. In question 9, the instructor will first point to one of the cells in the table—for instance, the cell referring to "Weight force/*x* direction"—and ask the class to hold up the flash-card letter of the appropriate response. In this way, all the cells in the table will be filled in, one by one. Finally, students may be asked to complete the problem by finding the answer to question 10 and checking it with those seated next to them, or with other student groups.

IV. Assessment

We encourage students to prepare for, attend, and participate attentively in class by offering frequent in-class assessment measures that contribute to students' overall grades. In addition to the traditional exams and quizzes, we have used several methods of having students solve quiz questions by working together in groups. Reference to notes, or to both notes and textbook, may be allowed. Students work in groups of two, three, or four, and groups may be allowed to confer with each other. Individual students may be permitted to "dissent" from a

Sample Problem

A 25.0-kg block has been sliding on a frictionless, horizontal ice surface at 2.00 m/s. Suddenly it encounters a large rough patch where the coefficient of kinetic friction is 0.05. How far does the block travel on this rough surface? *[Questions 1 through 10 refer to the motion on the rough surface.]*

1. How many *different* forces are now acting on the block? (Ignore air resistance.)
   A. 0 B. 1 C. 2 D. 3 ← E. 4 F. 5

2. What is the direction of the weight force? *(See Fig. 1.)*
   A. B. C. D. ← E. F.

3. What is the direction of the normal force?
   A. ← B. C. D. E. F.

4. What is the direction of the frictional force?
   A. B. C. D. E. F. ←

5. Is the block accelerating?
   A. Yes ←
   B. No
   C. Not enough information

6. What is the acceleration in the *y* direction?
   A. Greater than zero
   B. Less than zero
   C. Equal to zero ←
   D. Not enough information

7. What is the acceleration in the *x* direction?
   A. Greater than zero
   B. Less than zero ←
   C. Equal to zero
   D. Not enough information

8. How many forces are directly causing the acceleration in the *x* direction?
   A. 0 B. 1 ← C. 2 D. 3 E. 4 F. 5

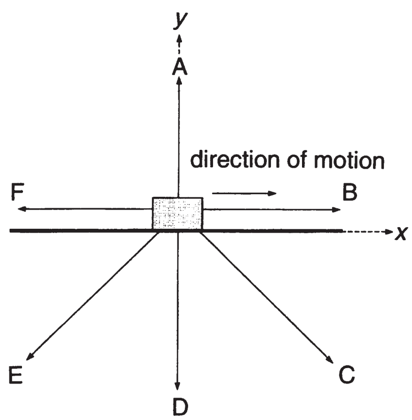


Fig. 1.

9. Put the appropriate letters in each box of the table:

	x direction	*y* direction
Weight Force	*[C]*	*[B]*
Normal Force	*[C]*	*[A]*
Friction Force	*[D]*	*[C]*
Total	*[F]*	*[C]*

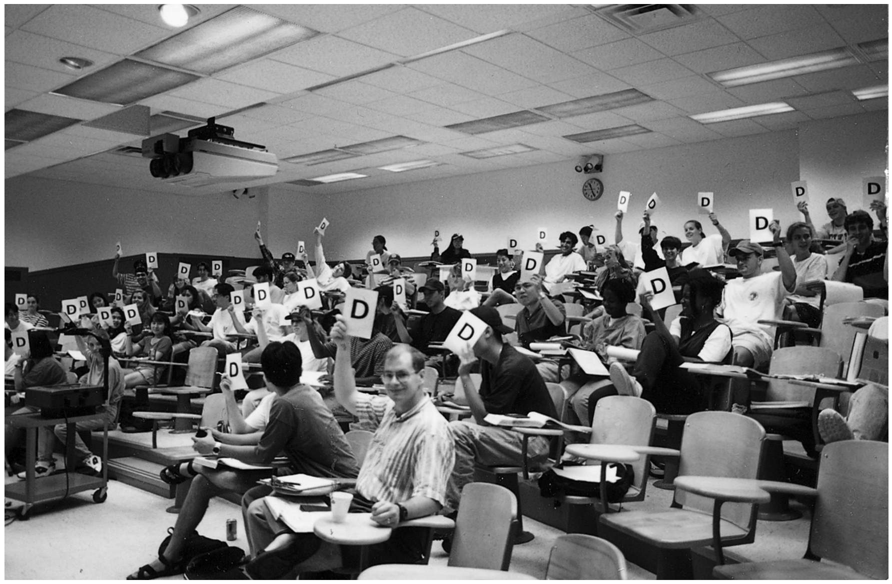
[Correct answer options are indicated by letters in brackets.]

A. + 245 N
B. − 245 N
C. 0 N
D. − 12.25 N
E. + 12.25 N
F. ma_x

10. Find the *x* component of the acceleration, and use it to determine the distance traveled. *[-0.49 m/s^2; 4.08 m]*

group response, handing in one of their own instead.

"Class Quizzes" are based solely on flash-card responses. If more than 50% of the class gives the correct response, each student in attendance receives credit for a 100% score on that quiz; otherwise, all receive a score of zero. "Group Quizzes" involve written responses that are handed in, with each person in the group getting the same grade. In "Challenge Quizzes," which generally involve more difficult questions, *each student* in a group is required to state how many points (up to 100% of the maximum possible quiz score) they want to "gamble" on their group's written response. Correct responses are awarded the number of points wagered, while incorrect responses result in a *loss* of that same number of points from the students' overall grade. (Typically, weaker students are not willing to put any points at risk.) When we use a multiple-choice format for the quizzes, students are often asked to report their responses by using the flash cards (after the quizzes have been collected). This allows instant feedback and discussion of the quiz problems.



Our Findings

Traditional Lecture Presentation Communicates Little to Students. We have found that many relatively simple concepts that are traditionally "covered" in a few minutes of lecture time turn out to be profoundly confusing to students even after extended thought and discussion. *[Example: The only force (ignoring air resistance) acting on a projectile during its flight is gravity, and the horizontal component of the projectile's acceleration is zero.]* Ideas that instructors may consider too trivial for more than a passing reference have been found to stump many students when they are asked to make use of them in problems. *(Example: Find the total momentum of a pair of objects sitting at rest.)* Results of using the interactive methods suggest that traditional methods of cursory treatment of important concepts during lecture yield little student understanding.

Instructors Must Have a Clear Concept of What They Intend Students to Learn. If the instructor's goal is for students to be trained to recognize certain types of quantitative problems, find the appropriate equation that may be used to solve the problem, and then use it to obtain a correct quantitative answer to a nearly identical problem presented to them—then these interactive methods may not be appropriate. If, however, the goal is for students to obtain a thorough understanding of certain basic concepts so that they may be able to devise novel solution methods for relatively *unfamiliar* problems in a variety of contexts, traditional methods do not appear to be very effective and the interactive methods may hold greater promise.

Outcome of Using Interactive Learning May Depend on Students' Level of Preparation. We have used these techniques both at Southeastern Louisiana University and at the University of Virginia at Charlottesville. The subjective response of the (typically much better prepared) students at UVa was more positive than of those at SLU. There is little doubt that the educational background of the students taking a particular introductory physics course is likely to have a significant effect on the outcome of interactive learning methods.

Students Accustomed to Traditional Methods May Be Suspicious of and Hostile Toward Interactive Learning.[14] Many students are accustomed to educational methods that emphasize memorization and formulaic learning. As a result, a significant number of the students in some of our classes showed a great distaste for—and were even resentful of—the inherent uncertainty and confusion that is an essential phase of the process of actively struggling to master difficult concepts. "Why can't you just tell us the answer?!" was a characteristic remark. Some students commented that the use of the flash cards was "a waste of time."

Interactive Methods Have Little Hope of Success If Used Only in Isolated Situations. Students who are accustomed exclusively to traditional memorization-based methods are unlikely to be receptive to highly interactive, concept-driven learning. Students who have little experience in pursuing *extended, time-consuming thought processes* to master difficult concepts—involving question-and-answer dialogue and discussion—tend to find such processes difficult, distasteful, frustrating, and confusing.

Conclusion

Interactive methods such as those described here focus on the goal of having substantial effective learning take place *during class time*. The objective is to ensure that students do not simply listen passively to the words spoken by the instructor, but that they become intensely involved in learning and applying targeted concepts. The physics lecture as a forum for "covering" large numbers of topics is sacrificed. What takes its place is an environment that becomes an expression of the instructor's skill in guiding and leading students through the complex thought processes required to understand and apply physics concepts. It is intended that these experiences in conceptual learning—particularly those few moments when the students can say, "Aha, *now* I see..."—will form a basis for students' out-of-class study that is at least as effective as the traditional lecture.

Acknowledgment

Remarks by Professor Susan Wyckoff during a group discussion catalyzed much of the work described in this paper.

References

1. P.W. Laws, "New approaches to undergraduate teaching: Introductory courses," in *Physics Departments in the 1990s,* edited by G.M. Crawley and B.V. Khoury (American Association of Physics Teachers and American Physical Society, College Park, MD, 1993), pp. 33–37.
2. J. M. Wilson, "The CUPLE physics studio," *Phys. Teach.* **32**, 518 (1994).
3. R.J. Beichner, "Improving the effectiveness of large-attendance lectures with animation-rich lecture notes," *AAPT Announcer* **24** (4), 64 (1994).
4. R. Littauer, "Instructional implications of a low-cost electronic student response system," *Educ. Tech.* (October 1972), p. 69; J. Mestre et al., "Turning passive lectures into interactive learning environments with a classroom communication system," *AAPT Announcer* **24** (4), 63 (1994); R. Swanson and C.J. Roddy, "Interactive classroom system promotes action and interaction," *AAPT Announcer* **24** (4), 64 (1994); J.C. Webb et al., "Collaborative learning and insights on students' thinking: Computer technology offers new possibilities for large classes," *AAPT Announcer* **24** (4), 64 (1994).
5. L.C. McDermott, "Guest Comment: How we teach and how students learn—a mismatch?", *Am. J. Phys.* **61**, 295 (1993).
6. F. Reif, "Millikan Lecture 1994: Understanding and teaching important scientific thought processes," *Am J. Phys.* **63**, 17 (1995).
7. D.P. Maloney, "Research on problem solving: Physics," in *Handbook of Research on Science Teaching and Learning*, edited by D.L. Gabel (Macmillan, New York, 1994), pp. 327–354.
8. D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," *Phys. Teach.* **30**, 141 (1992).
9. D. Hestenes and M. Wells, "A mechanics baseline test," *Phys. Teach.* **30**, 159 (1992).
10. *Cf.* C. Swartz, "Editorial: The pause that refreshes," *Phys. Teach.* **30**, 200 (1992); A. Van Heuvelen, *Am. J. Phys.* **59**, 891 (1991).
11. E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Englewood Cliffs, NJ, at press).
12. F. Reif, *Understanding Basic Mechanics Workbook* (Wiley, New York, 1995).
13. D.T. Marx and M.J. Comella, *Test Bank for* Physics, 3rd ed., by Cutnell and Johnson (Wiley, New York, 1995), p. 9.
14. For another perspective on this point, *cf.* R.W. Chabay and B.A. Sherwood, *Instructor's Manual to Accompany Electric and Magnetic Interactions* (Wiley, New York, 1995), pp. 8–9.

Increasing Active Student Participation in the Classroom Through the Use of "Flash Cards"

Kandiah Manivannan and David E. Meltzer

Department of Chemistry and Physics, Southeastern Louisiana University

Large lecture classes make it difficult to maintain high levels of student-faculty interaction; in these classes, students traditionally play a relatively passive role. We have been making use of techniques for increasing active student participation in the lecture classroom, and for raising the level of interaction between students and instructors. A central element in these methods is the use of "flash cards" which allow students to instantaneously indicate to the instructor their responses to multiple-choice questions. Students use 8.5 x 11 inch flash cards, labeled "A," "B," "C," "D," "E," and "F" to signal their responses to the instructor. Flash-card questions emphasize qualitative and proportional reasoning, solution strategies for problems, order of magnitude estimates, etc. Responses provide feedback to the instructor on student misconceptions, and pace of student understanding. Here we show an example of how we break down a conventional problem into conceptual elements—a so-called "problem dissection"—which can then be formed into flash-card questions. [Meltzer, D.E. and K. Manivannan, *Phys. Teach.* **34**, 72-76, 1996.]

PROBLEM DISSECTION TECHNIQUE

It is possible to take a fairly complicated problem, involving several different concepts, and break it down into conceptual elements. We work through the problem piece by piece, with constant interaction and feedback from the students through the use of flash cards. In the sample problem presented here, the essential steps leading to the solution are dealt with in questions 1 through 15. (It is important to note that each successive question is presented only after the preceding one has been answered and discussed.) After completing these, the students will proceed to the quantitative phase in the remaining questions.

SAMPLE PROBLEM

Four charges are arranged on a rectangle as shown in Figure 1. ($q_1 = q_3 = +10.0$ μC and $q_2 = q_4 = -15.0$ μC; a= 10.0 cm and b = 15.0 cm.) Find the magnitude and direction of the resultant electrostatic force on q_1.

CP399, *The Changing Role of Physics Departments in Modern Universities*: Proceedings of ICUPE, edited by E. F. Redish and J. S. Rigden

Figure 1. Diagram used for problem dissection

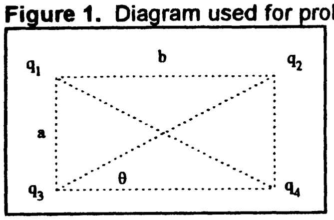
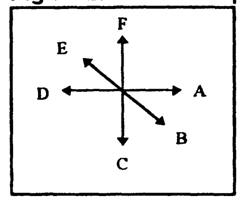


Figure 2. Direction options



1. How many forces (due to electrical interactions) are acting on charge q_1? **(A)** 0 **(B)** 1 **(C)** 2 **(D)** 3 **(E)** 4 **(F)** Not sure / Don't know. For questions **2,3** and **4**, refer to Figure 2 and pick a direction from the choices **A, B, C, D, E, F**. **2.** Direction of force on q_1 due to q_2. **3.** Direction of force on q_1 due to q_3. **4.** Direction of force on q_1 due to q_4. Let F_2, F_3 and F_4 be the *magnitudes* of the force on q_1 due to q_2, due to q_3 and due to q_4 respectively. **5.** F_2 is given by **(A)** $kq{:}q_2/a^2$ **(B)** kq_1q_2/b^2 **(C)** $kq_1q_2/(a^2+b^2)$ **(D)** $kq_1q_2/\_(a^2+b^2)$ **(E)** None of the above **(F)** Not sure/Don't know. Questions **6 & 7** are similar to question **5** with the subscript 2 changed to 3 and 4 respectively. At this point (after discussing questions **1** through **7**), the instructor draws the correct vector diagram showing all the forces acting on charge q_1 and asks the following questions: [For questions **8** through **13**, pick the answer from the list of six choices given below.] **(A)** F_2 **(B)** $-F_3\cos\theta$ **(C)** $F_3\sin\theta$ **(D)** $-F_4$ **(E)** 0 **(F)** None of the above. **8.** X-component of force on q_1 due to q_2; **9.** Y-component of force on q_1 due to q_2; **10.** X-component of force on q_1 due to q_3; **11.** Y- component of force on q_1 due to q_3; **12.** X-component of force on q_1 due to q_4; **13.** Y-component of force on q_1 due to q_4; **14.** *Write down* the X-component of the *net* force on q_1; **15.** *Write down* the Y-component of the *net* force on q_1; **16.** What is the value of angle θ? **(A)** 29° **(B)** 34° **(C)** 40° **(D)** 48° **(E)** 57° **(F)** Not sure/Don't know; **17.** Calculate the magnitude of the resultant force on q_1; **18.** Calculate the direction of the resultant force on q_1.

STUDENTS' ATTITUDES TOWARDS FLASH CARDS

These methods were employed in an algebra-based general physics course at the University of Virginia at Charlottesville. A questionnaire entitled "Flash Cards Student Survey" was distributed to all of the students enrolled in this course (N=41). The response of the students was very positive. On a 1-5 rating scale, students gave a mean response of 4.1 to the statements "gained better understanding," "paid more attention," and "instructor was more aware of problems." [4 = "Agree"; 5 = "Agree Strongly."] Mean response was 2.2 [2 = "Disagree," 3 = "Neutral"] to the statements "waste of time," "disliked holding up cards," and "disliked working in groups."

Nontraditional Approach to Algebra-Based General Physics

David E. Meltzer

Department of Chemistry and Physics, Southeastern Louisiana University

In order to improve the degree of conceptual learning in our algebra-based general physics course, the second semester (of a two-semester sequence) has been taught in a nontraditional format during the past year. The key characteristics of this course were: 1) Intense and continuous use of interactive-engagement methods and cooperative learning; 2) coverage of less than half of the conventional number of topics, 3) heavy emphasis on qualitative questions as opposed to quantitative problems, 4) adjustment of the pacing of the course based on continuous (twice per week) formative assessment.

The students enrolled in the course were relatively poorly prepared, with weak mathematical skills. Open-book quizzes stressing qualitative concepts in electricity and magnetism were given twice per week; most were given in "group quiz" format, allowing collaboration. Exams (also open-book) were all done individually. Most of the class time was taken up by quizzes, and by interactive discussion and group work related to quiz questions. New topics were not introduced until a majority of the class demonstrated competence in the topic under discussion.

Despite lengthy and intensive focus on qualitative, conceptual questions and simple quantitative problems, only a small minority of the class ultimately demonstrated mastery of the targeted concepts. Frequent testing and re-testing of the students on basic concepts disclosed tenacious persistence of misconceptions.

STUDENT PREPARATION

Students had completed the first semester of a two-semester sequence in algebra-based general physics. The first semester concentrated on vector concepts and Newtonian mechanics. Students had completed college algebra and trigonometry.

As measured by post-tests on the Force Concept Inventory, students' grasp of mechanics concepts was weak (mean post-test score = 39%). Students' mathematical skills were weak as well, as they generally had difficulty using elementary trigonometry (Pythagorean theorem, finding unknown sides of right triangles) and elementary algebra.

CP399, *The Changing Role of Physics Departments in Modern Universities*: Proceedings of ICUPE, edited by E. F. Redish and J. S. Rigden

"REDUCED SYLLABUS"

The number of topics covered was much smaller than the conventional number. Moreover, only the basic features of each topic were emphasized, with few details or applications:

1. Coulomb's Law; Electric field and force in two dimensions.
2. Electric potential and potential energy.
3. Capacitance; Ohm's law; Elementary series and parallel circuits; power.
4. Magnetic field of: straight current-carrying conductor; coil; current loop.
5. Magnetic force on straight current-carrying conductor and on moving charge.
6. Torque on current loop in magnetic field; induced currents.
7. Electromagnetic Waves (spectrum; $c = f\lambda$); Law of Reflection; Snell's Law.

Problem Types

- Emphasis on qualitative and "proportional reasoning" problems.
- Elementary quantitative applications: only elementary (high-school level) algebra and trigonometry used.
- Multiple-choice questions, typically with 7-12 answer options.
- High level of redundancy: slight variations of basic problems given repeatedly (up to ten times) throughout semester, on quizzes and exams.

Instructional Methods

1. Highly "interactive" classes:
   - Little lecture; much student-faculty interaction using "flash cards." [Cf. Meltzer, D.E. and K. Manivannan, *Phys. Teach.* **34**, 72-76,]
   - "Tutorial" format: students work in groups as instructor circulates throughout room and responds to requests for help.
   - Most of class time taken up by quizzes, and by interactive discussion and group work related to quiz questions.
2. Incessant formative assessment:
   - Quizzes twice per week (20% of class time).
   - Every quiz and exam "cumulative" (all topics may be covered).
   - All quizzes and exams "open book" and "open notes."
   - Most quizzes were "group quizzes" (done collaboratively); all exams done individually.

- Tremendous redundancy of questions: basic problems given in slightly revised form up to ten times over the course of the semester.

3. Exam and Quiz questions all had "extra credit" option: If student chose this option, correct answer counts extra, but incorrect answer results in 50% of value of question being ***deducted*** from exam score. This encouraged students to reflect critically on their confidence in their responses.
4. In addition to qualitative, conceptual questions being assigned for homework and review, selected "back of the chapter" quantitative problems were recommended for additional study.
5. Pacing of course continuously adjusted, guided by ongoing assessment. Concepts revisited repeatedly when testing revealed inadequate mastery.

Use of in-class physics demonstrations in highly interactive format

Kandiah Manivannan, Department of Physics, Astronomy, and Materials Science
Southwest Missouri State University, Springfield, MO 65804
David E. Meltzer, Department of Physics & Astronomy, Iowa State University, Ames, IA 50011

We show how traditional classroom demonstrations may be converted into active-learning experiences through linked multiple-choice question-and-answer sequences. Sample question sequences and worksheet materials are presented, as well as preliminary assessment data.

INTRODUCTION

In-class demonstrations have been considered by physics instructors to be a very important part of teaching physics. Physics demonstrations have been around for many years, and physics teachers and researchers have written numerous articles and books on classroom demonstrations. Demonstrations can certainly make physics classes fun and entertaining, and they may also stimulate students' interest and curiosity. Surveys conducted for the Introductory University Physics Project indicate that students believe demonstrations help them to better visualize and think about physics.

However, despite these positive aspects of physics demonstrations, there is a growing body of evidence suggesting that traditional in-class demonstrations are not very effective in promoting conceptual understanding of physics. One important factor is the lack of active participation and interaction of students during physics demonstrations. Recent research studies indicate that students who saw traditional physics demonstrations in a course fared no better than students who did not see the demonstrations.[1,2] The data do suggest, however, that there is at least a small improvement in performance when students have to predict the outcome of a demonstration before seeing it. Based on these and other studies, it has become increasingly clear that some form of interactive engagement is essential to maximize the effectiveness of classroom demonstrations.

At this time there are relatively few research-based curricular materials available for physics demonstrations. The pioneering work of Sokoloff and Thornton[3] on interactive lecture demonstrations (ILD) is probably the most comprehensive curricular material of that type available today. Their published results on ILDs indicate dramatic learning gains for students who were taught using ILDs compared to students who took a traditional course. ILDs require the use of Microcomputer-Based Laboratory equipment.

We have been developing new curricular materials on interactive physics demonstrations that would promote active learning in physics classes. Our goal is to produce activities that are suitable for any classroom setting and can easily be implemented without any additional resources or logistical support (such as computer hardware or teaching assistants). Our teaching strategy can be used with “high-tech” demonstrations as well as with those that are low-tech. The central feature is the use of the problem-dissection technique[4] to break a given physics demonstration into several conceptually linked mini-demonstrations. The demonstrations are presented to the class in a sequence while utilizing techniques (such as "flash cards") for acquiring immediate feedback from all the students in the class simultaneously. We find this approach very effective in helping students construct a deeper understanding of physical concepts through step-by-step confrontation with their alternate conceptions. Since these innovative elements are based on findings of physics education research, one may hope that student learning might be significantly more effective than in a course taught using

This paper appeared in *Proceedings of the Physics Education Research Conference, Rochester, New York, July 25-26, 2001*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (PERC, Rochester, New York, 2001), pp. 95-98.

traditional methods. In this paper we report our preliminary findings on a specific interactive demonstration activity involving free-falling objects.

METHOD

The most important features of the interactive demonstration curricular materials are as follows:

1. The curricular materials or worksheets are designed to strongly promote student-student as well as student-faculty interaction in the classroom.
2. The initial prediction of the outcome of the demonstration and the subsequent discussion among neighboring students – as well as the following class-wide discussion – are very important parts of the demonstration activity.
3. Activities are based on the premise that the explanation of even a very simple physics demonstration invariably hinges on a lengthy chain of concepts and reasoning.
4. The problem-dissection technique[4] is used to break a given physics demonstration into several conceptually linked mini-demonstrations.
5. The mini-demonstrations are presented as a sequence in a pre-determined order. Breaking down the main demonstration into smaller component demonstrations is very effective in helping students construct a deeper understanding of physical concepts through step-by-step confrontation with their alternate conceptions.
6. We utilize techniques (such as the use of flash cards, show of hands, or electronic wireless transmitters) for acquiring immediate feedback from all the students in the class. (See Ref. 4.)

Sample Interactive Demonstration

In order to explore the physics of freely falling objects, a dime and a quarter are dropped simultaneously from the same height. The question to be answered is: "Which object would hit the floor first?" We then design an interactive demonstration sequence consisting of several conceptually linked mini-demonstrations to address important conceptual issues associated with free-fall. A set of multiple-choice questions for this demonstration sequence was developed for use with flash cards. Worksheets were designed on which students were required to write predictions and draw motion diagrams. Excerpts from the questions and worksheets are shown below.

Initial Flash-Card Question

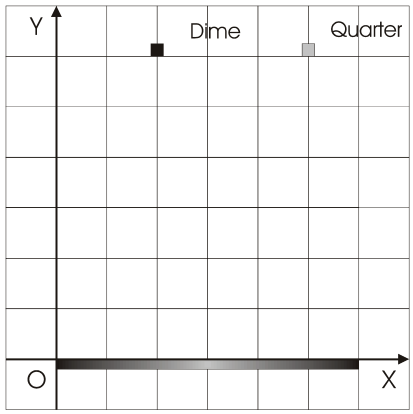
A dime and a quarter are dropped simultaneously from the same height. Which one will hit the floor first?

A. The dime will hit the floor first.
B. The quarter will hit the floor first.
C. Both hit the floor at the same time.
D. I am not sure/ I don't know.

Students are always required to make predictions of the outcome of the demonstration by holding up flash cards, or using wireless electronic transmitters. They may "vote" before and/or after talking to their neighbors. At the appropriate times, the instructor will provide assistance. Once the first demonstration is complete and students have finished their discussions and worksheet activities, the process is continued by asking (one by one) seven other closely related questions.

Follow-up Questions

These questions all follow the model of the first one, but in each case different pairs of objects are compared. (Each question is accompanied by a separate worksheet for student responses). The questions ask students to compare the rates of fall of the following pairs of items: (1) dime and piece of paper; (2) dime and piece of crumpled paper; (3) coffee filter and loaded coffee filter; (4) book and piece of paper; (5) piece of paper resting on top of book; (6) dime and piece of paper inside an open chamber; (7) dime and piece of paper inside closed, evacuated chamber.



(1) A dime and a quarter are dropped simultaneously from the same height. Which one will hit the floor first?

A. The dime will hit the floor first.
B. The quarter will hit the floor first.
C. Both hit the floor at the same time.
D. I am not sure/I don't know.

[Most of the following instructions were given either verbally or written on the blackboard:]

***Before* the demonstration:** In the space below, write down your prediction about the outcome of the demonstration. Write a few sentences to justify/explain your thinking.

***After* seeing the demonstration** *[sometimes, before seeing]*: Draw motion diagrams to represent the motion of the dime and the quarter. Do this directly on the diagram above.

Fig. 1. Sample activity and worksheet excerpts.

ANALYSIS OF CLASS-TEST DATA

The method described here has been implemented twice in the context of an introductory algebra-based mechanics class. Although the available assessment data are limited and inconclusive, they do suggest the possibility that significant improvements might be ascribed to interactive demonstrations.

Here we focus specifically on an analysis of students' responses to Question #1 on the Force Concept Inventory. This question targets precisely the concept that is the subject of the sample materials presented in this paper. For reference, we cite FCI #1:

> Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a single-story building at the same instant. The time it takes the balls to reach the ground below will be
>
> A. about half as long for the heavier ball as for the lighter one.
> B. about half as long for the lighter ball as for the heavier one.
> C. about the same for both balls.
> D. considerably less for the heavier ball, but not necessarily half as long.
> E. considerably less for the lighter ball, but not necessarily half as long.

The first author implemented interactive demonstrations as described here for the first time during Fall 1998 at Southeastern Louisiana University (SLU). They were also used during Summer 2000 at Southwest Missouri State University (SMS). Previously, he had taught the same course (algebra-based mechanics) three times; twice at SLU and once at the University of Virginia (UVa). Table I presents data from all five of these classes.

Table I. Assessment data (mean pre- and posttest scores, and Hake normalized gains <g>) for algebra-based mechanics courses taught by first author. Courses in **boldface** (**SLU 98** and **SMS 00**) used fully structured interactive demonstration (described in this paper) for free-falling objects. The other courses used more limited demonstrations involving only straightforward predictions.

	n	FCI #1 pre	FCI #1 post	FCI #1 <*g*>	Full FCI <*g*>
UVa 95	*55*	71%	82%	0.38	0.26
SLU 96	*75**	57%	78%	0.49	0.15
SLU 97	*66*	79%	85%	0.29	0.30
SLU 98	***31***	**65%**	**100%**	**1.00**	**0.34**
SMS 00	***22***	**36%**	**91%**	**0.86**	**0.50**

*non-matched sample; $n_{pre}=73$; $n_{post}=77$

All five of these classes were taught with interactive-engagement methods. In particular, they made heavy use of highly interactive "lectures" using the flash-card method. (These methods have been described in detail elsewhere; see Reference 4.) With specific regard to the concept of a freely falling object, there were significant differences in the instructional method used.

A full implementation of interactive demonstrations as described here, with a full sequence of conceptually linked questions, was used only for the SLU 98 and SMS 00 courses. The other courses used a format in which students' predictions were, indeed, solicited before the demonstration took place. However, in those cases there was little or no attempt to structure a series of tightly linked interactive demonstrations with a single theme as we have described here. (One of the consequences of using the full interactive method is that several additional minutes are required for the activity.)

This full implementation was only carried out for the concept of the free-falling object, and that is why data only for FCI #1 are examined in this section. The most remarkable result is that every single one of the 31 students in SLU 98 got that question correct on the posttest, while only 65% had it correct on the pretest. This performance was far better than that of SLU 97 or SLU 96. The 100% posttest score is significantly better than the 85% (p = 0.02) of SLU 97, the 78% (p = 0.004) of SLU 96, and the 82% (p = 0.01) of UVa, even though the pretest score for SLU 98 is relatively low. Although the overall FCI normalized gain of 0.34 for SLU 98 is nearly the same as that for SLU 97, its $<g>$ for FCI #1 is far higher. This very high gain for FCI #1 is nearly matched by the SMS 00 course, which also used the interactive demonstration method.

SUMMARY

We have described a method for implementing classroom demonstrations in a highly interactive fashion to promote active learning. The key aspects of this method are (1) create a carefully structured sequence of conceptually linked demonstrations, and (2) promote students' active engagement with the demonstrations by soliciting their input on multiple-choice questions using a classroom communication system, such as flash cards. Preliminary data from classroom testing are promising and suggest that this method may be able to produce significant learning gains.

We believe that the "highly interactive" format may increase the pedagogical effectiveness of the demonstrations, as is suggested by the data in Table I. Since the *same* instructor employing interactive-engagement methods taught *all five* courses, it is unlikely that "teaching to the test" produced this increase. In this regard, we intend to examine correlations among performances on FCI questions 1, 3 and 13 (all related to free-fall). Although we do not now have student achievement or attitude data to correlate with FCI data, in the future we plan to investigate these variables. We plan to develop and test additional interactive demonstrations on other topics to further explore the potential of this method.

REFERENCES

1. Pamela A. Kraus (1997). *Promoting Active Learning in Lecture-Based Courses: Demonstrations, Tutorials, and Interactive Tutorial Lectures.* Ph.D. dissertation, U. Washington, UMI #9736313.

2. J. P. Callan, C. H. Crouch and E. Mazur, "Classroom demonstrations: education or mere entertainment?" AAPT Announcer **29(4)**, 89 (1999).

3. D. R. Sokoloff and R. K. Thornton, "Using interactive lecture demonstrations to create an active learning environment," Phys. Teach. **35**(10), 340-347 (1997).

4. D. E. Meltzer and K. Manivannan, "Promoting interactivity in physics lecture classes," Phys. Teach. **34**(2), 72-76 (1996).

K.M. acknowledges the support of a USDE-PT3 Implementation grant P342A99041 and FFT grants.

Transforming the lecture-hall environment: The fully interactive physics lecture

David E. Meltzer[a)]
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

Kandiah Manivannan
Department of Physics, Astronomy, and Materials Science, Southwest Missouri State University, 901 South National Avenue, Springfield, Missouri 65804

(Received 19 September 2001; accepted 29 January 2002)

Numerous reports suggest that learning gains in introductory university physics courses may be increased by ‘‘active-learning’’ instructional methods. These methods engender greater mental engagement and more extensive student–student and student–instructor interaction than does a typical lecture class. It is particularly challenging to transfer these methodologies to the large-enrollment lecture hall. We report on seven years of development and testing of a variant of Peer Instruction as pioneered by Mazur that aims at achieving virtually continuous instructor–student interaction through a ‘‘fully interactive’’ physics lecture. This method is most clearly distinguished by instructor–student dialogues that closely resemble one-on-one instruction. We present and analyze a detailed example of such classroom dialogues, and describe the format, procedures, and curricular materials required for creating the desired lecture-room environment. We also discuss a variety of assessment data that indicate strong gains in student learning, consistent with other researchers. We conclude that interactive-lecture methods in physics instruction are practical, effective, and amenable to widespread implementation.
[DOI: 10.1119/1.1463739]

I. INTRODUCTION

Numerous investigations in recent years have shown active-learning methods to be effective in increasing student learning of physics concepts. These methods aim at promoting substantially greater engagement of students during in-class activities than occurs, for instance, in a traditional physics lecture. A long-standing problem has been that of transporting active-learning methods to large-enrollment classes in which 50–300 students sit together in a single classroom.

An important breakthrough in addressing this problem was the 1991 introduction of the Peer Instruction method by Eric Mazur at Harvard University.[1] This now widely adopted method restructures the traditional lecture class into a series of short lecture presentations punctuated by a series of ‘‘ConcepTests.’’ These are qualitative multiple-choice questions to which all students in the class simultaneously respond, both before and after discussion.

In this paper we describe a variant of Peer Instruction that we have developed and tested. It carries the transformation of the physics lecture-room environment several steps further, aiming at the achievement of a virtually continuous dialogue between students and instructor of a type ordinarily characteristic only of one-on-one (or one-on-few) instruction that takes place, for example, in the instructor’s office with a handful of students present. This ‘‘fully interactive lecture’’ offers a useful option for physics instructors who want to maximize the potential for instructor–student interaction in the large-classroom environment. We have employed these methods in our classes over the past seven years at Southeastern Louisiana University (SLU), the University of Virginia, Iowa State University (ISU), and Southwest Missouri State University.

The basic elements of an interactive lecture strategy have been described by Mazur.[1] In this paper we broaden and extend that discussion, explaining in detail how the lecture component in large-classroom instruction may be almost eliminated. Depending on the preferences of the instructor and the specific student population, this strategy may yield worthwhile learning outcomes. To carry out the rapid back-and-forth dialogue observed in one-on-one instruction in large-enrollment classes requires a variety of specific instructional strategies, an unusual form of preparation by the instructor, and specific characteristics of the curricular materials.

In Sec. II we review the research related to student learning in physics lecture classes. In Sec. III we give an overview of our general strategy for creating interactive lectures, and the student response systems necessary to that strategy are discussed in Sec. IV. In Sec. V we outline the format of the fully interactive lecture class, while Sec. VI contains a detailed, almost verbatim, excerpt from an actual class. This excerpt is analyzed in Sec. VII. In Sec. VIII we discuss the printed curricular materials that have been developed for use with these instructional methods. In Sec. IX we discuss implementation issues, and in Sec. X we discuss the analysis of assessment data related to student learning in our classes. We offer some concluding remarks in Sec. XI.

II. A LONG-STANDING CHALLENGE: PROMOTING ACTIVE-LEARNING IN LARGE LECTURE CLASSES

A. Motivation: Student–instructor disconnect in large-enrollment classes

Recent research has cast serious doubt on the effectiveness of instruction for the majority of students enrolled in introductory physics courses, the most common setting for large-enrollment, lecture-based instruction. Not surprisingly, the large-enrollment lecture class is among the most challenging environments in which to achieve improved learning gains. It is very difficult for instructors to assess student learning and to implement any needed alterations in instruction in "real-time." Moreover, the high student/instructor ratio makes it difficult for instructors to engage students in instructional activities that go much beyond passive listening.

B. Limitations of the lecture approach: The case of physics

An increasing body of evidence suggests that instruction utilizing *only* lecture classes and standard recitations and labs results in relatively small increases in most students' understanding of fundamental concepts.[2–8] Complex scientific concepts are often not effectively communicated to students simply by lecturing about them—however clearly and logically the concepts may be presented.[9–12] Students taught exclusively through lecture-based curricula are inclined to short-circuit the highly complex scientific thought process.[13,14] Lectures that are particularly clear and well-organized may, ironically, contribute to students' tendency to confuse the *results* of science with the scientific process itself. Students who avoid the intense mental struggle that often accompanies growth in personal understanding may never succeed in developing mastery over a concept.[15] In other words, students do not absorb physics concepts simply by being told (*or* shown) that they are true, and they must be guided to resolve conceptual confusion through a process that maximizes the active engagement of their mental faculties.

A term that is often used to characterize an instructional process of this type is "active learning," and the term "interactive engagement" (IE) has been used to describe the type of physics instruction that most effectively engenders active learning through discussion with peers and/or instructors.[4] Active learners are relatively efficient at learning physics concepts. They are perhaps most easily characterized as students who continuously and actively probe their own understanding in the process of learning new concepts. They frequently formulate and pose questions to themselves, constantly testing their knowledge. They scrutinize implicit assumptions, examine systems in varied contexts, and are sensitive to areas of confusion in their understanding. By contrast, the majority of students in introductory physics courses are unable to do efficient active learning on their own. In essence, they don't know the questions they need to ask. They are often unable to recognize when their own understanding is inadequate, and tend to lack confidence in their ability to resolve confusion. In order to carry through the learning process effectively, they require substantial guidance by instructors and aid from appropriate curricular materials.

There is good evidence that, in addition to improving learning by students who may not be natural active learners, interactive-engagement methods result in significant learning gains by the best students as well.[16,17] Pedagogical models that engage students in a process of investigation and discovery—often oriented around activities in the instructional laboratory—are specific types of interactive-engagement methods found to be effective.[5,6,10,17–20] The targeted concepts are in general not told to the students before they have the opportunity to follow through chains of reasoning that might lead them to synthesize the concept on their own. It is especially challenging to develop effective active-learning methods that lack a laboratory component, and the large lecture class is an inherently difficult environment in which to establish active learning.

C. Recent approaches to active learning in large physics lecture classes

The issue of how to increase attention and engagement of students during lecture courses is not unique to physics. Various systems have been designed that allow students in large classes to provide instantaneous responses to instructors' questions.[21–23] Other influential methods include "think-pair-share"[24] (periodic interruption of lectures for student discussion), and the "minute paper"[25] (students' written comments during the last minute of class). Various strategies have been reviewed by Bonwell and Eison.[26]

Physics educators have explicitly addressed the challenge of the large-class learning environment. Van Heuvelen[27,28] has developed "active-learning problem sheets"[29,30] for student use during class meetings in the lecture hall.[31] Mazur has achieved great success in popularizing Peer Instruction[1,32–35] by suspending a lecture at regular intervals with challenging conceptual questions posed to the whole class. Other early strategies for lecture classes have been described.[36–38] More recently, the group at the University of Massachusetts has developed and popularized interactive-lecture methods employing an electronic response system.[39–41] Poulis *et al.*[42] have also made use of interactive lecturing with an electronic system, and other electronic communication systems for use in lectures have been discussed by Shapiro,[43] and by Burnstein and Lederman.[44]

Other strategies for implementing active learning in large-enrollment classes have been described by Beichner *et al.*[45] and by Zollman.[46] Sokoloff and Thornton have adapted their very popular microcomputer-based laboratory materials, originated in collaboration with Priscilla Laws,[6,18,47] for use in large lecture classes in the form of "interactive lecture demonstrations."[48,49] Assessment data from several groups support the effectiveness of this method.[6,17,18] Novak and collaborators[50] have developed the "just-in-time teaching" method in physics lecture courses, incorporating some techniques similar to those used by Hestenes and his collaborators in the "modeling instruction"[51] method. Textbooks and workbooks with a high interactive component, usable in large classes, include those by Chabay and Sherwood[52,53] and by Knight.[54] There is good evidence for the effectiveness of both of these innovative curricular materials.[55,56] The interactive-lecture strategies to be discussed in this paper build on the recent history of efforts to improve instruction in large physics classes. Preliminary reports have been published,[57,58] and several workshops have been presented.[59]

Other important pedagogical reform methods focus more particularly on activities that occur in small-class laboratories or recitation sections associated with lecture courses. Among the most prominent are the *Tutorials in Introductory Physics*,[60,61] Collaborative Group Problem Solving,[62–65] and *RealTime Physics*,[66] along with its close relative, *Workshop Physics*.[67] Important research results related to instruction in large-enrollment physics classes have been reported by Kraus,[68] and Cummings *et al.*[17] have described a careful investigation of a technology-rich studio environment.[69]

III. TRANSFORMING THE LECTURE-ROOM ENVIRONMENT

Our goal is the transformation of the lecture class, to the furthest extent possible, to the type of instructional environment that exists in an instructor's office. When physics instructors have one or two students in their office, they would likely speak for just a few minutes, solicit some feedback, then continue the discussion based on that feedback. In the office, instructors can get a sense of where students are conceptually and of how well they are following the discussion. It is possible to tailor one's presentation to the students' actual pace of understanding. By asking students to consider each other's ideas, the instructor helps them to think critically about their own ideas. The key issue is whether it is *practical* to do this in a room filled with 100 or more students.

We (and others) have found that it *is* practical to bring about this transformation to a very great extent. Success hinges on two key strategies: (1) students need to be guided in a deliberate, step-by-step process to think about, discuss, and then respond to a carefully designed sequence of questions and exercises; (2) there must be a system for the instructor to obtain *instantaneous* responses from *all* of the students in the class *simultaneously*. This system allows instructors to gauge their students' thinking and to rapidly modify their presentation, subsequent questioning, and discussion of students' ideas. Our methods are a variant of Peer Instruction,[1,32–35] and are similar to methods used at the University of Massachusetts[39–41] and at Eindhoven.[42]

The basic objective is to drastically increase the quantity and quality of interaction that occurs in class between the instructor and the students and among the students themselves. To this end, the instructor poses many questions. Students decide on an answer, discuss their ideas with each other, and provide their responses using a classroom communication system. The instructor makes immediate use of these responses by tailoring the succeeding questions and discussion to most effectively match the students' pace of understanding.

In attempting to address the insufficiencies of the traditional lecture, the fully interactive lecture method that we employ essentially abandons any effort to utilize class time for presenting detailed and comprehensive explanations and derivations of physics principles. Instead, that time is used in much the same way as in one-on-one tutoring: there is a continual interchange of questions and answers between instructor and students. The instructor guides the students in step-by-step fashion to consider certain problems; the students listen, think, write or calculate, and then receive immediate feedback regarding the correctness of their responses, both from their classmates and from the instructor.

In abandoning lecture's traditional role of providing extensive and detailed background information, we must evidently utilize other means for achieving that objective. The burden of providing a detailed compendium of facts, derivations, and explanations is carried by a set of lecture notes; these largely substitute for the traditional textbook. Students are expected to read and refer to the lecture notes for background information and sample problems. Although we do review during class the concepts developed in the lecture notes, we do not find it productive to spend extensive amounts of time on that activity.

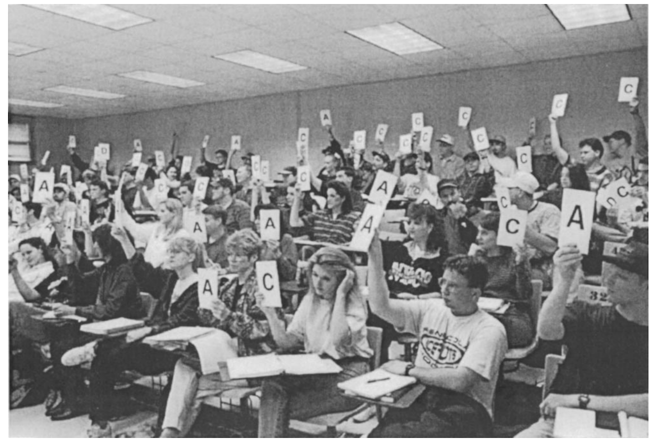


Fig. 1. Students signaling their response to instructor's question using flash cards.

IV. STUDENT RESPONSE SYSTEM

There are a number of student response systems available for use with interactive-lecture methods, including commercially available electronic systems.[70–72] Our method employs flash cards on which oversize letters of the alphabet are printed. Flash cards are less expensive and easier to implement, although they lack useful features of the electronic systems such as instant graphical displays of responses. We emphasize that almost everything we discuss in this paper may be implemented equally effectively with electronic response systems.

With the use of the flash-card system, we are able to ask many questions during class and no longer have to wait for one daring individual to respond. Every student in the class has a pack of six large cards ($5\frac{1}{2}$ in.$\times 8\frac{1}{2}$ in.), each printed with one of the letters A, B, C, D, E, or F. Students bring the cards every day, and extra sets are always available. During class we repeatedly present multiple-choice questions. Often, the questions stress qualitative concepts involving comparison of magnitudes, directions, or trends (for example, ''Will it decrease, remain the same, or increase?''). These questions are difficult to answer by plugging numbers into an equation. We give the students time to consider their response, 15 s to 1 min depending on the difficulty. Then we ask them to signal their response by holding up one of the cards, everybody at once (see Fig. 1). We can easily see all the cards from the front of the room. Immediately, we can tell whether most of the students have the answer we were seeking—or if, instead, there is a ''split vote,'' that is, part of the class with one answer, part with another—or perhaps more than one other. (One of them, it is hoped, is the right answer!)

One of the advantages of this system is that it allows the instructor to observe the students' body language. We can see whether the students held up their cards quickly, with confidence, or if instead they brought them up slowly, with confused looks on their faces. Do a large number of students delay their response, finally holding up an upside-down F? This is our signal for ''I don't really know the answer, and I can't even give a very good guess.'' It is not particularly easy for students to see each others' cards and so there is a fair degree of anonymity in their responses. Students' comfort in signaling answers with the cards seems to increase as the course progresses.

V. FORMAT OF THE FULLY INTERACTIVE LECTURE CLASS

A. Overview

Although there is considerable flexibility in the actual format of a fully interactive lecture class, it is possible to describe a characteristic pattern. The actual length and sequencing of the individual phases will vary depending on the activities of the previous class and those planned for the succeeding days. A typical class proceeds in three phases.

(1) A brief introduction/review of the basic concepts is presented at the blackboard, a sort of mini-lecture usually lasting around 3–7 min.

(2) A sequence of multiple-choice questions is posed to the class. These emphasize qualitative reasoning, proceeding from relatively simple to more challenging, and are closely linked to each other to explore just one or two concepts from a multitude of perspectives, using a variety of representations.[73] Students provide responses by using the flash cards. *Every opportunity is taken to interrupt the sequence of multiple-choice questions with brief free-response exercises, for example, drawing simple diagrams or performing elementary calculations*.[74]

(3) Follow-up activities are carried out. These vary and may consist of interactive demonstrations, group work using free-response worksheets, or another mini-lecture and question sequence.

At ISU, in addition to the class meetings (3 h/week) in the lecture hall, we make use of a once-per-week 50-min recitation session, which has been converted into a full-fledged tutorial in the style developed at the University of Washington.[10,60] Students spend the entire session working in small groups on carefully structured printed worksheets, guided by Socratic questioning from the instructors. Worksheets used in these tutorials have been designed by us and also form part of the *Workbook for Introductory Physics*.[75] At ISU we also have been able to make use of four of the weekly, 2-h laboratory periods to do additional active-learning instruction. In these we use *Tutorials for Introductory Physics*[61] and materials from the text *Electric and Magnetic Interactions*.[52]

B. Mini-lecture

The instructor begins by taking a few minutes to outline the principles and concepts underlying that day's activities. One or two key ideas are sketched, along with relevant diagrams and mathematical formulations. A demonstration might be shown (soliciting students' predictions of the outcome) and an example problem solved at the board. From then on the ball moves to the students' court.

C. Interactive-question sequence

The instructor proceeds to ask a series of questions to which the students all respond. We might use questions printed in the *Workbook* (which students always bring to class[76]) or present questions on the board or with an overhead transparency. The sequence starts with *easy* questions, in order to build confidence. Students consider the question on their own, taking perhaps 15–30 s. At a certain moment, all are asked to give their responses simultaneously. Because the first few questions are simple, the responses should be overwhelmingly correct. Gradually, the questions become more challenging. The instructor takes any available opportunity to interject a question requiring a ''free response,'' such as a simple sketch. As the students work on the free-response questions, the instructor circulates around the room and observes their work.

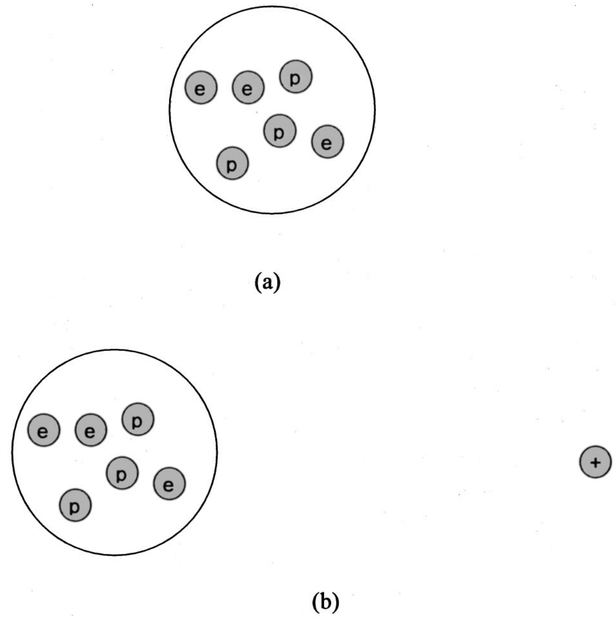


Fig. 2. Diagrams used in interactive-question sequence: (a) initial diagram, requiring flash-card response; (b) follow-up diagram, requiring free response by students writing at their desks.

As an example, the diagram in Fig. 2(a) was presented to the class (*e* represents an electron, *p* a proton). Students were first asked about the net electrical charge on the object represented by the circle; is it (A) greater than zero, (B) equal to zero, or (C) less than zero. Most students quickly responded with the correct answer, B. The instructor then drew in a nearby positive charge [Fig. 2(b)], inviting students to consider the nature of the interaction between the circled object and the positive charge (assuming the electrons and protons are fixed in position). He asked the students to sketch a set of arrows representing all electrical forces acting on that positive charge due to each of the protons and electrons. As the students worked at their desks, the instructor walked around the room, and quickly assessed how well the students were handling the assignment; he stepped to the board for a few moments to offer some hints. This entire process took less than 1 min. The instructor then asked the students whether the net interaction force implied by the collection of force vectors they had drawn was (A) toward the right, (B) toward the left, or (C) approximately equal to zero.

As an example of a more extended sequence, consider the series of electric field questions in Fig. 3. Question 1 is fairly easy; a large majority of students gave the correct answer (B) without needing to discuss it with their neighbors. When we came to question 2, however, we found that students were split in their choices; in addition to the correct answer (B), a significant fraction of the class held up the A card. When we came to question 3, the class response was very split; each of the options received some support. (Later, question 4 was given as a follow-up question in a different context.)

At some point, there is likely to be a significant split in opinion reflected in the students' responses. Perhaps 50%–70% give one answer (for example, A), while the remainder give a different answer (let's say, C). The instructor informs

1. An electron is placed at the origin, and then released and allowed to move freely. It begins to move along the positive x axis. From this one can conclude that the electric field at the origin points:
   A. toward the positive x direction.
   B. toward the negative x direction.
   C. toward the positive y direction.
   D. toward the negative y direction.
   E. in a direction not along either x or y axes.
   F. There is not enough information to determine the direction of the electric field at the origin.

2. In a certain region of space, the electric field is zero everywhere. This means that, if a charged particle is located anywhere in this region:
   A. the particle can not be moving.
   B. the particle experiences no net electrical force.
   C. the particle experiences a repulsive electrical force.
   D. the particle experiences an attractive electrical force.
   E. the particle is always forced back toward one particular location.

3. Throughout a certain region of space, the electric field has uniform magnitude and direction. This means that wherever a particular charged particle is placed at rest in this region and then allowed to move freely, it will always:
   i. remain motionless.
   ii. move with constant speed.
   iii. move in an unchanging direction.
   iv. move with constant magnitude of acceleration.
   A. i only
   B. ii only
   C. ii and iii only
   D. iii and iv only

4. Charge A is 2C and charge B is –4C. They are sitting in a uniform electric field. Which of these diagrams correctly shows the forces that are exerted on charges A and B ***by the electric field*** (<u>*not*</u> by each other)?

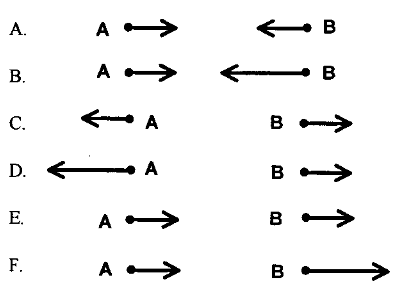


Fig. 3. Excerpt from interactive-question sequence.

the class of the difference of opinion: ''We have A's and C's, perhaps a few more A's. Why don't you take a few seconds to discuss it with each other?'' The students are expected to discuss the question with whoever is at a convenient distance. Almost always, an animated class-wide discussion ensues; nearly all students are actively engaged in comparing their answers, arguing for their point of view, and listening critically to their neighbors' reasoning. The instructor does not rush to press for an answer. A minute or more might elapse before a decreased intensity of discussion is noticed. Perhaps the instructor gives a warning, ''another 30 seconds.'' At a certain point, all students are asked to give their response. Often, the students will have reached a consensus: nearly everyone now has the same answer. Sometimes, however, the split in opinion persists; that is a signal that more discussion—with some additional exercises and questions—is probably needed.

If student opinion remains divided and a split vote persists despite the student discussion, we will often ask for an A supporter to present his/her argument, followed by a proponent of the C viewpoint. If necessary, we will eventually step in to alleviate the confusion. By this time, most of the students will have carefully thought through the problem. If they haven't already figured it out by themselves, they will now at least be in an excellent position to make sense out of any argument we offer to them. Before those minutes of hard thinking, we could have made the same argument and watched as almost every student in the class gave the wrong answer to some simple question. We know this to be true because we have tried it often enough.

One of the results of using interactive lectures is that the instructor begins to acquire startling new insights into what the students are *really* getting out of a typical lecture. One can present a straightforward concept (from the instructor point of view) and a simple example, and then—instead of proceeding rapidly to the next topic as would a traditional lecturer—present a short set of questions for the students to answer. One often discovers that the students are deeply mired in confusion. This is precisely what might occur in the office setting when, in the course of leading the student(s) through a series of questions, the instructor uncovers an unexpected and serious conceptual confusion. A tactical retreat is usually necessary, backtracking to simpler concepts that are more firmly understood by the student; one can then lead once again from the new starting point. This process takes some time but is necessary, because the student could not hope to master the new idea without consolidating his or her understanding of the foundation concepts.

This process is exactly what may be replicated through a fully interactive lecture. By using a properly thought out sequence of questions (often developed on the fly without having been scripted in advance) along with the student response system, the instructor is able to identify an area of conceptual confusion. Recognizing the need to retreat, the instructor offers another question that refers back to concepts previously discussed. One may then probe to locate a region of relatively firm understanding that can serve as a new launching point toward the original target.

As we work our way through a series of intermediate questions, at each step, we get a reading on our class: Do they respond quickly? With confidence? Mostly correctly? Then we comment briefly and move forward. Otherwise, we pause for a longer discussion. Instead of disposing of the entire topic in less than 2 min of traditional lecture, we now might take 10–15 min, struggling together with our students as they work their way through a conceptual minefield.

D. Follow-up activities

The sequence of interactive questions may be followed by another such sequence, perhaps preceded by a new mini-lecture. Mini-lectures may also be judiciously sprinkled into a class at various moments, allowing an opportunity for motivational or philosophical comments, or simply to provide a break from problem solving. We also expend considerable amounts of time on student group work using printed worksheets, included as an integral component of the *Workbook*; an excerpt is in Appendix A. Another method that we have

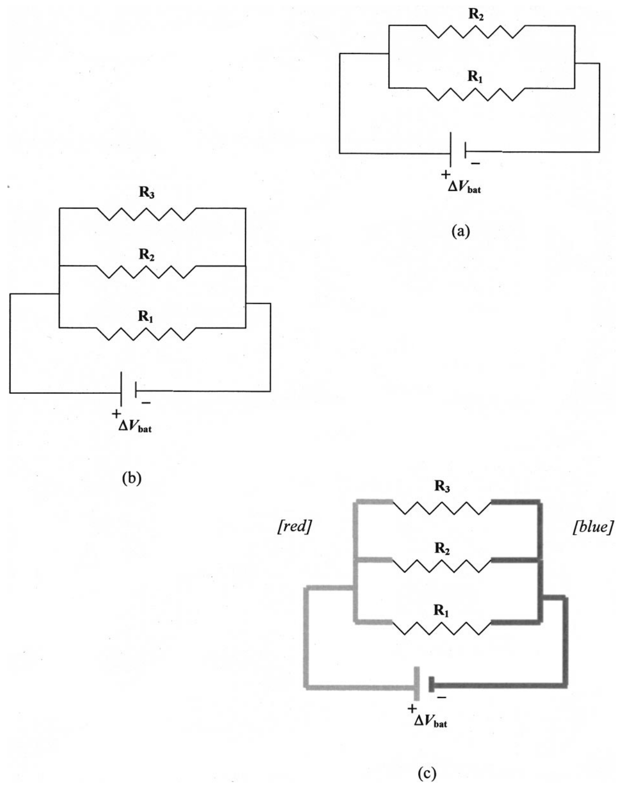


Fig. 4. Diagrams used in sample interactive-question sequence (Sec. VI): (a) initial diagram; (b) follow-up diagram; (c) diagram representing mnemonic for circuit potential map (instructor statement 9): left-side conducting segments (light shade; referred to as "red" by instructor) are at potential of positive battery terminal, right-side conducting segments (dark shade; referred to as "blue") are at potential of negative battery terminal.

used with great success is to convert the standard physics lecture demonstration into a fully interactive sequence.[77]

Our worksheets designed for use in large-enrollment classes focus on qualitative questions or problems that require only elementary algebraic calculations. Responses required from students include simple sketches, diagrams, graphs, and elementary numerical or algebraic expressions. Such responses may be easily and rapidly scanned and evaluated by an instructor who walks through the room.[78] By quickly sampling a significant fraction of the class, the instructor is able to recognize common difficulties and offer appropriate hints or other guidance.

VI. SAMPLE INTERACTIVE-QUESTION SEQUENCE

The instructional sequence that follows below occurred during the first half of an actual class. After having already studied series and parallel circuits, as well as electrical power, the students had started a new worksheet in the tutorial session on the previous day. The teaching assistant had reported substantial confusion, and so the instructor began class this day by posing a question (Instructor Statement 1) regarding battery power in a parallel circuit.

The instructor asks students to consider the two-resistor parallel circuit shown in Fig. 4(a), and then proceeds to ask a sequence of questions as follows.

(1) *Instructor*: Suppose an additional resistor is added in parallel to the circuit shown [in Fig. 4(a)], and so we get the circuit shown [in Fig. 4(b)]. Will the power produced by the battery (A) increase, (B) decrease, or (C) remain the same? [The instructor writes the question and the three response options on the board, and follows the same procedure with all questions cited in this segment.]

Students' responses are split approximately equally among the three options.

(2) *Instructor*: Will the current through the battery (A) increase, (B) decrease, or (C) remain the same?

Student responses are split approximately equally between (A) increase, and (B) decrease.

(3) *Instructor*: Okay, how about this: is ΔV_{R_3} (A) greater than, (B) less than, or (C) equal to ΔV_{R_2}? Note: ΔV_{R_3} represents the absolute value of the potential drop across resistor R_3, etc.

Students are slow to show their flash cards; responses are still very split among the options.

(4) *Instructor*: Okay, let's go back to the two-resistor circuit [Fig. 4(a)]. Is ΔV_{R_2} (A) greater than, (B) less than, or (C) equal to ΔV_{R_1}?

Student question: Is $R_2 = R_1$?

(5) *Instructor*: Let's assume they are.

The large majority of students correctly answer (C).

(6) *Instructor*: Okay, now assume that $R_2 > R_1$; what will be the answer in that case?

Again, the large majority of students correctly answer (C).

(7) *Instructor*: What happens to I_1 if we increase R_2, will it (A) increase, (B) decrease, or (C) remain the same? I_1 represents the current through resistor R_1, etc.

The large majority of students correctly answer (C).

(8) *Instructor*: All right, now let's go back to the three-resistor case. Is ΔV_{R_3} (A) greater than, (B) less than, or (C) equal to ΔV_{R_2}?

Flash cards are slow coming up, responses are mixed.

(9) *Instructor*: All right, here's a hint. [Instructor uses red chalk to highlight all conducting segments connected directly to positive terminal of battery, and uses blue chalk to highlight all segments connected to negative terminal [Fig. 4(c)]; this mnemonic had been introduced in previous classes to emphasize that the potential difference between any point in the red region and any point in the blue region was equal to the potential difference between the battery terminals, that is, that $V_{\mathrm{red}} - V_{\mathrm{blue}} = \Delta V_{\mathrm{bat}}$.]

Now, the large majority of students hold up the correct answer (C).

(10) *Instructor*: And how about compared to ΔV_{R_1}, is ΔV_{R_3}(A) greater than, (B) less than, or (C) equal to ΔV_{R_1}?

The large majority of students again hold up correct answer (C).

(11) *Instructor*: Okay.

Student question: So what changes? Doesn't something change?

(12) *Instructor*: Yes, but *not* ΔV. Okay, let's assume that all three resistors are equal, $R_1 = R_2 = R_3$, and let me ask you about the current. Is I_3 (A) greater than, (B) less than, or (C) equal to I_2?

Nearly all students correctly answer (C).

(13) *Instructor*: And is I_3 (A) greater than, (B) less than, or (C) equal to I_1?

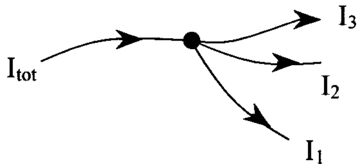


Fig. 5. Diagram used in sample interactive-question sequence (Sec. VI), drawn on board after instructor statement 16.

Again, nearly all students correctly answer (C).

(14) *Instructor*: Okay, now if we start with that initial two-resistor circuit and add the third resistor, will the total current through the battery (A) increase, (B) decrease, or (C) remain the same?

Student response is approximately 50% for (A), 40% for (B), and 10% for (C).

(15) *Instructor*: Okay, we still have a split vote. Will somebody explain why they think the answer is (A)?

Student: It's (A) because the equivalent resistance of the circuit will decrease.

(16) *Instructor*: And now will somebody explain why they think the answer is (B)?

Nobody volunteers to defend answer (B). Instructor now draws on board diagram shown in Fig. 5.

(17) *Instructor*: Okay, once again: If we start with that initial two-resistor circuit and add the third resistor, will the total current through the battery (A) increase, (B) decrease, or (C) remain the same?

Now there is a much larger proportion of correct (A) responses.

(18) *Instructor*: (A) is correct. I guess that still seems weird.

Several students agree out loud that it does seem weird. Instructor reminds students that they have observed and discussed experiments in the laboratory that are consistent with this conclusion.

Student question: How far can the battery go and still keep putting out more current?

(19) *Instructor*: I don't know. It basically depends on the equipment you're using.

Student: But aren't you increasing the equivalent resistance, since $R_{\mathrm{equiv}}=R_1+R_2+R_3$?

(20) *Instructor*: Ah. No, that's only for series circuits. It's not true for parallel circuits. Okay, let's go back to our original question. If we add a resistor in parallel to the original two-resistor parallel circuit, will the power produced by the battery (A) increase, (B) decrease, or (C) remain the same?

A full two minutes elapse before the students are asked for a response. The large majority of students correctly answer (A).

(21) *Instructor*: Okay, (A) is correct.

VII. DISCUSSION OF SAMPLE INTERACTIVE-QUESTION SEQUENCE

The sequence in Sec. VI is a representative example of how closely a fully interactive lecture may resemble a one-on-one tutorial session, and how little it resembles a traditional lecture. The role of the instructor is essentially that of asking questions, providing hints, and guiding discussion. The instructor also confirms answers on which the class has achieved consensus. Here we discuss key elements of the fully interactive lecture exemplified by this sequence.

A. *The frequency of questioning may be as high as several per minute*. During this relatively brief sequence, which took only approximately 20 min, the students were asked to use their flash cards to respond to 13 separate questions. During portions of the segment, there were two or three (easy) questions in a single minute. This rate is similar to the rhythm of one-on-one tutoring, in which there is often a rapid exchange of questions and answers between students and instructor.

B. *The instructor must often create unscripted questions on the spot*. All of the questions were improvised by the instructor without previous scripting or preparation. In just the way that an instructor must come up with appropriate extemporaneous questions when doing one-on-one teaching, an instructor in a fully interactive lecture must be prepared to respond to the flow of the large-class discussion. It is important to write both the question and the answer options on the board so students may refer back to them. However, it may be useful to delay writing the answer options for a few moments to first give students time to consider their own response.

C. *Easy questions are used to maintain the flow of the discussion*. Many of the questions are easy for the students to answer, and they receive overwhelmingly correct responses. Crouch and Mazur[35] note that questions with correct-response rates over 70% tend to produce less useful discussions than do more difficult questions. However, we find that they build student confidence and are important signals to the instructor of students' current knowledge baseline. Often enough, questions thought by the instructor to be simple turn out not to be, requiring some backtracking. Because of that inherent degree of unpredictability, some proportion of the questions asked will turn out to be quite easy for the students. This small conceptual "step-size" allows more precise fine tuning of the class discussion.

D. *Virtually any system offers a rich array of possible question variants*. Almost any physics problem may be turned into an appropriate conceptual question. By using the basic question paradigms "increase, decrease, remain the same," "greater than, less than, equal to," and "left, right, up, down, in, out," along with obvious variations, it is possible to rapidly create many questions that probe students' qualitative thinking about the system. By introducing minor alterations in a physical system (adding a force, increasing a resistance, etc.), students can be guided to apply their conceptual understanding in a variety of contexts. In this way, the instructor is able to provide a vivid model of the mental approach needed for active learning.

E. *The instructor must be prepared to approach a given problem with a variety of possible questioning strategies*. It often is found that students do not respond in an expected manner, and that their knowledge base for a particular problem is shakier than anticipated. Just as in one-on-one tutoring, the instructor must be ready to pose easier questions set in less complex physical settings, and to offer appropriate hints to guide the students toward the target concept. By remaining observant of students' rapidity in offering responses, body language in showing the flash cards, and expressions on their faces, the instructor should be able to judge which questions might require additional response time.

VIII. STUDENT WORKBOOK

A. Elements of the Workbook

As our experience in implementing these methods has evolved, we have found it increasingly necessary to abandon traditional curricular materials and to develop our own in order to support the instructional techniques. The first need was for a large stock of appropriate multiple-choice questions to be used in the fully interactive lectures. Despite the excellent set of ConcepTests provided in Mazur's book, our methods required many more questions covering a wider range of difficulty levels than were available in Mazur's book or in other sources. The materials we eventually developed for the second semester of the algebra-based general physics course now form the *Workbook for Introductory Physics*.[79]

Our early attempts to rely on standard textbooks as a course reference eventually foundered due to the sharp clash between the heavily mathematical approach of such texts, and our strong focus on qualitative and conceptual problems. This clash led to abandonment of a standard text for use in our second-semester course, and the creation of a set of lecture notes as a substitute. These notes, now included as an integral component of the *Workbook*, emphasize qualitative reasoning, make heavy use of sketches and diagrams, and—though treating fewer topics than standard texts—go into far greater depth on those key concepts chosen for emphasis in our course.

Another key element that was found to be necessary for our *Workbook* was the creation of numerous free-response worksheets (see, for example, Appendix A). The worksheets emphasize qualitative questions, often require explanations of reasoning, and target learning difficulties that have been identified in the research literature as well as those familiar to us from our own experience. In addition to in-class use, the worksheets also serve as a primary source of homework exercises. Although superb worksheets based on extensive research are available in the *Tutorials for Introductory Physics*,[61] there was simply not enough to satisfy our need for every-day use in the algebra-based course, covering the full range of topics in that course and appropriate for students even with very low levels of preparation. (Other sources of worksheets of a somewhat different type are now also available.[29,30,80])

A final element now included in the *Workbook* is a large collection of quizzes and exams (and solution sets for the exams) that have been given in previous years. These form an invaluable source of additional flash-card questions, free-response exercises, and material for homework assignments and student review. They also respond directly to incessant student demands for samples of previous exams for exam preparation and review.

B. Nature of the curricular materials

The materials are designed based on the assumption that the solution of even very simple physics problems invariably hinges on a lengthy chain of concepts and reasoning. The question sequences guide the student to lay bare these chains of reasoning, and to construct in-depth understanding of physical concepts by step-by-step confrontation with conceptual sticking points. Carefully linked sequences of activities first lead the student to confront the conceptual difficulties, and then to resolve them. This strategy was developed at the University of Washington.[8–10] Complex physical problems are broken down into conceptual elements, allowing students to grapple with each one in turn and then return to synthesize a unifying perspective.

Over several years the flash-card questions, worksheets, and quiz and exam problems have undergone a continuous (and unending) process of testing and revision in actual classroom situations. Constant in-class use discloses ambiguous and confusing wording which is then rapidly corrected in new printings of the materials—sometimes the same day, for use in a later tutorial session. Analysis of assessment data provides additional guidance for revisions.

IX. IMPLEMENTATION ISSUES

A. Constraints on topical coverage

The single greatest concern for most instructors who are considering implementing interactive-lecture methods is that of coverage: can one cover the same amount of material as in a traditional course? The short answer is no. That is, the instructor will not be able to present, at the board, the same amount of material as in a standard course, and there will not be enough time during class to discuss the usual wide variety of topics. It is helpful to be very clear about this fundamental reality.

However, that short answer only scratches the surface of the issue. For one thing, there is extensive evidence that although instructors in introductory physics courses might cover many topics, the majority of students do not gain any significant degree of mastery over most of the material. Assessment data from our courses and from many others show convincingly that student learning of basic concepts is improved with interactive-engagement methods. Moreover, as much as we might wish to give a clear-cut answer to the question of coverage, there really does not exist an answer that is both accurate and general. The amount of material that can be covered is critically dependent on the student population. We found, for instance, that an amount of material requiring virtually the entirety of a fifteen-week semester at one institution could be effectively covered before the midterm date at a different institution. There, the better-prepared students were able to master the concepts more quickly.

The best response to this question is that instructors are free to cover as many topics as they wish. The real issue is *depth* of coverage. We choose certain concepts from each topic—the big ideas in our view—and focus in-depth class discussion on those concepts. We are content to discuss only briefly, if at all, other concepts contained within the same topical area. For instance, we cover dc circuits, but not ac circuits or multiloop circuits requiring analysis with simultaneous linear equations. We cover interference, but not diffraction, the optics of lenses, but not of mirrors or optical instruments. We omit topics such as special relativity, particle physics, and astrophysics. (On a time-per-topic basis, our second-semester course spends approximately 75% on electricity and magnetism and about 25% on optics and modern physics.) If it is necessary for some reason to cover certain topics, there is nothing to prevent an instructor from devoting a few traditional lectures to those subjects; that will ensure rapid coverage indeed!

B. Consistency of implementation

In a traditional lecture class the initiative lies entirely with the instructor; the student is free to relax, listen, and passively observe the instructor's board work. In the fully interactive lecture the student is continually being forced to think hard about difficult concepts, commit to decisions about problem solutions, and interact with classmates to discuss challenging questions. At the very least this interaction requires a significant investment of thought and energy in a course most students take merely to satisfy a requirement. Many students who find themselves in this situation do *not* automatically welcome the opportunity to engage in a learning experience that is far more intensive than normal.[81,82]

Largely for these reasons, we and others have found that it is critical to the success of these methods that they be implemented consistently throughout the course, beginning with the very first day. For example, our students pick up their sets of flash cards as they walk in the first day of class, and the first set of flash-card questions begins within the first minute of class. (These are questions such as "Did you take high-school physics?," etc.) We explain that these methods have been repeatedly demonstrated to yield positive results, and reassure students that the impact on grades is usually found to be favorable. Virtually every class period includes interactive questions and related activities. Instructors need to be aware that attempting to introduce these or other forms of active-learning methods mid-semester, after students have already settled down into the routine of a traditional lecture course, could be disastrous for student (and instructor) morale.

C. Grade-related assessment

As has been pointed out by many educators, it is absolutely *essential* to the success of any instructional method that students be examined and graded in a manner consistent with the form of instruction. In our second-semester course, we give a written in-class quiz twice per week; the majority of questions are very similar to the flash-card questions. Indeed, actual past quiz questions are frequently used as part of the flash-card question sequences. Exams also focus heavily on qualitative questions, and on problems that involve little algebra but require good conceptual knowledge and proportional reasoning skills. Some problems require explanations of students' reasoning. To help promote a cooperative atmosphere among the students, an absolute grading scale is used so that any student accumulating a preset point total is guaranteed in advance, at the minimum, a certain corresponding letter grade.

D. Student attitudes

We and others[35] have found that during the first few weeks many students are unsettled and uncomfortable with interactive lecture classes. It takes time for them to become accustomed to the new routine and to appreciate its benefits. We find that by the end of the course, most students have positive attitudes. End-of-course surveys show that most students react favorably to the instructional methods, with approximately 30%–40% giving maximum ratings on evaluations. (Sample comment: "... best physics instructor I have ever had.... He makes physics fun and interesting to learn... .") Most of the remainder are positive or neutral, but there is often a core of less than 10% that *despises* these methods. (Sample comment from the same class: "... has a new way of teaching he is trying to develop. It doesn't work...going to lecture was pointless other than to take required quizzes.") During the Fall 2000 semester at ISU the number of responses in this most unsatisfied category dropped to zero but it appears, unfortunately, that that was only an anomaly.

E. Demands on the instructor

Teaching a physics course using fully interactive lectures is not an easy task; it requires much energy and commitment. The instructor needs to come to class with a clear plan—tentative though it may be—for that day's intended sequence of questions and activities. Pre-scripted questions must be selected, and additional questions must be prepared as needed. During class the instructor must be attentive to student reactions, willing to walk around the room and check on student work, and prepared to shift gears and redirect discussion on short notice. (When we find ourselves lecturing for more than ten minutes at a time, it indicates that we have not prepared adequately for that day's class.)

X. ASSESSMENT OF STUDENT LEARNING

We first note a remarkable effect that we have consistently observed, that is, a very small number of dropouts, typically 1%–3% after the first week. Attendance is $\approx$90% on virtually every class day (no doubt largely due to frequent graded quizzes). We should also acknowledge that, although we believe that the techniques would scale well with larger classes, we have not personally tested these methods in classes with over 100 students.

The *Workbook* has been used for the past five years at SLU and ISU and has undergone continuous development. The course at SLU consisted only of the interactive lectures, while that at ISU has the very substantial additional element of a weekly tutorial session. There are still other important elements of the ISU course that certainly contribute to the learning gains, including the four active-engagement laboratory sessions. (We have no way of apportioning learning gain contributions among the various course elements.) Our full implementation model has been used only for the second-semester algebra-based course, and data from that course are reported here.

We discuss the results of the Conceptual Survey in Electricity (CSE), the Conceptual Survey in Electricity and Magnetism (CSEM),[7] electric circuit concept questions, and quantitative problem solving. Since 1997, an abridged version of the CSE has been administered on both the first and last days of class. The CSE is a 33-item multiple-choice test that surveys knowledge related to electrical fields and forces. About half of the items are identical (or nearly so) to questions included on the CSEM. The items on the CSE and CSEM are almost entirely qualitative and probe knowledge both of physics concepts and aspects of related formalism.[7] On the pre-test, students answered all questions, but on the post-test they were instructed to respond only to a 23-item subset.[83] We refer to this subset of the CSE as the Abridged CSE. Only the 23 designated items were graded, both on the pre- and post-test. Table I gives these scores for the five courses in which we administered the test; only students who took *both* tests are included (that is, data are "matched"). Despite the addition of tutorials, along with expansions and improvements in the curricular materials, we cannot conclusively state that the improvements in post-test scores and normalized gain[84] (that is, Hake's $\langle g \rangle$) observed at ISU can be entirely attributed to changes in instruction. ("Normalized gain" is defined as the actual pre-test to post-test increase in exam score, divided by the maximum possible increase.) As

Table I. Scores on the 23-item abridged Conceptual Survey in Electricity (CSE).

Sample	N	CSE mean pre-test score	CSE mean post-test score	$\langle g \rangle$[a]
SLU 1997 (lecture only)	58	29%	62%	0.46
SLU 1998 (lecture only)	50	27%	66%	0.53
ISU 1998 (lecture+tutorial)	70	34%	76%	0.64
ISU 1999 (lecture+tutorial)	87	30%	78%	0.69
ISU 2000 (lecture+tutorial)	66	34%	79%	0.69

[a]Calculated using exact (unrounded) pre-test and post-test scores.

one of us has shown (DEM) in a recent report, various other factors probably play a significant role in determining student performance as reflected in assessment data of this type.[85]

In all cases, our pre-test to post-test gains are quite high by most standard measures such as normalized gain $\langle g \rangle$ (0.46–0.69) and effect size. (''Effect size'' is the change in exam score divided by the standard deviation of the scores.) By way of comparison, it has been found in mechanics courses that typical values of normalized gain on the Force Concept Inventory are $\langle g \rangle \approx 0.25$ for traditional courses, and $0.35 \leqslant \langle g \rangle \leqslant 0.70$ for interactive engagement courses.[4,19] (The Force Concept Inventory is a very widely used mechanics diagnostic test.) For the three ISU samples, treating the ''pre-test'' and ''post-test'' populations as distinct, we find effect size $d > 3.0$, while values of $d \approx 0.8$ are ordinarily considered large.[86]

Although our post-test and $\langle g \rangle$ values are far higher than comparable values found in a national survey of CSE results,[7] it would not be proper to attempt a direct comparison between our abridged-CSE data and other data reflecting administration of the full CSE. Table II shows mean pre-test, post-test, and normalized learning gain values for a 14-item subset that consists of all questions included on both the abridged CSE and on the CSEM; only ISU data are available. Also shown are comparable values from the national survey data.[7] (Note that these latter data are not matched.) These data show that although ISU pre-test scores are very nearly equal to those in the algebra-based courses in the national sample, post-test scores and normalized learning gains are dramatically higher than both algebra-based and calculus-based courses in that sample, with mean normalized learning gains (mean $\langle g \rangle = 0.68$) triple those found in the national survey ($\langle g \rangle = 0.22$).[87] We note also that our students' scores on final-exam magnetism questions drawn from the CSEM—well above those of the national sample post-tests—are quite consistent with the data shown in Table II.

In Table III we present data on electric circuit questions that have been administered on our final exams for the past four years; these questions (Fig. 6) are drawn from the study of Shaffer and McDermott.[88] The authors report assessment data on these questions for several different courses, including both traditional courses and courses that used the electric circuit tutorials from *Tutorials in Introductory Physics*. Although we find significant year-to-year variations in the scores of students in our courses, all of our eight scores are higher than the comparable scores in traditional courses.[88]

Our course differs from most traditional courses in three key ways: (1) use of fully interactive lecture and highly interactive tutorials, (2) strong emphasis on conceptual problems, and (3) coverage of a smaller number of topics than most courses. Our data do not allow us to estimate the *relative* contribution of these three factors to the assessment results reported here. In relation to item (3), we note that Hake has concluded that the fraction of course time devoted to the study of mechanics topics is not significantly related to superior learning gains on the Force Concept Inventory reported for IE courses.[89] He also notes that only *partial* implementation of interactive methods—even when there may be some emphasis on conceptual problems—is correlated with poorer learning gains than those achieved in courses with full implementation of those methods.[90] However, a study by Greene suggests that improved learning gains may be possible even in a relatively traditional noninteractive course in which conceptual examples and problems are strongly emphasized on homework assignments and exams.[91]

An important issue for many students in the algebra-based physics course is preparation for pre-professional exams such as the Medical College Admissions Test (MCAT). The most recent versions of the MCAT put substantial emphasis both on qualitative physics questions and on the analysis of complex reading passages requiring application of fundamental physics concepts in unfamiliar contexts. Physics courses that emphasize conceptual understanding might well provide superior preparation for this type of exam. Careful studies of MCAT performance for students enrolled in such a course at the University of California at Davis provide support for this hypothesis.[92]

An important concern of many physics instructors is the

Table II. Scores of CSEM subset of 14 electricity questions.

Sample	N[a]	CSEM electricity subset mean pre-test score	CSEM electricity subset mean post-test score	$\langle g \rangle$[b]
National sample (algebra-based courses)	402	27%	43%	0.22
National sample (calculus-based courses)	1496	37%	51%	0.22
ISU 1998	70	30%	75%	0.64
ISU 1999	87	26%	79%	0.71
ISU 2000	66	29%	79%	0.70

[a]N for national sample is mean of values reported for each of the 14 individual questions, both pre and post; data from Ref. 7.
[b]Calculated using exact (unrounded) pre-test and post-test scores.

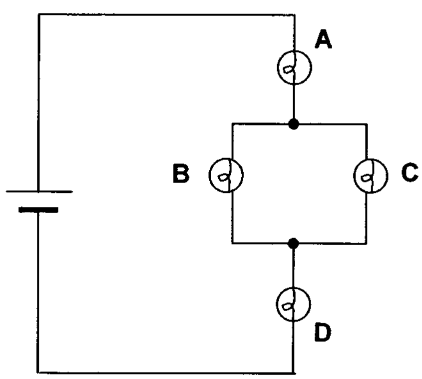
Table III. Post-instruction scores on circuit questions.

Sample	N	Four-bulb question [Fig. 6(a)]: correct with correct explanation	Five-bulb question [Fig. 6(b)]: correct with correct explanation
Traditional, algebra-based, university	[a]	$<50\%$	⋯
Traditional, calculus-based, university	[a]	$<50\%$	15%[b]
Tutorial, calculus-based, university	[a]	$>75\%$	45%[c]
Tutorial, calculus-based, college	[a]	65%	⋯
SLU 1998	61	54%	59%[d]
ISU 1998	76	75%	33%
ISU 1999	86	59%	31%
ISU 2000	79	86%	46%

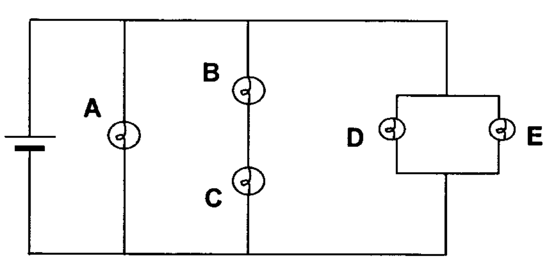
[a]Four-bulb question, four classes, total $N \approx 500$; Five-bulb question: see notes (b) and (c); data as reported in Ref. 88.
[b]$N \approx 50$; administered in subsequent course.
[c]$N \approx 50$; administered in subsequent course.
[d]Explanation not required.

extent to which a course's focus on conceptual questions may detract from students' ability to solve standard quantitative problems. (We stress, though, that our course's emphasis on qualitative problems is accompanied by extensive practice with some fairly standard quantitative problems, albeit ones requiring only a modest degree of algebraic manipulation; see Appendix B.[93]) We have attempted to address this concern by including on our final exam problems drawn directly from the traditional calculus-based introductory physics course at ISU (omitting problems using calculus). In 1998 we used six questions copied directly from two different final exams in the calculus-based course; in 1999 and 2000 we included three of those same six questions. (All six are shown in Appendix C.) The data in Table IV show that students in our algebra-based course outperformed the students in the calculus-based course on those questions; they also show that results on the three-item subset were virtually identical to those on the full six-item set.

Our results are consistent with those of others who have implemented research-based instructional methods. That is, students' ability to solve quantitative problems is maintained or even slightly improved. At the same time, at the cost of a modest restriction of topical coverage, students are able to meet substantially more rigorous standards on qualitative problem solving.[94]



(a)



(b)

Fig. 6. Questions used to assess understanding of circuits (from Ref. 88). All bulbs are identical, and all batteries are ideal. Students are asked to rank relative brightness of bulbs, and to explain their reasoning; (a) Answer: A $=$ D $>$ B $=$ C; (b) Answer: A $=$ D $=$ E $>$ B $=$ C.

Table IV. Scores on quantitative problems, ISU courses.

Sample	N	Mean score
Traditional calculus-based course, 1997 and 1998 six final exam questions	320	56%
Interactive-lecture course, 1998 (algebra-based) six final exam questions	76	77%
Traditional calculus-based course, 1997 and 1998 three-question subset	372	59%
Interactive-lecture course, 1998, 1999, 2000 (algebra-based) three-question subset	241	78%

XI. CONCLUSION

Our objective is to transform the large-enrollment lecture classroom, as much as possible, to one that is more typical of small-group instruction. We try to achieve this objective by obtaining simultaneous responses from all students to carefully designed sequences of questions emphasizing qualitative reasoning. The students' responses allow us to modify the pacing and direction of further class discussion and questioning. Curricular materials designed to facilitate this instructional method have been developed, tested, and assembled into a student workbook. Assessment data regarding student learning show gains far higher than those reported in national surveys of comparable courses.

Our experience and those of others makes it clear that interactive lectures are now a practical and tested option, available for immediate use by physics instructors virtually anywhere. As with any other novel teaching method, there is a learning curve for both students and instructors, but most practitioners have found that a commitment to use the methods on an extended basis almost always results in at least some degree of success.

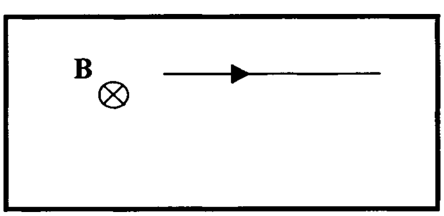
ACKNOWLEDGMENTS

Dan McCarthy edited an early version of the *Workbook* with painstaking care and offered many helpful suggestions for improvements. We are very grateful to the teaching assistants at Iowa State University: Michael Fitzpatrick, Tina Fanetti, Jack Dostal, Ngoc-Loan Nguyen, Agnès Kim, Sarah Orley, and David Oesper. Through close and dedicated work with students in their tutorial sessions, they were able to provide much valuable insight into student thinking and many important suggestions for improvements to the curricular materials.

APPENDIX A: EXCERPT FROM FREE-RESPONSE WORKSHEET

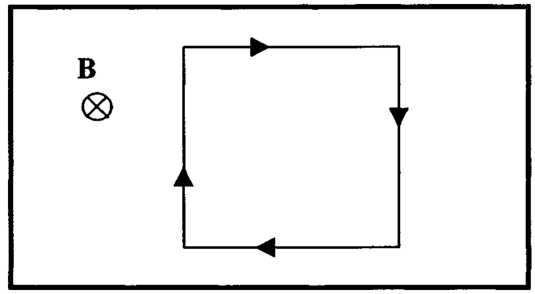
Torque on a Current Loop in a Magnetic Field

1. All *throughout* the boxed region below, there is a uniform magnetic field pointing *into* the page (as indicated by the cross). [This field is created by source currents outside of the region.] A wire segment carrying a current in the direction shown is placed inside the region. [Wires leading to the battery are not shown in this or in any subsequent figure.]

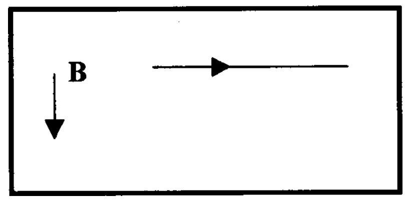


Indicate the direction of the force on the wire segment, using either arrows or the "dot" or "cross" symbols. If the force is zero, write "zero."

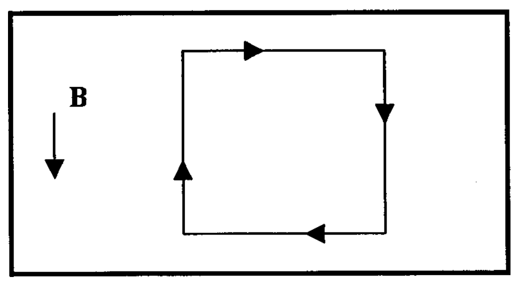
2. Now, a square wire loop carrying a steady clockwise current is placed in the region. (Current in each of the four sides is equal.) On each of the four sides of the loop, indicate the direction of the magnetic force (if there is one) or write "zero." Is there a *net force* acting on the loop as a whole? If so, state its direction. If not, explain how you can tell.



3. In this region, a uniform magnetic field is present that points toward the *bottom* of the page. A wire segment carrying a current in the direction shown is placed in the region. Indicate the direction of the force on the wire segment, using either arrows or the "dot" or "cross" symbols. If the force is zero, write "zero."



4. Now, a square wire loop carrying a clockwise current is placed in the region. On each of the four sides of the loop, indicate the direction of the magnetic force (if there is one.) Is there a *net force* acting on the loop as a whole? If so, state its direction. If not, explain how you can tell.



APPENDIX B: REPRESENTATIVE SAMPLE OF QUIZ AND EXAM PROBLEMS USED IN COURSE

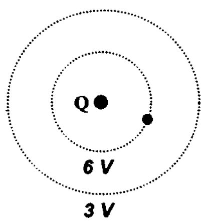
1. An electron is located at (0 m, +1 m) and two protons are located at (0 m, −2 m). A +2-C charge is located at the origin. What is the magnitude of the net electric *field* experienced by the charge at the origin, produced by the electron and the protons?

2. Current flows out of a battery and into resistor A (2 ohms). When the current flows out of resistor A it branches, with part of it going through resistor B (2 ohms) and the rest going through resistor C (4 ohms). The current then recombines and returns to the battery. If the voltage drop across resistor A is ΔV_A, what is the voltage drop across resistor C?

A. 1/3 ΔV_A
B. $\frac{1}{2}$ ΔV_A
C. 2/3 ΔV_A
D. $\frac{3}{4}$ ΔV_A
E. ΔV_A
F. 3/2 ΔV_A
G. 4/3 ΔV_A
H. 2 ΔV_A
I. 3 ΔV_A

3. A charge Q is fixed at the origin. An object with mass 3 kg and charge 9 C is held motionless on the 6-V equipotential circle (a distance r from the origin), and then released. (See diagram.) Which of these will be closest to the velocity attained by the object when it is *very far* (more than 1,000 r) from the origin?

A. 0 m/s
B. 2 m/s
C. 3 m/s
D. 4 m/s
E. 6 m/s
F. 36 m/s
G. 54 m/s

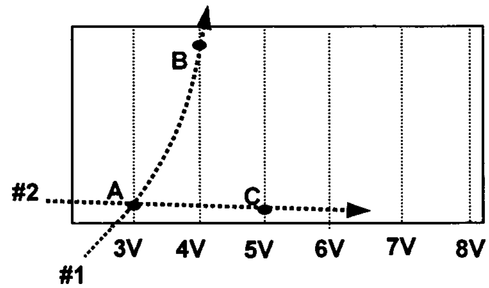


4. A 5-ohm and a 2-ohm resistor are connected in series to a battery. In a separate circuit, a 5-ohm and a 2-ohm resistor are connected in parallel to a battery with the *same* voltage. In which resistor is the *most power* being dissipated?

A. The 5-ohm resistor in the series circuit.
B. The 5-ohm resistor in the parallel circuit.
C. The 2-ohm resistor in the series circuit.
D. The 2-ohm resistor in the parallel circuit.
E. Both resistors in the series circuit, which dissipate the same amount of power.
F. Both resistors in the parallel circuit, which dissipate the same amount of power.
G. All four resistors dissipate the same amount of power.

5. A positive charge q is shot into a region in which there is a uniform electric field (see diagram). First, it is shot along path #1; then it is shot in again along path #2. *CHOOSE TWO CORRECT STATEMENTS (half credit for each).*

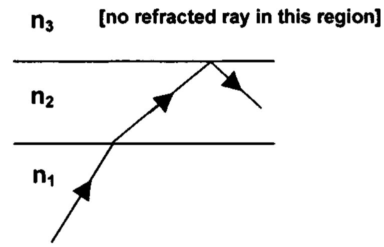
A. It gains kinetic energy while traveling inside this region.
B. It loses kinetic energy while traveling inside this region.
C. Its kinetic energy is constant while traveling inside this region.
D. The kinetic energy change from [A to B] is *greater than* the kinetic energy change from [A to C].
E. The kinetic energy change from [A to B] is *less than* the kinetic energy change from [A to C].
F. The kinetic energy change from [A to B] is *the same as* the kinetic energy change from [A to C].



6. The diagram shows part of the path traveled by a particular light ray as it strikes a piece of three-layer material. The different layers have different indices of refraction (n_1, n_2, and n_3) as indicated. Note that no ray is observed in the n_3 region.

What is the correct ranking (largest to smallest) of the three indices of refraction?

largest____________________smallest



APPENDIX C: QUESTIONS FROM ISU CALCULUS-BASED PHYSICS EXAM

Questions 1–6 were given on the 1998 final exam in the interactive-lecture course. Questions 1, 2, and 4 were also given on the 1999 and 2000 final exams in that course; the format of question 4 was slightly modified to increase its difficulty.

1. Two point charges $+7.00\times10^{-9}$ C and $+9.00\times10^{-9}$ C are located 4.00 m apart. The electric field intensity (in N/C) halfway between them is:

A. 0 B. 1.1 C. 4.5 D. 9 E. 36

2. Two particles, X and Y, are 4 m apart. X has a charge of 2Q and Y has a charge of Q. A third charged particle Z is placed midway between X and Y. The ratio of the magnitude of the electrostatic force on Z from X to that on Z from Y (F_{zx}: F_{zy}) is:

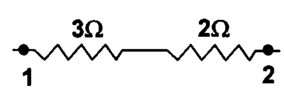
A. 4:1 B. 2:1 C. 1:1 D. 1:2 E. 1:4

3. An unknown resistor dissipates 0.50 W when connected to a 3.0 V potential difference. When connected to a 1.0 V potential difference, this resistor will dissipate:

A. 0.50 W B. 0.17 W C. 1.5 W D. 0.056 W
E. None of these.

4. In the diagram, the current in the 3.0-Ω resistor is 4.0 A. The potential difference between points 1 and 2 is:

A. 0.75 V B. 0.8 V C. 1.25 V D. 12 V E. 20 V

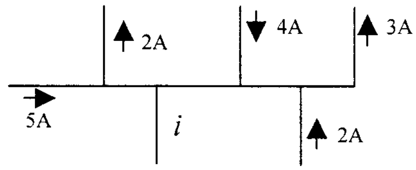


5. The electric field at a distance of 10 cm from an isolated point charge of 2×10^{-9} C is:

A. 0.18 N/C B. 1.8 N/C C. 18 N/C D. 180 N/C
E. None of these

6. A portion of a circuit is shown, with the values of the currents given for some branches. What is the direction and value of the current i?

A. ↓, 6 A B. ↑, 6 A C. ↓, 4 A D. ↑, 4 A E. ↓, 2 A



[a]Electronic mail: dem@iastate.edu

[1]Eric Mazur, *Peer Instruction: A User's Manual* (Prentice–Hall, Upper Saddle River, NJ, 1997).

[2]Ibrahim Abou Halloun and David Hestenes, ''The initial knowledge state of college physics students,'' Am. J. Phys. **53**, 1043–1055 (1985).

[3]David Hestenes, Malcolm Wells, and Gregg Swackhamer, ''Force concept inventory,'' Phys. Teach. **30**, 141–158 (1992).

[4]Richard R. Hake, ''Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,'' Am. J. Phys. **66**, 64–74 (1998).

[5]Edward F. Redish, Jeffery M. Saul, and Richard N. Steinberg, ''On the effectiveness of active-engagement microcomputer-based laboratories,'' Am. J. Phys. **65**, 45–54 (1997).

[6]Ronald K. Thornton and David R. Sokoloff, ''Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula,'' Am. J. Phys. **66**, 338–352 (1998).

[7]David P. Maloney, Thomas L. O'Kuma, Curtis J. Hieggelke, and Alan Van Heuvelen, ''Surveying students' conceptual knowledge of electricity and magnetism,'' Phys. Educ. Res., Am. J. Phys. Suppl. **69**, S12–S23 (2001).

[8]Lillian C. McDermott, ''Millikan Lecture 1990: What we teach and what is learned—closing the gap,'' Am. J. Phys. **59**, 301–315 (1991).

[9]Lillian C. McDermott, ''Guest comment: How we teach and how students learn—a mismatch?,'' Am. J. Phys. **61**, 295–298 (1993).

[10]Lillian C. McDermott, ''Bridging the gap between teaching and learning: The role of research,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 139–165 (1997)], pt. 1.

[11]E. F. Redish, ''Implications of cognitive studies for teaching physics,'' Am. J. Phys. **62**, 796–803 (1994).

[12]Arnold B. Arons, *Teaching Introductory Physics* (Wiley, New York, 1997).

[13]David P. Maloney, ''Research on problem solving: Physics,'' in *Handbook of Research on Science Teaching and Learning*, edited by Dorothy L. Gabel (Macmillan, New York, 1994), pp. 327–354.

[14]Frederick Reif, ''Millikan Lecture 1994: Understanding and teaching important scientific thought processes,'' Am. J. Phys. **63**, 17–32 (1995).

[15]It is widely acknowledged among physics educators that, one way or the other—whether through homework, class work, or individual discussion—most students must be guided to exert intense mental efforts in order to learn physics effectively. Belief in this principle by no means is held only by nontraditional physics educators. Preliminary results from a large-scale interview study clearly suggest that these beliefs characterize the views of many physics instructors regarding student learning of problem solving skills: Patricia Heller, Kenneth Heller, Charles Henderson, Vince H. Kuo, and Edit Yerushalmi, ''Instructors' beliefs and values about learning problem solving,'' in *Proceedings of the 2001 Physics Education Research Conference*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (Rochester, New York, 2001), pp. 75–78.

[16]John R. Thompson, Paula R. L. Heron, Peter S. Shaffer, and Lillian C. McDermott, ''Development and assessment of curriculum,'' AAPT Announcer **28** (2), 80 (1998).

[17]Karen Cummings, Jeffrey Marx, Ronald Thornton, and Dennis Kuhl, ''Evaluating innovations in studio physics,'' Phys. Educ. Res., Am. J. Phys. Suppl. **67**, S38–S44 (1999). Also see Ref. 62.

[18]Ronald K. Thornton and David R. Sokoloff, ''Learning motion concepts using real-time microcomputer-based laboratory tools,'' Am. J. Phys. **58**, 858–867 (1990).

[19]Edward F. Redish and Richard N. Steinberg, ''Teaching physics: Figuring out what works,'' Phys. Today **52**, 24–30 (1999).

[20]Fred Goldberg, ''Constructing physics understanding in a computer-supported learning environment,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 903–911 (1997)], pt. 2.

[21]Raphael Littauer, ''Instructional implications of a low-cost electronic student response system,'' Educational Technology (October 1972), p. 69.

[22]E. A. Lumsden, ''Use of student feedback cards for diagnostic purposes during classroom lectures,'' in *Improving College and University Teaching Yearbook 1976* (Oregon State U.P., Corvallis, 1976), p. 39.

[23]A form of the ''feedback card'' method was described by Thomas A. Moore in his *Six Ideas That Shaped Physics* (WCB McGraw–Hill, Boston, 1998).

[24]F. Lyman, ''The responsive class discussion,'' in *Mainstreaming Digest*, edited by A. S. Anderson (College of Education, University of Maryland, College Park, 1981); F. Lyman, ''Think-pair-share, thinktrix, thinklinks, and weird facts,'' in *Enhancing Thinking Through Cooperative Learning*, edited by N. Davidson and T. Worsham (New York, Teacher's College Press, 1992).

[25]The ''minute paper'' has long been ascribed to Wilson as he was apparently the first to describe it in the literature [R. C. Wilson, ''Improving faculty teaching: Effective use of student evaluations and consultants,'' J. Higher Educ. **57**, 192–211 (1986)]. More recently, it has been acknowledged that the original source of the idea was Berkeley physicist C. Schwartz. See Barbara Gross Davis, Lynn Wood, and Robert C. Wilson, *A Berkeley Compendium of Suggestions for Teaching with Excellence* (University of California, Berkeley, 1983, available at http://teaching.berkeley.edu/compendium/, Suggestion #95).

[26]Charles C. Bonwell and James A. Eison, ''Active learning: Creating excitement in the classroom,'' ASHE-ERIC Higher Education Report No. 1, The George Washington University, School of Education and Human Development, Washington, DC, 1991), pp. 7–19 and *passim*, and references therein.

[27]Alan Van Heuvelen, ''Learning to think like a physicist: A review of research-based instructional strategies,'' Am. J. Phys. **59**, 891–897 (1991).

[28]Alan Van Heuvelen, ''Overview, case study physics,'' Am. J. Phys. **59**, 898–907 (1991).

[29]Alan Van Heuvelen, *ALPS Kit: Active Learning Problem Sheets, Mechanics* (Hayden–McNeil, Plymouth, MI, 1990).

[30]Alan Van Heuvelen, *ALPS Kit: Active Learning Problem Sheets, Electricity and Magnetism* (Hayden–McNeil, Plymouth, MI, 1990).

[31]Van Heuvelen has updated his methods by incorporating multimedia tools and innovative group activities: Alan Van Heuvelen, ''Experiment problems for mechanics,'' Phys. Teach. **3**, 176–180 (1995); ''Using interactive simulations to enhance conceptual development and problem solving skills,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 1119–1135 (1997)], pt. 2; Alan Van Heuvelen, Leith Allen, and Pavlos Mihas, ''Experiment problems for electricity and magnetism,'' Phys. Teach. **37**, 482–485 (1999).

[32]Eric Mazur, ''Understanding or memorization: Are we teaching the right thing?,'' in *Conference on the Introductory Physics Course*, edited by Jack Wilson (Wiley, New York, 1997), pp. 113–123.

[33]Eric Mazur, ''Peer Instruction: getting students to think in class,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 981–988 (1997)], pt. 2.

[34]Catherine H. Crouch, ''Peer Instruction: An interactive approach for large classes,'' Opt. Photonics News **9** (9), 37–41 (September 1998); Adam P. Fagen, Catherine H. Crouch, and Eric Mazur, ''Peer instruction results

from a range of classrooms,'' Phys. Teach. **40**(4), 206–209 (2002).

[35]Catherine H. Crouch and Eric Mazur, ''Peer instruction: Ten years of experience and results,'' Am. J. Phys. **69**, 970–977 (2001); http://mazur-www.harvard.edu/library

[36]Laurent Hodges, ''Changes in the introductory calculus-based physics course at Iowa State University,'' in *Conference on the Introductory Physics Course*, edited by Jack Wilson (Wiley, New York, 1997), pp. 291–294.

[37]R. G. Fuller, ''Using interactive lecture methods to teach physics,'' American Physical Society Forum on Education, Spring 1994.

[38]H. T. Hudson, ''Teaching physics to a large lecture section,'' Phys. Teach. **23**, 81–84 (1985).

[39]Robert J. Dufresne, William J. Gerace, William J. Leonard, Jose P. Mestre, and Laura Wenk, ''*Classtalk*: A classroom communication system for active learning,'' J. Comp. Higher Educ. **7**, 3–47 and http://umperg.physics.umass.edu/gemsFolder/umperg2/CT_Paper.pdf (1996).

[40]Jose P. Mestre, William J. Gerace, Robert J. Dufresne, and William J. Leonard, ''Promoting active learning in large classes using a classroom communication system,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 1019–1036 (1997)], pt. 2.

[41]Laura Wenk, Robert Dufresne, William Gerace, William Leonard, and Jose Mestre, ''Technology-assisted active learning in large lectures,'' in *Student-Active Science, Models of Innovation in College Science Teaching*, edited by Ann P. McNeal and Charlene D'Avanzo (Saunders College Publishing, Fort Worth, 1997), pp. 431–452.

[42]J. Poulis, C. Massen, E. Robens, and M. Gilbert, ''Physics lecturing with audience paced feedback,'' Am. J. Phys. **66**, 439–441 (1998).

[43]Joel A. Shapiro, ''Electronic student response found feasible in large science lecture hall,'' J. Coll. Sci. Teach. **26**, 408–412 (1997).

[44]Ray A. Burnstein and Leon M. Lederman, ''Using wireless keypads in lecture classes,'' Phys. Teach. **39**, 8–11 (2001).

[45]Robert J. Beichner, Jeffery M. Saul, Rhett J. Allain, Duane L. Deardorff, and David S. Abbott, ''Introduction to SCALE-UP: Student-centered activities for large enrollment university physics,'' Proceedings of the 2000 Annual Meeting of the American Society for Engineering Education, Session 2380, 2000 and http://www2.ncsu.edu/ncsu/pams/physics/Physics_Ed/index.html.

[46]Dean Zollman, ''Learning Cycle physics,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 1137–1149 (1997)], pt. 2.

[47]Priscilla W. Laws, ''Calculus-based physics without lectures,'' Phys. Today **44** (12), 24–31 (1991); ''Millikan Lecture 1996: Promoting active learning based on physics education research in introductory physics courses,'' Am. J. Phys. **65**, 14–21 (1997).

[48]David R. Sokoloff and Ronald K. Thornton, ''Using interactive lecture demonstrations to create an active learning environment,'' Phys. Teach. **35** (10), 340–347 (1997).

[49]David R. Sokoloff and Ronald K. Thornton, ''Using interactive lecture demonstrations to create an active learning environment,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 1061–1074 (1997)], pt. 2.

[50]Gregor M. Novak, Evelyn T. Patterson, Andrew D. Gavrin, and Wolfgang Christian, *Just-In-Time Teaching: Blending Active Learning with Web Technology* (Prentice–Hall, Upper Saddle River, NJ, 1999).

[51]David Hestenes, ''Modeling methodology for physics teachers,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 935–957 (1997)], pt. 2.

[52]Ruth W. Chabay and Bruce A. Sherwood, *Electric and Magnetic Interactions* (Wiley, New York, 1995).

[53]Bruce A. Sherwood and Ruth W. Chabay, ''Integrating theory and experiment in lecture using desktop experiments,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 1053–1060 (1997)], pt. 2.

[54]Randall D. Knight, *Physics: A Contemporary Perspective* (Addison–Wesley, Reading, MA, 1997), Vols. 1,2, preliminary ed. Also see Ref. 80.

[55]Ruth W. Chabay, ''Qualitative understanding and retention,'' AAPT Announcer **27** (2), 96 (1997).

[56]Randall D. Knight, ''Studio Physics at Cal Poly: What have we learned?'' AAPT Announcer **30** (2), 132 (2000).

[57]David E. Meltzer and Kandiah Manivannan, ''Promoting interactivity in physics lecture classes,'' Phys. Teach. **34**, 72–76 (1996).

[58]Kandiah Manivannan and David E. Meltzer, ''Increasing active student participation in the classroom through the use of 'flash cards,' '' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 821–822 (1997)], pt. 1; David E. Meltzer, ''Nontraditional approach to algebra-based general physics,'' *ibid.* **399**, 823–824 (1997).

[59]David E. Meltzer and Kandiah Manivannan, ''Interactive methods for large classes: Workshop W29 at the AAPT Winter Meeting, New Orleans, Louisiana, January 4, 1998''; David E. Meltzer and Kandiah Manivannan, ''Interactive methods for large classes: Workshop W36 at the AAPT Summer Meeting, Lincoln, Nebraska, August 4, 1998,'' AAPT Announcer **28** (2), 66 (1998); ''Workshop W04 at the AAPT Summer Meeting, San Antonio, Texas, August 3, 1999,'' **29** (2), 57 (1999); ''Workshop W11 at the AAPT Summer Meeting, Guelph, Ontario, Canada, July 29, 2000,'' **30** (2), 53 (2000).

[60]Lillian C. McDermott, Stamatis Vokos, and Peter S. Shaffer, ''Sample class on Tutorials in Introductory Physics,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 1007–1018 (1997)], pt. 2.

[61]Lillian C. McDermott, Peter S. Shaffer, and the Physics Education Group, *Tutorials in Introductory Physics* (Prentice–Hall, Upper Saddle River, NJ, 2002).

[62]Patricia Heller, Ronald Keith, and Scott Anderson ''Teaching problem solving through cooperative grouping. Group versus individual problem solving,'' Am. J. Phys. **60**, 627–636 (1992).

[63]Patricia Heller and Mark Hollabaugh, ''Teaching problem solving through cooperative grouping. 2. Designing problems and structuring groups,'' Am. J. Phys. **60**, 637–644 (1992).

[64]Patricia Heller, Thomas Foster, and Kenneth Heller, ''Cooperative group problem solving laboratories for introductory classes,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 913–933 (1997)], pt. 2.

[65]Patrica Heller and Kenneth Heller, *Cooperative Group Problem Solving in Physics* (University of Minnesota, Minneapolis, 1999) and http://www.physics.umn.edu/groups/physed/Research/CGPS/GreenBook.html.

[66]David R. Sokoloff, Ronald K. Thornton, and Priscilla W. Laws, *RealTime Physics, Active Learning Laboratories; Module 1: Mechanics; Module 2: Heat and Thermodynamics* (Wiley, New York, 1999).

[67]Priscilla W. Laws, with R. J. Boyle, P. J. Cooney, K. L. Laws, J. W. Luetzelschwab, D. R. Sokoloff, and R. K. Thornton, *Workshop Physics Activity Guide, The Core Volume with Module 1: Mechanics 1* (Wiley, New York, 1997).

[68]Pamela Ann Kraus, ''Promoting Active Learning in Lecture-Based Courses: Demonstrations, Tutorials, and Interactive Tutorial Lectures,'' Ph.D. dissertation, University of Washington (UMI, Ann Arbor, MI, 1997), UMI #9736313.

[69]Jack M. Wilson, ''The CUPLE physics studio,'' Phys. Teach. **32**, 518–523 (1994).

[70]*Classtalk*, a hard-wired system, may still be available from Interactive Classroom Consulting, http://www.bedu.com/ICC.html.

[71]A system based on infrared signaling is distributed by EduCue, 351 Alplaus Ave, Alplaus, NY 12008, http://www.educue.com; a different system is available from TI, http://education.ti.com/product/tech/tinav/overview/overview.html.

[72]A rf-based system is available from Socratec.com, http://www.socratec.com/.

[73]It is possible to use questions that focus on quantitative understanding or methods of calculation. We do so on occasion, depending on the nature of the topic and the course itself. In this regard, see Refs. 40 and 58. The strategy of using an ''easy-to-hard'' conceptual sequence has also been discussed by Mestre *et al.*, Ref. 40.

[74]This strategy is also emphasized by Crouch and Mazur, Ref. 35.

[75]See Sec. VIII. We refer to these sessions as ''tutorials'' because their

format and philosophy very closely match those developed at the University of Washington. However, for the most part, the materials we employ during these sessions are not the actual *Tutorials in Introductory Physics* cited in Ref. 61. The latter are used during three of the laboratory periods at ISU.

[76]The *Workbook* is distributed in three-hole-punched format for ring binders, so students do not have to bring all of the materials every day.

[77]Kandiah Manivannan and David E. Meltzer, ''Use of in-class physics demonstrations in highly interactive format,'' in *Proceedings of the 2001 Physics Education Research Conference*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (Rochester, New York, 2001), pp. 95–98.

[78]The idea of using specially designed worksheets in large lecture classes has also been discussed by Van Heuvelen (Refs. 27 and 28) and by Kraus (Ref. 68). Many questions in our worksheets also ask for explanations of students' reasoning. These explanations are emphasized and carefully checked during the tutorial sessions, but not so much so during the interactive lectures. We have not found it practical to make a rigid separation between worksheets used in lecture and those used in tutorials. In fact, which worksheets get used where is variable, and is essentially a function of day-to-day class scheduling.

[79]A preliminary edition of the *Workbook* is available from the authors in CD-ROM format. There is now also a vastly expanded inventory of ConcepTests available at the Project Galileo web site, http://galileo.harvard.edu/.

[80]Randall D. Knight, *Physics: A Contemporary Perspective, Student Workbook* (Addison–Wesley Longman, Reading, MA, 1997), Vols. 1, 2, preliminary ed.

[81]In fact, this type of student resistance to interactive-engagement physics courses has been discussed in the literature by a number of practitioners, for example, Ruth W. Chabay and Bruce A. Sherwood, *Instructor's Manual to Accompany Electric and Magnetic Interactions* (Wiley, New York, 1995), pp. 8 and 9.

[82]Additional insight regarding possible gender-related disparities in student responses are discussed by Priscilla W. Laws, Pamela J. Rosborough, and Frances J. Poodry, ''Women's responses to an activity-based introductory physics program,'' Phys. Educ. Res., Am. J. Phys. Suppl. **67**, S32–S37 (1999).

[83]The CSE was used in this abridged form, omitting some items, for various reasons. In some cases, the notational conventions differed from what was used in class (for example, electric field lines are used on the CSE, but only field vectors are used in class). In other cases, the questions dealt with material that was covered peripherally or not at all in class. This abridged subset consisted of the following item numbers from CSE Form G [corresponding CSEM item numbers in brackets]: 3, 4, 5, 6, 7, 8, 9[7], 10[8], 11[9], 12, 13[10], 14[11], 15, 16[12], 23[17], 24[18], 25[19], 26, 27[16], 28[20], 31[3], 32[4], 33[5]. (In several cases, there are minor differences between the CSE questions and the corresponding CSEM items.)

[84]We follow Hake's definition (Ref. 4) of ''normalized learning gain'' g, where $g=[$(post-test score-pre-test score)/(maximum possible score-pre-test score)]; $\langle g\rangle$ is calculated by using class-mean values for pre-test and post-test scores in the formula for g.

[85]David E. Meltzer, ''The relationship between mathematics preparation and conceptual learning gains in physics: A possible 'hidden variable' in diagnostic pretest scores?,'' Phys. Educ. Res., Am. J. Phys. Suppl. (submitted), and available at http://www.public.iastate.edu/~per/articles/ms/ms.pdf.

[86]The effect size d is a widely used measure in education research that quantifies the nonoverlap of two populations, typically including one that has, and another that has not received some specified pedagogical intervention. (Higher values of d correspond to greater nonoverlap, that is, larger treatment ''effect.'') See, for example, Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum, Hillsdale, NJ, 1988), 2nd ed., Chap. 2. The definitions given by Cohen are widely, though not universally, adopted: $d=|m_A-m_B|/\sigma_{\text{rms}}$; $\sigma_{\text{rms}}=\sqrt{(\sigma_A^2+\sigma_B^2)/2}$, where m_A and σ_A are the mean score and standard deviation of population A, and m_B and σ_B are those corresponding to population B. As an example, for the ISU 2000 sample we have $m_{\text{pre-test}}=33.7\%$ and $\sigma_{\text{pre-test}}=16.0\%$, $m_{\text{post-test}}=79.4\%$ and $\sigma_{\text{post-test}}=14.3\%$, $\sigma_{\text{rms}}=15.2\%$, and $d=45.7/15.2=3.01$.

[87]Of the 223 students in the three ISU samples combined, only 7 had individual values of $g\leqslant 0.22$.

[88]Peter S. Shaffer and Lillian C. McDermott, ''Research as a guide for curriculum development: An example from introductory electricity. II. Design of instructional strategies,'' Am. J. Phys. **60**, 1003–1013 (1992).

[89]Reference 4, Sec. V B 3.

[90]Reference 4, Sec. III A.

[91]Ronald L. Greene, ''Illuminating physics via web-based self-study,'' Phys. Teach. **39**, 356–360 (2001). In Greene's course, very strong emphasis was placed on qualitative, conceptual problems in examples, homework assignments, quizzes, and exams. Although interactive methods were not used during lectures, it is important to note that the extensive homework assignments made heavy use of a nontraditional, highly interactive web-based methodology which itself incorporated IE techniques such as immediate feedback.

[92]C. J. DeLeone, W. H. Potter, and L. B. Coleman, ''Comparisons of student MCAT performance: Traditional lecture/laboratory course vs. Physics Education Research based course,'' APS Centennial Meeting Program, Session LB20, Abstract LB20.07 (1999).

[93]The ratio of quantitative to qualitative problems on our quizzes and exams is approximately 1/1. Many problems are of a combined nature, involving both qualitative and quantitative elements; they often emphasize proportional reasoning in various contexts. During a semester, including quizzes, homework, and exams, students solve approximately 400 problems for grade credit.

[94]Lillian Christie McDermott, ''Oersted Medal Lecture 2001: Physics Education Research—The key to student learning,'' Am. J. Phys. **69**, 1127–1137 (2001).

IOWA STATE UNIVERSITY
Center for Teaching Excellence
Enhancing Student Learning and the Scholarship of Teaching

TEACHING AT ISU

November/December 2002
Vol. 15, No. 2

Using Active-Engagement Teaching Methods in Large-Enrollment Classes to Improve Student Learning

by David E. Meltzer
Assistant Professor
Department of Physics and Astronomy,
and 2002-03 CTE Teaching Scholar

I have always enjoyed learning about scientific concepts and explaining them to other people, and I used to spend a great deal of time and effort preparing extremely clear and detailed lectures. After a while, though, I could not avoid the realization that most of my students were not learning physics very well, despite my painstaking efforts to present concepts clearly, completely, and methodically. Although physics is a difficult subject, I felt that I should be doing a better job of communicating its ideas.

I became aware that university faculty engaged in physics education research were having success with instructional methods that employed "active engagement." In these methods, most often applied in instructional laboratories or small classes, instructors avoid giving students a fully worked-out set of answers and explanations right at the beginning. Instead, they guide students to figure out concepts on their own – as much as possible – through hands-on laboratory investigations or closely guided theoretical reasoning. Instructors guide students to follow productive lines of reasoning through a form of Socratic dialogue, asking many leading questions.

But can these instructional methods be employed in a lecture hall with 80 or more students? The answer is yes. Two effective techniques are: (1) guide students through a sequence of multiple-choice questions that force them to think deeply about the targeted concept, and use a classroom communication system to obtain instantaneous responses from all students simultaneously; (2) allow students to work in small groups on problems requiring non-multiple-choice responses such as diagrams or short answers. Responses to properly designed questions can be very quickly checked by the instructor who circulates around the lecture hall, examining the work of students near the aisles and front row.

The communication system I use is flash cards: each student is given six 5 x 8 cards on which the letters A, B, C, D, E, or F are printed. I write questions on the board along with several possible answers or provide pre-printed questions, and I'll usually give students 15-30 seconds to consider their answer. If they have trouble responding, or if there is much disagreement on the answers (for instance, half with "A" and half with "C") I'll give them another minute (or more) so they can discuss it with each other. This method allows a virtually continuous exchange of questions and answers between instructor and students.

Professor David Meltzer

I have done careful assessment of my students' learning over the years, using several standard conceptual tests as well as questions borrowed from other instructors' exams. I measure students' learning gains, that is, improvement from a pretest given on the first day of instruction to a post-test given the very last day. My students' gains are consistently above those reported in classes using more traditional forms of lecture instruction.

continued on page 2

INSIDE

They are exposed to fewer topics than in a traditional class, but seem to learn the concepts they study in much greater depth. They also learn to analyze problems qualitatively, and not simply by relying on equations. Course evaluations suggest that most students enjoy this method of instruction. Many more details about the assessments and the instructional methods can be found on the website of the ISU Physics Education Research Group, http://www.physics.iastate.edu/per/.

References:

1. David E. Meltzer and Kandiah Manivannan, "Transforming the lecture-hall environment: The fully interactive physics lecture," *American Journal of Physics* **70**, 639-654 (2002).

2. Eric Mazur, *Peer Instruction: A User's Manual* (Upper Saddle River, NJ, Prentice Hall, 1997).

Enhancing Active Learning in Large-Enrollment Physics Courses

David E. Meltzer, Assistant Professor, Department of Physics and Astronomy,
Iowa State University, Ames, IA
dem@iastate.edu

IINTRODUCTION

I have taught physics courses for physics majors, engineering students, and life-sciences majors, as well as for students planning careers as public-school teachers in both elementary and secondary schools. A common theme in all of these courses is “active learning,” that is: guiding students to maximum intellectual engagement with the material. A key strategy is to promote intensive interaction both between students and the instructor, and among the students themselves. This holds true whether one has a dozen elementary-education majors in a lab room, or 200 engineering students in a large lecture hall.

Much research suggests that learning of science concepts is enhanced when students are guided to analyze and draw conclusions from their own observations of physical phenomena (McDermott, 1991). Instead of instructors providing worked-out solutions and pre-packaged explanations, students are guided to “figure things out for themselves” with a minimum of intervention. When an instructor is working with just one or two students, this task might be relatively easy to accomplish. But when one faces 100 or more students simultaneously, the challenge of promoting maximum intellectual engagement can be extreme.

In this paper I describe methods I have used with great success to promote active learning in large-enrollment physics classes. The strategies are based on *guiding students along productive lines of reasoning through a question-and-answer process in a group-learning environment.* In a different context I have used this same strategy in small classes for pre-service elementary teachers. Although the specific techniques described here might differ from those used in a small class, the overall strategy is essentially the same: help students learn efficiently by aiding them to ask and answer intellectually provocative questions. The goal is to catalyze, in the students’ own mind, the conceptual breakthroughs needed for understanding of scientific concepts.

THE PROBLEM: LARGE CLASSES

Imagine you are beginning your lecture in a room filled with 150 students. Many of them—perhaps most—appear to be attentive and expectant. You start your carefully prepared presentation, striving to be as clear as possible. Every now and then you ask a question of the class, pause and wait for someone to answer, and then comment on their response. Repeatedly, you ask if anyone has questions; only rarely does anyone respond. You’re a bit uneasy about the lack of questions—surely they’re not finding your explanations to be all *that* clear? You wonder how well your students actually understood your lecture. Were you able to clear up the tricky points you knew would cause them trouble? You can wait until the exam and see how well they do, but does this really tell

To be published in *Best Practices and Lessons Learned: Highlights from the NSF Collaboratives for Excellence in Teacher Preparation Program and Other Innovative Programs Around the Country*, edited by Diane Smith and Elisabeth Swanson (Montana State University, Bozeman, MT, in press)

you whether they got anything out of your lecture? For years, I wondered whether there was some way out of this frustrating dilemma. Eventually, I decided that indeed there was a way.

In the following paragraphs, I will describe methods developed in close collaboration with Kandiah Manivannan of Southwest Missouri State University (Meltzer and Manivannan, 2002a; Meltzer and Manivannan, 1996; see also http://www.physics.iastate.edu/per/index.html). I have used these methods primarily in the second semester of the algebra-based general physics course, a course taken mostly by students in the life sciences including pre-medical and pre-veterinary students. The majority of enrolled students are female. I have taught this course at Southeastern Louisiana University (Physics 192: Fall 1995-Spring 1998) and Iowa State University (Physics 112: Fall 1998-Fall 2002). Both institutions are typical in that their large student enrollments result in many large lecture courses. In physics, this means that an instructor teaching an introductory course might face anywhere from 50 to 250 students at one time. Both students and instructors are often dissatisfied with the "anonymous" atmosphere of such classes, and have a common interest in improving the effectiveness of the learning environment in these large lecture courses.

THE SOLUTION: INTERACTIVE ENGAGEMENT

Our basic strategy is to drastically increase the quantity and quality of interaction that occurs in class between the instructor and the students, and among the students themselves. To this end, the instructor poses many questions. All of the students must decide on an answer to the question, discuss their ideas with each other, and provide their responses to the instructor. The instructor makes immediate use of these responses by tailoring the succeeding questions and discussion to most effectively match the students' pace of understanding. Our methods are, in effect, a variant of "Peer Instruction," which was developed by Eric Mazur at Harvard University (Mazur, 1997; Crouch and Mazur, 2001). Instructional methods that emphasize interaction among students and instructors combined with rapid feedback have been referred to by Richard Hake as "interactive engagement" (Hake, 1998).

As a model of this learning environment, consider the instructor's office. When you have one or two students in your office asking for help, do you lecture to them for 50 minutes, pausing occasionally to ask a question? More likely you speak for just a few minutes, sketching diagrams and writing a few simple equations. Then you stop and ask for some feedback. Maybe you pose a simple question or sketch out a problem for them to try, or ask one student to comment on an answer given by the other. In the office, you are able to get an ongoing sense of where your students are at conceptually, and how well they are following the ideas you're presenting. By getting continual feedback from them, you're able to tailor your presentation to their actual pace of understanding. By asking them to consider each other's ideas, you help them to think critically about their own ideas. But is it practical to do this in a room filled with over 100 students?

My answer is that it *is* practical. It is possible to recreate in the lecture hall much of the learning environment that exists in the instructor's office. One can transform—to a substantial extent—the environment of the lecture hall into that of a small seminar room in which *all* of the students are actively engaged in the discussion. It takes preparation and practice to do it well, but any instructor who is committed to the effort should be able to succeed. Here I will describe the methods I use in my large lecture classes.

THE FULLY INTERACTIVE LECTURE

To begin with, I give up the idea of delivering long lectures. As much as I used to love to lecture, I hardly do it anymore because I have become painfully aware of how ineffective it is. I used to enjoy carefully and precisely outlining my hard-won insights about difficult physics concepts. I would present these concepts slowly and painstakingly, with great clarity, never glossing over confusing points. As long as students were paying close attention, it was simply inconceivable to me that anyone could fail to follow my crystal-clear logic. Inconceivable, that is, until I really began to interact with my students in the lecture hall. I realized, to my dismay, that most of my students were *not* understanding my beautifully clear lectures—not at all. My carefully crafted arguments flew right over their heads, leaving only confusion. Sometimes they convinced themselves that they understood my words—but, in fact, they were usually wrong. What I did to discover that this was true any instructor can do, and I suspect they would come to a similar realization.

I now get instantaneous feedback simultaneously from all the students in the class. I ask questions during class—many questions—and no longer have to wait for one brave soul to dare to offer a response. Every single student in the class has a pack of six large "flash cards" (5½″ × 8½″), each printed with one of the letters A, B, C, D, E, or F. They bring the cards every day, and I always have extras in case someone forgets. Repeatedly during class I will present a multiple-choice question to the students. The questions stress qualitative concepts involving comparison of magnitudes (e.g., "*Which is larger: A, B, or C?"*), direction (*"Which way will it move?"*) and trends (*"Will it decrease, remain the same, or increase?"*). These kinds of questions are hard to answer by plugging numbers into an equation. I give the students some time to consider their response, 15 seconds to a minute depending on the difficulty of the question. Then I ask them to signal their response by holding up one of the cards, everybody at once. I can easily see all the cards from the front of the room. Immediately, I can tell whether most of the students have the answer I was seeking – or if, instead, there is a "split vote," some with one answer, some with another. (I hope that one is the right answer!)

I can see whether the class held up their cards quickly, with confidence, or if instead they brought them up slowly, with confused looks on their faces. If there is a split vote, I ask them to talk to each other. I allow about a minute for those who think the answer is, say, "A" to try to persuade those who believe it is "C" to change their views. And, of course, the "C" supporters argue for their side of the case. Then I ask for another vote. If it is still split, I'll ask for an "A" supporter to stand and present their argument, followed (in alphabetical order) by a proponent of the "C" point of view. Eventually, if necessary, I will step in to—I hope—alleviate the confusion. But by this time, most of the students will have thought through the concept that was causing the problem because they will have tried to convince their neighbors that they were right. And, if they haven't already figured things out by themselves, they will now at least be in an excellent position to make sense out of any argument I offer to them. Before that minute or two of hard thinking, though, I could have made the same argument and then watched as almost every student in the class gave the wrong answer to some simple question. I know this is true, because I have tried it often enough.

By now I have had many opportunities to ask my students questions during my lecture that I would once have considered "trivial." These questions pertain to concepts that I—and most instructors—would have covered in a few seconds or a minute of clear, logical reasoning. I would have said that it was impossible for my students to get these simple questions wrong, or have any difficulty with them. But in fact they do, and now I know it. I pose a question that, I think, is a completely straightforward application of a principle I just presented. For instance: *If a two-resistor parallel circuit is increased to three resistors in parallel, what happens to the total power provided by the battery?* The logic points inescapably toward only one possibility. I wait as my students study the question, debating the answer with each other, looking around. Slowly, after a minute, the cards come up: half are "A" (decreases), and nearly a third are "B" (remains the same). But the correct answer is "C" (increases), a choice selected by perhaps one student out of five.

I realize that I need to retreat, and I offer another question—perhaps I make it up on the spot—that goes back to a concept discussed last week. Then we work our way through a series of intermediate questions, back to the one that started the trouble. At each step, I get a reading on my class: Do they respond quickly? With confidence? *Mostly* correctly? Then I comment briefly and move forward. Otherwise, I pause for a longer discussion. In the old days I would have disposed of this entire topic in less than two minutes of lecture, and have been well satisfied that I made my points clearly and effectively. Now I take 10 to 15 minutes, and struggle together with my students as they work their way through a conceptual minefield. But this time, I believe, my students really do construct a basis for understanding the material. And, I realize, the self-satisfaction of the old days was no more than wishful thinking and self-deception.

CLASS FORMAT

A typical class proceeds in three phases:

(1) A brief introduction/review of the basic concepts is presented at the blackboard, a sort of "mini-lecture" lasting three to seven minutes.

(2) A sequence of about a half-dozen multiple-choice questions (sometimes more) is posed to the class; these questions emphasize qualitative understanding, proceed from easier to more challenging, and are closely linked to each other to explore just one or two concepts from a multitude of perspectives. They frequently employ graphs, diagrams, and verbal descriptions. Students provide responses to these questions using the flash cards as described above.

(3) The students then proceed to work on free-response questions in the form of integrated worksheets, which again stress diagrammatic and graphical representations. The students work in groups while the instructor circulates throughout the room, rapidly scanning the students' work by looking over

their shoulder. It is easy to quickly assess the graphs, diagrams, and short answers that comprise the bulk of the responses.

This method is crucially dependent on having at one's disposal a large number of carefully constructed sequences of conceptual multiple-choice questions. The purpose of emphasizing non-numerical questions is to prevent students short-circuiting the thinking process by blindly plugging numbers into poorly understood equations. Although some collections of such problems exist in the literature (Mazur, 1997; Novak et al., 1999), we have had to construct our own set to meet the needs of a full one-semester course (Meltzer and Manivannan, 2002b). It is the preparation and testing of such question sets that is among the most time-consuming prerequisites for this instruction. Our questions are based, as much as possible, on the physics education research literature (McDermott and Redish, 1999).

The free-response questions are also presented in a highly structured sequence, designed to lead students to think deeply about fundamental conceptual issues. These worksheets are largely designed after the model of the University of Washington Tutorials (McDermott et al., 2002), although here adapted for large classes by somewhat more gentle pacing. Both the multiple-choice question sets and the free-response worksheets are provided to the students in the form of a three-hole-punched workbook, and they are required to bring relevant sections to class every day. I have also written a complete set of lecture notes which are now bound together with the workbook. These notes offer concise reference materials that heavily emphasize qualitative understanding, and provide numerous sample questions of the type used on quizzes and exams.

Another critical course element is the continual—almost relentless—feedback. Written quizzes are given every Monday and Friday and count for 1/3 of the total grade. Additional group-quiz points are available on Wednesday. Homework must be handed in during the Thursday "tutorial" (recitation) meetings. (Tutorials consist of group work on worksheets while two teaching assistants circulate throughout the room.) The net result of these incentives is a consistent 90% attendance rate for both lectures and recitations.

INSTRUCTIONAL OUTCOMES

I have found that overall learning gains by the students in this course are very high in relation to comparable courses nationwide. For the past several years I have given the "Conceptual Survey of Electricity," a diagnostic instrument that assesses qualitative understanding. My students' pretest scores (about 30%) are nearly identical to those reported in comparable algebra-based courses, and substantially lower than those in a nationwide sample of about 1500 students in calculus-based courses. However, the average post-test scores of my students in Physics 112 at Iowa State (taught five times from Fall 1998 to Fall 2002) were in the 75-79% range, while those of the nationwide sample range from around 43% in the comparable algebra-based course to approximately 51% for students in the calculus-based class (Meltzer and Manivannan, 2002a; Maloney et al., 2001). Other assessment data are consistent with these results. Moreover, on quantitative problems borrowed from exams given in the calculus-based course at Iowa State (Physics 221), students in my algebra-based course do comparably well, or better.

One of the most dramatic consequences of this instructional method is a very small number of dropouts, typically 1-3% after the first week. The low dropout rate combined with the strong evidence of good learning gains are, for me, the key test of the instructional methods. However, it is also important to note that the majority of students seem to react favorably to the instructional methods, as shown by their responses to end-of-semester surveys. Their feelings are reflected in their evaluations of the instructor and their comments on the instructional methods. From 1998-2002, 75% gave top ratings of 4 or 5 on a 1-5 scale. (Sample comment: *". . . best physics instructor I have ever had. I liked the way he had class interaction and explained things. He makes physics fun and interesting to learn, whereas most physics instructors just babble inanely during lecture"*). Most of the remainder are neutral, but a persistent core of 10% or less *despises* these methods and is vocal about that fact. (Sample comment from the same class: *". . . has a new way of teaching he is trying to develop. It doesn't work. He relies too heavily on the students to help each other, when all we want is to learn the material . . . going to lecture was pointless other than to take required quizzes."*)

CONCLUSION

The overall result of these methods is, for me, little short of a revelation regarding student learning. By exposing what I believe to be a realistic picture of how my students learn during lectures, I feel that I have been able to transform the classroom experience for them. Previously, this experience—while enjoyable for the instructor and (perhaps) entertaining for the students—served to do little more than inform them of the topics they needed to study on their own. I now believe that my students are actually *learning* during class, and building a much firmer basis for their out-of-class work.

My collaborator, Kandiah Manivannan, and I have given many workshops for other instructors to help them learn about our instructional methods, and we have published very detailed accounts of the methods that have been disseminated widely. Our CD of the instructional materials (Meltzer and Manivannan, 2002b) has been distributed free to many hundreds of physics instructors worldwide, on request, and many of them have told us that they have used our methods and materials successfully in their own classes. With support from the National Science Foundation, we are now engaged in developing additional materials for other topics in the introductory physics curriculum. We are hopeful that we will be able to achieve learning gains in other areas of the curriculum that are comparable to what we have documented in our previous work.

REFERENCES

Crouch, Catherine H. and Mazur, Eric. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics* 69, 970-977.

Hake, Richard R. (1998). Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 66, 64-74.

Maloney, David P., O'Kuma, Thomas L., Hieggelke, Curtis J., and Van Heuvelen, Alan. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics* 69, S12-S23.

Mazur, Eric. (1997). *Peer Instruction: A User's Manual.* Prentice Hall, Upper Saddle River, New Jersey.

McDermott, Lillian Christie. (1991). Millikan Lecture 1990: What we teach and what is learned—Closing the gap. *American Journal of Physics* 59, 301-315.

McDermott, Lillian C. and Redish, Edward F. (1999). Resource Letter: PER1: Physics Education Research. *American Journal of Physics* 67, 755-767.

McDermott, Lillian C., Shaffer, Peter S., and the Physics Education Group. (2002). *Tutorials in Introductory Physics.* Prentice-Hall, Upper Saddle River, New Jersey.

Meltzer, David E. and Manivannan, Kandiah. (1996). Promoting interactivity in physics lecture classes. *The Physics Teacher* 34 (2), 72-76.

Meltzer, David E. and Manivannan, Kandiah. (2002a). Transforming the lecture-hall environment: The fully interactive physics lecture. *American Journal of Physics* 70, 639-654.

Meltzer, David E. and Manivannan, Kandiah. (2002b). *Workbook for Introductory Physics: Part II*. CD-ROM available from the author.

Novak, Gregor M., Patterson, Evelyn T., Gavrin, Andrew D., and Christian, Wolfgang. (1999). *Just-In-Time Teaching: Blending Active Learning with Web Technology.* Prentice-Hall, Upper Saddle River, New Jersey.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Nos. DUE #0243258 and DUE #0311450. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

IV.

Student Learning and Reasoning in Thermodynamics

Student reasoning regarding work, heat, and the first law of thermodynamics in an introductory physics course

David E. Meltzer, Department of Physics & Astronomy, Iowa State University, Ames, IA 50011

Abstract: Written quiz responses of 653 students in three separate courses are analyzed in detail.

There has been relatively little research on student learning of thermodynamics in physics courses at the university level. A recent study by Loverude *et al.*[1] has made it evident that students at the introductory level (and beyond) face many significant difficulties in learning fundamental thermodynamic concepts such as the first law of thermodynamics.

I have been engaged in an ongoing project with T. J. Greenbowe to investigate student learning of thermodynamics in both physics and chemistry courses.[2] As part of that investigation, a short diagnostic quiz has been administered over the past two years in the calculus-based introductory physics course at Iowa State University (ISU). This quiz focuses on heat, work, and the first law of thermodynamics.

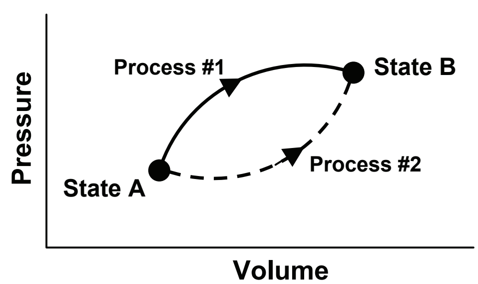
At ISU, thermodynamics is studied at the end of the second semester of the two-semester sequence in calculus-based introductory general physics. This course is taught in a traditional manner, with large lecture classes (up to 250 students), weekly recitation sections (about 25 students), and weekly labs taught by graduate students. Homework is assigned and graded every week. Thermal physics comprises 18-20% of the course coverage, and includes a wide variety of topics such as calorimetry, heat conduction, kinetic theory, laws of thermodynamics, heat engines, entropy, etc.

The diagnostic quiz used in this study is shown below; it has been administered in three separate classes. The version shown here was administered in May 2001; the other two versions (December 1999 and December 2000) had very minor variations from the one shown here. (There were one or two additional questions on these quizzes which are not discussed here.)

The 1999 and 2000 classes were taught by the same instructor, using a different textbook in each course. The 2001 course was taught by a different instructor, using the same text that was employed in the 1999 course. Both instructors are very experienced and have taught introductory physics at ISU for many years.

The quiz was administered in two different ways: in 1999 and 2001, it was given as a practice quiz in the final recitation session (last week of class). In almost all cases it was ungraded; one instructor used it as a graded quiz. In 2000 the quiz was administered as an ungraded practice quiz in the very last lecture class of the year.

This *p-V* diagram represents a system consisting of a fixed amount of ideal gas that undergoes two ***different*** processes in going from state A to state B:



[In these questions, ***W*** represents the work done ***by*** the system during a process; ***Q*** represents the heat ***absorbed*** by the system during a process.]

1. Is *W* for Process #1 ***greater than, less than***, or ***equal to*** that for Process #2? Explain.
2. Is *Q* for Process #1 ***greater than, less than***, or ***equal to*** that for Process #2? Please explain your answer.

Fig. 1. Thermodynamics diagnostic quiz

This paper appeared in *Proceedings of the Physics Education Research Conference, Rochester, New York, July 25-26, 2001*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (PERC, Rochester, New York, 2001), pp. 107-110.

Answers:

1. $W = \int_{V_A}^{V_B} pdV$ *= the area under the curve in the p-V diagram, so* $W_1 > W_2$.

2. $\Delta E_1 = \Delta E_2 \Rightarrow Q_1 - W_1 = Q_2 - W_2 \Rightarrow Q_1 - Q_2 = W_1 - W_2$. Therefore, $W_1 > W_2 \Rightarrow Q_1 > Q_2$. (Since system #1 loses more energy by doing more work, it must gain more energy through heat absorption to have the same net change in internal energy.)

Correct explanations for #1 were considered to be virtually anything that mentioned "area under the curve," the integral $\int_{V_A}^{V_B} pdV$, "working against higher pressure," etc.

A liberal standard was used in assessing answers to #2; examples of answers considered correct:

"$\Delta E = Q - W$. For the same ΔE, the system with more work done must have more Q input so process #1 is greater."

"Q is greater for process 1 since Q = E + W and W is greater for process 1."

"Q is greater for process one because it does more work, the energy to do this work comes from the Q_{in}."

An analysis of students' responses on the quiz is shown in Tables I and II.

Table I: Students' reasoning on Work question (*Note: explanations not required in 1999)

	1999 (*n*=186)	2000 (*n*=188)	2001 (*n*=279)
$W_1 > W_2$	**73%**	**70%**	**61%**
Correct or partially correct explanation	*	56%	48%
Incorrect or missing explanation	*	14%	13%
$W_1 = W_2$	**25%**	**26%**	**35%**
Because work is independent of path	*	14%	23%
Other reason, or none	*	12%	13%
$W_1 < W_2$	**2%**	**4%**	**4%**

Table II: Students' reasoning on Heat question

	1999 (*n*=186)	2000 (*n*=188)	2001 (*n*=279)
$Q_1 > Q_2$	**56%**	**40%**	**40%**
Correct or partially correct explanation	14%	10%	10%
Q is higher because pressure is higher	12%	7%	8%
Other incorrect, or missing explanation	31%	24%	22%
$Q_1 = Q_2$	**31%**	**43%**	**41%**
Because heat is independent of path	21%	23%	20%
Other explanation, or none	10%	18%	20%
$Q_1 < Q_2$	**13%**	**12%**	**17%**
Nearly correct, sign error only	4%	4%	4%
Other explanation, or none	10%	8%	13%
No response	**0%**	**4%**	**3%**

CONCEPTUAL DIFFICULTIES IDENTIFIED IN STUDENTS' RESPONSES

1. Difficulty interpreting work as "area under the curve" on a *p-V* diagram. Although most students correctly responded that $W_1 > W_2$, only about 50% of all students were able to give an acceptable explanation. This basic geometrical interpretation is usually the very first topic discussed in connection with *p-V* diagrams, and it is difficult to make efficient use of such diagrams without understanding this idea.

2. Belief that work done is independent of process. A substantial number (15-25%) of students are under the impression that work is (or behaves as) a state function, and that the work done during a process depends only on the initial and final states. Many students state this very explicitly in their written explanations. Others do not have such a clearly expressed notion, but still identify the work done by the two processes in the diagram as being equal to each other.

3. Belief that heat absorbed is independent of process. About 20-25% of all students explicitly state a belief that the heat absorbed during a process depends only on the initial and final states. (Answers categorized as "Because heat is independent of path" include those stating that both processes reached the same final state, had the same initial and final states, etc.) In addition, the claim that $Q_1 = Q_2$ was justified by a wide variety of other explanations.

4. Association of greater heat absorption with higher pressure. The most popular alternative explanation for $Q_1 > Q_2$ was that higher pressures were involved in Process #1. It was clear, though, that students were not considering the process as a whole (omitting, e.g., any consideration of initial and final states), and were simply associating "heat" with "pressure," often through appeals to the ideal gas law.

5. Use of a "compensation" argument, e.g., "more work implies less heat," etc. A significant number of students attempted to employ an argument that states, roughly speaking, "more heat (or work) implies less work (or heat)." For instance, only 5% of students who claimed $W_1 = W_2$ also argued that $Q_1 < Q_2$; however, that argument was made by 20% of students who had correctly answered $W_1 > W_2$. In some cases, it was clear that students were employing the first law of thermodynamics in the form $\Delta E = Q + W$ (i.e., W being defined as work done ***on*** the system). This was *not* the convention used in their physics class, although it is typically the one used in chemistry courses. An analogous argument was used by other students who explicitly employed $\Delta E = Q - W$; these students were often making a simple sign error (and are categorized as "Nearly correct, sign error only" in Table II). The "compensation" argument was also seen in the explanations of the (very few) students who stated that $W_1 < W_2$; most of them went on to argue that $Q_1 > Q_2$.

6. Inability to make use of the first law of thermodynamics. Even including students who made sign errors (as described above), only about 15% of all 653 students were able to give a correct answer with a correct explanation based on the first law of thermodynamics. There was almost no variation in this proportion from one class to the next, despite changes in instructors and textbooks.

CLUES REGARDING CONCEPTUAL DYNAMICS

Among the most interesting and important aspects of students' reasoning (from the instructor's standpoint) is the *path* along which learning takes place.[3] By this I mean the sequences of ideas that lead either to productive or unproductive lines of thought from the standpoint of yielding good learning outcomes. In the present case we have an observation of student thinking at only a single point in time. Therefore, any hypotheses we induce from the data must be tested through sequential observations and student interviews. Nonetheless, there are several provocative aspects of the data that are consistent over all the observations.

A. Patterns underlying students' responses

1. Although a belief in path-independence of heat is somewhat more common among students who answer $W_1 = W_2$, more than one third of those who correctly answer $W_1 > W_2$ *also* claim that $Q_1 = Q_2$. About half of the students who answer $W_1 = W_2$ also state that $Q_1 = Q_2$ (*1999:* 40%; *2000:* 51%; *2001:* 53%). However, a very substantial number of those who realize that work is dependent on process (and correctly answer $W_1 > W_2$) also seem to believe that heat is *not* process dependent. This is implied by the fact that more than one third of those who answer $W_1 > W_2$ also claim that $Q_1 = Q_2$: *1999:* 29%; *2000:* 41%; *2001:* 34%. This somewhat unexpected result is made more provocative by the following observation.

2. Students are more likely to justify a $Q_1 = Q_2$ answer by explicitly asserting that "Q is path-independent" ***if they answered the Work question correctly.*** Students who answered the Work question incorrectly and who also stated $Q_1 = Q_2$ often gave no explanation for their answer to the Heat question. Only infrequently did they claim that heat was "independent of process" or use words to that effect (e.g., "both processes ended at the same point," "had the same initial and final points," etc.). By contrast, students who answered the Work question *correctly* but stated that $Q_1 = Q_2$ usually *did* explicitly claim that heat was independent of process. (See Tables III, IV.)

Table III. Students who answer $Q_1 = Q_2$ (2000)

2000	**Correct on work question (*n* = 54)**	**Incorrect on work question (*n* = 27)**
Explain by claiming "heat is independent of path"	61%	36%
Explain with other reasons, or no explanation given	39%	63%

Table IV. Students who answer $Q_1 = Q_2$ (2001)

2001	**Correct on work question (*n* = 58)**	**Incorrect on work question (*n* = 55)**
Explain by claiming "heat is independent of path"	66%	35%
Explain with other reasons, or no explanation given	34%	65%

B. Conjectures on conceptual dynamics

1. Belief that heat is process-independent may not be strongly affected by realization that work is *not* process-independent. The process-dependence of both heat and work are fundamental concepts in thermodynamics. Because the formalism of *p-V* diagrams is ubiquitous in physics instruction, a very natural representation of the idea of process dependence is that different paths, representing different processes, are characterized by different amounts of work done ("areas under the curve"). It might seem then that the process-dependence of work should be easier to grasp, at least at the formal level, than that of heat. One might think that when a student gains this perception about work, the idea of heat *also* being dependent on process would not be such a big leap. The data suggest that the linkage between these concepts in instruction may not be as close as one might guess.

2. Understanding the process-dependence of work may strengthen belief that heat is *independent* of process. Various interpretations of the data in Tables III and IV are possible. For instance, students who have a good grasp on the concept that "work is area under the curve" may also have a clearer perception than do other students that *something*, at least, is independent of process in thermodynamics. If they have not yet clearly grasped the idea of internal energy change, they may too readily transfer that perception, mistakenly, to *heat*. On the other hand, these data may simply reflect a better ability to express their (incorrect) ideas on the part of students who correctly answer the Work question.

This material is based upon work supported by the National Science Foundation under Grant Number DUE-9981140.

REFERENCES

1. M.E. Loverude, P.R.L. Heron, and C.H. Kautz, "Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas," Phys. Educ. Res., Am. J. Phys. Suppl. (*to be published*).
2. D.E. Meltzer and T.J. Greenbowe, "Recurrent areas of confusion in student learning of thermodynamics," AAPT Announcer **31**(2), 81 (2001).
3. R.K. Thornton, "Conceptual Dynamics: following changing student views of force and motion," in AIP Conf. Proc. **399,** E.F. Redish and J.S. Rigden, eds. (AIP, N.Y., 1997), pp. 241-266.

PHYSICS EDUCATION RESEARCH SECTION

All submissions to PERS should be sent (preferably electronically) to the Editorial Office of AJP, and then they will be forwarded to the PERS editor for consideration.

Investigation of students' reasoning regarding heat, work, and the first law of thermodynamics in an introductory calculus-based general physics course

David E. Meltzer[a)]
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

(Received 17 July 2002; accepted 13 July 2004)

Students in an introductory university physics course were found to share many substantial difficulties related to learning fundamental topics in thermal physics. Responses to written questions by 653 students in three separate courses were consistent with the results of detailed individual interviews with 32 students in a fourth course. Although most students seemed to acquire a reasonable grasp of the state-function concept, it was found that there was a widespread and persistent tendency to improperly over-generalize this concept to apply to both work and heat. A large majority of interviewed students thought that net work done or net heat absorbed by a system undergoing a cyclic process must be zero, and only 20% or fewer were able to make effective use of the first law of thermodynamics even after instruction. Students' difficulties seemed to stem in part from the fact that heat, work, and internal energy share the same units. The results were consistent with those of previously published studies of students in the U.S. and Europe, but portray a pervasiveness of confusion regarding process-dependent quantities that has been previously unreported. Significant enhancements of current standard instruction may be required for students to master basic thermodynamic concepts.
[DOI: 10.1119/1.1789161]

I. INTRODUCTION

Thermodynamics has a wide-ranging impact, as is demonstrated by the number of different fields in which it plays a fundamental role both in practice and in instruction. The broad-based and interdisciplinary nature of the subject has motivated us to engage in a project to develop improved curricular materials that will increase the effectiveness of instruction in thermodynamics. We are initially investigating the effectiveness of current, standard instruction in order to pinpoint student learning difficulties that might potentially be addressed with alternate instructional approaches.

Given the fundamental importance of thermodynamics, it is surprising that there has been little research into student learning of this subject at the university level. Although there have been hundreds of investigations into student learning of the more elementary foundational concepts of thermodynamics (such as heat, heat conduction, temperature, and phase changes) at the secondary and pre-secondary level, the number of published studies that focus on university-level instruction on the first and second laws of thermodynamics is on the order of ten, of which only one was devoted to physics students at U.S. universities.[1]

Prior work has demonstrated convincingly that pre-university students face enormous obstacles in learning to distinguish among the concepts of heat, temperature, internal energy, and thermal conductivity. In physics, heat (or heat transfer) is a process-dependent variable and represents a *transfer* of a certain amount of energy between systems due to a temperature difference. By contrast, in the kinetic theory of a gas, temperature is a measure of the average kinetic energy of the molecules in a system. However, among beginning science students heat is frequently interpreted as a mass-independent *property* of an object and temperature is interpreted as a measure of its intensity. Often, temperature and heat are thought to be synonymous. Alternatively, heat often is interpreted as a specific quantity of energy possessed by a body with temperature a measure of that quantity.[2–4] Objects made of materials that are good thermal conductors are believed by students to be hotter or colder than other objects at the same temperature, due to the sensations experienced when the objects are touched.[5] Instructors at the university level often have noted similar ideas among their own students,[6] and investigations that have probed university students' thinking about these concepts have recently appeared.[7]

A few investigations have been reported that examined pre-university students' understanding of the concept of entropy and the second law of thermodynamics.[3,8] Several reports have examined student learning of thermodynamics concepts in university chemistry courses.[9–15] Some of these studies have touched on first- and second-law concepts in addition to topics more specific to the chemistry context. Among the investigations directed at university-level physics instruction, one in France focused on oversimplified reasoning patterns used by students when thinking about thermodynamics, particularly when explaining multivariable phenomena with reference to the ideal gas law.[16] A German study examined the learning of basic thermal physics concepts by students preparing to become physics teachers.[17] There also was a very brief report of a survey of entrants to a British university,[18] and a study related to U.S. students' concepts of entropy and the second law of thermodynamics.[19]

The first detailed investigation of university physics students' learning of heat, work, and the first law of thermodynamics was published by Loverude, Kautz, and Heron in 2002.[20] (Additional details are in Loverude's dissertation.[21]) This study incorporated extensive data collected from observations at three major U.S. universities and documented serious and numerous learning difficulties related to fundamental concepts in thermodynamics. It was found that many students had a very weak understanding of the work concept and were unable to distinguish among fundamental quantities such as heat, temperature, work, and internal energy. Only a small proportion of students in introductory courses were found to be able to make use of the first law of thermodynamics to solve simple problems in real-world contexts.

The present investigation includes an independent examination of some of the same research questions analyzed in Ref. 20 and other, related questions. A preliminary report of the work described here appeared in 2001.[22]

Our findings include several previously unreported aspects of students' reasoning about introductory thermodynamics. In contrast to at least one previous report,[11] it was found that students have a reasonably good grasp of the state-function concept. However, students' understanding of process-dependent quantities was seriously flawed, as sizeable numbers of students persistently ascribe state-function properties to both work *and* heat. This confusion regarding work and heat is associated with a strong tendency to believe that the net work done and the net heat absorbed by a system undergoing a cyclic process are both zero. Interview data disclosed unanticipated levels of confusion regarding the definition of thermodynamic work and heretofore unreported difficulties with the concept of heat transfer during isothermal processes. Consistent results over several years of observations enabled us to make a high-confidence estimate of the prevalence of difficulties with the first law of thermodynamics among students in the calculus-based general physics course. Our findings should help provide instructors of introductory physics with a solid basis on which to plan future instruction in thermodynamics.

II. CONTEXT OF THE INVESTIGATION

Our data were collected during 1999–2002 and were in three forms: (1) a written free-response quiz that was administered to a total of 653 students in three separate offerings (Fall 1999, Fall 2000, Spring 2001) of the calculus-based introductory physics course at Iowa State University (ISU); (2) a multiple-choice question that was administered to 407 students on the final exam during the 2001 course offering; and (3) one-on-one interviews that were conducted with 32 student volunteers who were enrolled in a fourth offering of the same course in Spring 2002.

A. Written diagnostic

Thermodynamics is studied at ISU during the second semester of the two-semester sequence in calculus-based introductory general physics, which is offered during both the fall and spring semesters. Most students taking this course are engineering majors. The course is taught in a traditional manner, with large lecture classes (up to 250 students), weekly recitation sections (about 25 students), and weekly labs taught predominantly by graduate students. Homework is assigned and graded every week. Thermal physics comprises 18–25% of the course coverage, and includes a wide variety of topics such as calorimetry, heat conduction, kinetic theory, laws of thermodynamics, heat engines, and entropy.

The 1999 and 2000 classes were taught by the same instructor, using a different textbook in each course. The 2001 course was taught by a different instructor, using the same text (later edition) that was employed in the 1999 course.[23] Both instructors are very experienced and have taught introductory physics at ISU for many years. (The author was not involved in the instruction in any of the courses that served as a basis for this study.)

A written diagnostic quiz (described in Sec. IV) was administered in two different ways: in 1999 and 2001, it was given as a practice quiz in the final recitation session (last week of class). In nearly all cases it was ungraded, although one recitation instructor used it as a graded quiz. In 2000 the quiz was administered as an ungraded practice quiz in the last lecture class of the semester. In addition, a multiple-choice problem similar to those on the diagnostic quiz was administered on the final exam of the 2001 course.

B. Interviews

During the Spring 2002 offering of this course, instead of administering a written diagnostic quiz, student volunteers were solicited to participate in one-on-one problem-solving interviews in which their reasoning processes were probed in depth. This course was taught by the same instructor as the Spring 2001 course. Thermal physics topics occupied 25% of the class lectures, and a different text[24] was used than in the previous courses. Due to travel obligations, two different faculty members (the professor in charge of the course, plus another very experienced instructor) were responsible for presenting the thermodynamics lectures.

Exam questions and assigned homework problems included calculations of work done, heat transferred, and changes in internal energy during various processes (some represented on P-V diagrams), including adiabatic, isothermal, isobaric, and numerous cyclic processes. Other questions related to the temperature/kinetic energy/internal energy relationship, and to the efficiency of heat engines and refrigerators. (There also were many problems related to the other thermal physics topics covered during the course.)

All lectures and homework assignments related to thermal physics were completed before the second midterm exam. This exam included questions related to the role of the thermal reservoir in an isothermal expansion, changes in internal energy during a cyclic process, and many questions related to entropy, engines, and the second law of thermodynamics.

Interviews began five weeks after the second midterm exam, and continued over a three-week period through the week of final exams. A new set of questions was developed for the interviews. (These are the Interview Questions shown in the Appendix and discussed in Sec. IV.) The average duration of each interview was over 1 h, including time for the students to work by themselves. Many interviews extended longer than that period, and a few were shorter. All were recorded on audiotape. Students were asked to explain as best they could how they obtained their answers to the questions. When inconsistencies appeared in their responses, they were urged to address them. This often led to changes in responses, often from incorrect to correct, sometimes from one incorrect answer to a different one, but only very rarely from a correct response to one that was incorrect. Substantial efforts were exerted to ensure that students very clearly understood the meaning of the questions, diagrams, and spe-

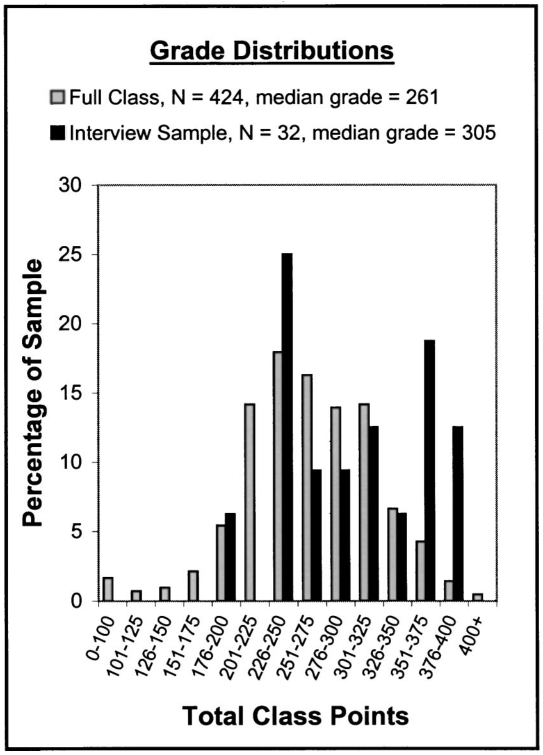


Fig. 1. Grade distributions for the interview sample ($N=32$) and for the full class from which the interview sample was drawn ($N=424$). Grades based on total class points (nominal maximum $=400$). The interview sample mean score (300) and median score (305) are well above the corresponding scores for the full class (mean score $=261$, standard deviation $=59$; median score $=261$).

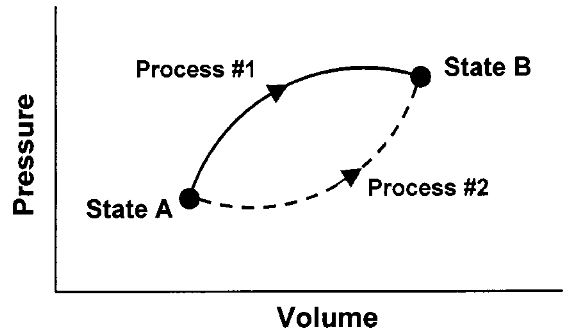
cific terminology employed. Any apparent ambiguities in the students' interpretations of the questions were explicitly addressed by the interviewer (the author).

III. CHARACTERIZATION OF THE INTERVIEW SAMPLE

There were 32 students in the interview sample. They were drawn from 13 different recitation sections (out of a total of 20), taught by seven different recitation instructors (out of a total of nine), and 66% were engineering majors. Other majors with at least two representatives were computer science, chemistry, and meteorology; there was one physics major. All but one had studied physics while in high school, and many had taken Advanced Placement physics or a community college physics course while in high school.

The grading in the course was based on exam scores (three midterm exams and a final) plus a recitation-laboratory grade; the nominal maximum total points available was 400. The distributions of total class points (out of 400) both for the full class ($N=424$) and the interview sample ($N=32$) are plotted in Fig. 1 as a percentage of each population. It can be seen that the scores of the students in the interview sample are strongly skewed toward the top end of the class. More than one third of the interview sample scored above the 91st percentile of the class, and half scored above the 81st percentile; only two students in the interview sample fell below the 25th percentile. It is evident that the average level of knowledge demonstrated by the interview sample is very unlikely to be lower than that of the class population as a whole.

This *P-V* diagram represents a system consisting of a fixed amount of ideal gas that undergoes two ***different*** processes in going from state A to state B:



[In these questions, ***W*** represents the work done ***by*** the system during a process; ***Q*** represents the heat ***absorbed*** by the system during a process.]

1. Is *W* for Process #1 ***greater than***, ***less than***, or ***equal to*** that for Process #2? Explain.

2. Is *Q* for Process #1 ***greater than***, ***less than***, or ***equal to*** that for Process #2? Please explain your answer.

3. Which would produce the largest change in the total energy of all the atoms in the system: ***Process #1***, ***Process #2***, or ***both processes produce the same change?***

Fig. 2. Written quiz used in investigation, referred to as "Diagnostic Questions." This version was administered in Spring 2001. Responses to this quiz are shown in Tables I and II.

IV. DIAGNOSTIC QUESTIONS AND INTERVIEW QUESTIONS

The written diagnostic quiz is shown in Fig. 2; it was administered in four separate courses. The version shown here was administered in Spring 2001, and it was also used (with minor wording changes to match the terminology of the course textbook) during the interviews conducted in Spring 2002. The Fall 1999 and Fall 2000 versions had very minor variations from the one shown in Fig. 2 with respect to Questions #1 and #2. A different version of Question #3 was used in 1999, and it was omitted entirely in 2000.

For the interviews, an additional separate set of questions was developed consisting of eight sequential questions related to two cyclic processes. (Before being presented with the questions, interview subjects were first asked to respond to the written diagnostic quiz.) The questions are shown in the Appendix. A *P-V* diagram corresponding to the processes described in these questions is shown in Fig. 3; this diagram was not given to the students. (Note that this process is the same as depicted in Fig. 4 of Ref. 20, although traversed in the opposite direction.) Students were asked to

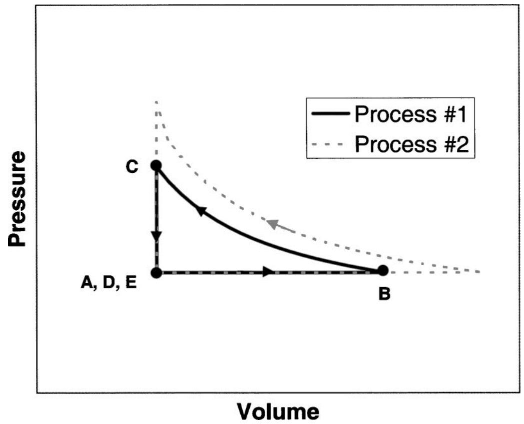


Fig. 3. A P-V diagram corresponding to processes described in the Interview Questions. (This diagram was not shown to the students.)

circle their answers to these questions and verbally explain the reasoning they used to obtain their answers. (Several minor changes in wording to the questions were made to improve clarity during the course of the series of interviews.)

The multiple-choice question administered on the 2001 final exam will be described in Sec. VI.

V. THERMAL PHYSICS CONCEPTS: PREDOMINANT THEMES OF STUDENTS' REASONING

The students' responses to items #1 and #2 of the diagnostic questions are shown in Tables I and II, respectively. The responses in the 1999, 2000, and 2001 samples were very consistent from one year to the next. They also are consistent with the verbal and written responses given to the same questions by students in the interview sample. In Table III, the responses of students in the interview sample to the questions in the Appendix are tabulated.

In the following, I will examine in detail the most prevalent concepts in students' thinking. In each case the subheading refers to a reasoning pattern common to a minimum of 20–25% of all students in the respective samples.

A. Relation between temperature and molecular kinetic energy

A fundamental link between the macroscopic and microscopic models of thermodynamics lies in the proportionality between temperature and the average molecular kinetic energy of a gas. Almost all introductory texts use the kinetic theory of gases to provide a derivation of the relation $KE_{tot} = (3/2)nRT$ for the total molecular kinetic energy contained within n moles of a monatomic ideal gas. Interview Question #3 asks students about possible changes in the total kinetic energy of the molecules of the system during the isothermal compression occurring from time B to time C. No deep understanding is required to respond that this energy remains unchanged during the process. Although a slim majority (56%) of students give this answer, nearly one third assert that the total molecular kinetic energy will increase. This difficulty in matching an isothermal ideal-gas process with no change in molecular kinetic energy has not been previously reported.

During the interviews, students who asserted that the molecular kinetic energy would change during the isothermal process were usually asked to explain what role, if any, the temperature had played in their reasoning. The most common line of reasoning is typified by these responses:

(The designation ''S11'' refers to student #11, using an arbitrary numbering system for students in the interview sample.)

> ''[S11] There's a higher pressure; the molecules are moving faster, hitting the sides faster, which creates a larger pressure. And so since they're moving faster, they have a higher kinetic energy.''
>
> ''[S21] When the volume decreases, something has to make up for it. In this case the pressure's going to increase. If you add more pressure you're going to increase the collisions of the particles, and so ... the kinetic energy will increase because of that. They're moving faster; kinetic energy is related to the speed of the particles ... *Interviewer: Did the temperature play any part of this, any consideration here*? Yes ... If you're going to increase the pressure, the temperature also increases ... *Interviewer: I should point out that ... the temperature is the same as at time B* ... In that case then, the temperature would not have a factor on kinetic energy ... The kinetic energy varies with the temperature, but the temperature doesn't change; it won't affect the kinetic energy. In this case, the pressure's the only part of the $PV=nRT$ equation that's going to affect the kinetic energy.''

Reference 20 pointed out that students frequently invoked a ''collision'' argument similar to that used by these two students, to account for temperature increases during adiabatic compression. The same observation was made by Ro-

Table I. Responses to diagnostic Question #1 (work question).

	1999 ($N=186$)	2000 ($N=188$)	2001 ($N=279$)	2002 Interview Sample ($N=32$)
$W_1>W_2$	**73%**	**70%**	**61%**	**69%**
Correct or partially correct explanation	[a]	56%	48%	66%
Incorrect or missing explanation	[a]	14%	13%	3%
$W_1=W_2$	**25%**	**26%**	**35%**	**22%**
Because work is independent of path	[a]	14%	23%	22%
Other reason, or none	[a]	12%	13%	0%
$W_1<W_2$	**2%**	**4%**	**4%**	**9%**

[a]Explanations not required in 1999.

Table II. Responses to diagnostic Question #2 (heat question).

	1999 (N=186)	2000 (N=188)	2001 (N=279)	2002 Interview Sample (N=32)
$Q_1 > Q_2$	**56%**	**40%**	**40%**	**34%**
Correct or partially correct explanation	14%	10%	10%	19%
Q is higher because pressure is higher	12%	7%	8%	9%
Other incorrect, or missing explanation	31%	24%	22%	6%
$Q_1 = Q_2$	**31%**	**43%**	**41%**	**47%**
Because heat is independent of path	21%	23%	20%	44%
Other explanation, or none	10%	18%	20%	3%
$Q_1 < Q_2$	**13%**	**12%**	**17%**	**13%**
Nearly correct, sign error only	4%	4%	4%	3%
Other explanation, or none	10%	8%	13%	9%
No response	**0%**	**4%**	**3%**	**6%**

zier and Viennot in their study of French university students.[25] In the present study, it is seen for the first time that the argument that molecular collisions produce a net increase in molecular kinetic energy is so compelling for many students that they apply it even in the case of an isothermal process, persisting even after acknowledging the existence of a relation between temperature and kinetic energy. For many students, the relationship between temperature and the molecular kinetic energy of an ideal gas—considered virtually axiomatic by many instructors—is one that is only vaguely understood.

B. The concept of state function in the context of energy

The concepts of state and state function are fundamental to thermal physics and provide a starting point for the analysis of all thermodynamic phenomena and processes. Question #3 on the written quiz probes understanding of these concepts. (This question was not administered in 1999 and 2000.) In the 2001 sample, 73% responded correctly to this question, saying that the total energy change in the two processes would be the same. In the interview sample, 88% provided this correct response. Of the students in the latter sample, 78% provided an acceptable explanation of their answer, that is, they either associated the energy change of the atoms with the temperature change and noted that these changes would be equal for the two processes, or they explicitly stated that the energy (or internal energy) was a state function and depended only on initial and final states, was independent of path, etc. A similar problem dealing with this issue is Interview Question #7. As shown in Table III, 90% of students in the interview sample gave a correct answer to this question with an acceptable explanation.

In 1999, instead of Question #3 as shown in Fig. 2, the following question was presented: ‘‘Consider a system that begins in State A, undergoes Process #1 to arrive at State B, and then undergoes the *reverse* of Process #2, thereby arriving once again at State A. During this entire back-and-forth process ($A \rightarrow B \rightarrow A$), does the internal energy of the system ($E_{\rm int}$) undergo a *net increase*, a *net decrease*, or *no net change*? Explain your answer.’’

Of the 186 students in the 1999 sample, 85% correctly answered that the internal energy of the system would undergo no net change in the cyclic process described; 70% gave an acceptable explanation for their answer. These results along with those from 2001 suggest that students become comfortable with the idea that a thermodynamic system might be in one or another state, where a state is characterized by a certain value for the total energy contained within the system. They seem to realize that in making a transition from one state to another, the particular process involved in the transition does not affect the net energy change, and that the net change is determined only by the initial and final states. When the system follows a route that brings it back to that initial state, they are able to see that the total energy also must return to its initial value.

During the course of the interviews, it was evident that students associated not only a specific energy value with a given thermodynamic state, but realized that each state was characterized by well-defined values for the pressure, volume, and temperature as well. Although very few students spontaneously articulated a precise definition of ‘‘state,’’ state function, or internal energy, they solved problems and provided explanations in a manner that was consistent with at least a rudimentary understanding of those concepts. (This conclusion is in marked contrast to the conclusions of Kaper and Goedhart in relation to Dutch chemistry students in a thermodynamics course.[11])

Many of the conceptual difficulties encountered by students in the context of thermal physics seemed to stem from an overgeneralization of the concept of state function. In thermal physics, quantities (such as heat transfer and work) which are *not* state functions, but instead characterize specific thermodynamic processes, are equally as important as state functions to understanding and applying thermodynamic principles. Most of our remaining discussion will be devoted to analyzing students' reasoning regarding these process-dependent quantities, as well as the first law of thermodynamics which relates these quantities to the internal energy.

C. Work as a mechanism of energy transfer

An elementary notion in thermal physics is that if a system characterized by a well-defined pressure undergoes a quasistatic process in which a boundary is displaced, energy is transferred between the system and the surrounding environment in the form of work. If the volume of the system increases, internal energy of the system is transferred to the environment and we say that work is done *by* the system; conversely, if the volume decreases, work is done *on* the system and energy is transferred *to* it. The critical distinction

Table III. Responses to Interview Questions ($N=32$).

Question	Response	Proportion giving response
#1		
	Work is done *on* the gas	31%
	Work is done *by* the gas (*correct*)	69%
#2		
	Increases by x Joules	47%
	Increases by less than x Joules	41%
	with correct explanation	*28%*
	with incorrect explanation	*13%*
	Remains unchanged	9%
	Uncertain	3%
#3		
	Increase	31%
	Decrease	13%
	Remain unchanged (*correct*)	56%
#4		
	No	59%
	Yes, from water to gas	3%
	Yes, from gas to water	38%
	with correct explanation	*31%*
	with incorrect explanation	*6%*
#5		
	Decreases by less than y Joules	16%
	Decreases by y Joules (*correct*)	84%
#6, i		
	Greater than zero	16%
	Equal to zero	63%
	Less than zero (*correct*)	19%
	No response	3%
#6, ii		
	Greater than zero	9%
	Equal to zero	69%
	Less than zero	16%
	with correct explanation	*13%*
	with incorrect explanation	*3%*
	Uncertain	6%
#7[a]		
	All equal (*correct*)	90%
	Other response, or none	10%
#8[b]		
	$\lvert W_1\rvert=\lvert Q_1\rvert=0$	50%
	$\lvert W_1\rvert=\lvert Q_1\rvert\neq 0$ (*correct*)	16%
	Uncertain	6%
	Other response	28%

[a] $N=30$.
[b] Responses regarding Process #1 only.

is not so much in recognizing whether the words ‘‘by’’ or ‘‘on’’ should be used in a particular instance; rather, it is essential to recognize whether energy is transferred *into* or *out of* a system as a result of the process.

Loverude *et al.* have described and documented many of the difficulties students encounter when studying the concept of work, both in the context of mechanics and in that of thermal physics.[20] They showed that few students were spontaneously able to invoke the concept of work when discussing the adiabatic compression of an ideal gas. Students were unable to understand that an entity called work could bring about a change in the internal energy of a system. There was a tendency to treat the concept of work as superfluous, as unconnected to temperature changes in gases, or on the other hand, as being essentially synonymous with heat. Many students were unable to recognize that heat and work are independent means of energy transfer.

The results of our investigation fully support their conclusions and offer additional insight into the nature of student reasoning regarding work in the context of thermodynamics. Responses given during the interviews to Questions #1 and #2 reveal that approximately 1/3 to 1/2 of the students in the interview sample have a substantial confusion regarding this concept.

Interview Question #1 asks students whether positive work is done on or by the gas during the isobaric expansion process from time A to time B. To answer, a student must recognize that the expansion of a system corresponds to positive work being done by the system on the surrounding environment. However, 31% of the students in the interview sample said that the expansion process described in Question #1 corresponded to positive work being done *on* the gas by the environment. They backed up their answer with explanations that made it clear that this error was not merely a semantic confusion:

> ‘‘[S31] The gas is expanding and for it to expand, heat or energy or something had to be put into it to get it to expand. And, since the only option of putting stuff into the gas is ‘*a*’ [*positive work done on the gas by the environment*], that’s why I picked ‘*a*.’ ’’
>
> ‘‘[S20] The environment would be water and stuff ... water would be part of that, and since it moved the piston up ... the environment did work on the gas, since it made the gas expand and the piston moved up ... water was heating up, doing work on the gas, making it expand.’’

These and similar responses suggest that many students simply do not realize that as the gas expands against its surrounding environment, the gas *loses* energy as a result of the work done during the process. They realize that there is energy transfer to the gas in the form of heat, but do not seem to recognize that there is energy transfer away from the gas in the form of work. Instead, as previously pointed out in Ref. 20, students make a fundamental error by identifying ‘‘work’’ with energy transfer in the form of heat, and in general they have difficulty distinguishing between the two quantities. In the case of adiabatic compression, students in the Loverude *et al.*[20] study had used ‘‘heat’’ when ‘‘work’’ would have been appropriate. Analogously, in the case of isobaric expansion, students often use the word ‘‘work’’ to refer to a heating process. The belief that positive work is done *on* a system by the environment during an expansion process has not been previously reported.

It is interesting to compare this observation to results of a study by Goldring and Osborne[26] of students taking A-level physics in London secondary schools. (This level is roughly equivalent to introductory college physics in the U.S.) They found that more than half of the students in their study claimed that work is done both when an object is heated and also whenever energy is transferred. Similarly, nearly half said that heat is always created when work is done.

The problem of not recognizing the energy-transfer aspect of macroscopic work plays an even more significant role in students’ responses to Interview Question #2, and it is this set of responses that validates the interpretation of students’ thinking proposed above in connection with Question #1. Students are told that the gas absorbs x Joules of energy from the water during the heating-expansion process, and are asked what will happen to the total kinetic energy of all the gas molecules. The correct answer (''increases, but by less than x Joules'') was given by 41% of the students, but only 28% could provide a correct explanation such as this student's answer:

> ''[S9] Some heat energy that comes in goes to expanding, and some goes to increasing the kinetic energy of the gas.''

Almost half of the students (47%) answered that ''the total kinetic energy of all of the gas molecules increases by x Joules,'' with explanations such as

> ''[S3] For it to increase by less than x Joules that energy would have to go somewhere, so that would say that the potential energy of the gas had increased, and I don't see how that would be happening.''
> ''[S4] There would be conservation of energy. If you add that much, it's going to have to increase by that much.''
> ''[S5] Kinetic energy is going to increase by x Joules because, I assume that there's no work done by expansion, that it doesn't take any kind of energy to expand the cylinder, which means that all of my energy is translated into temperature change.''

This fundamental confusion regarding the energy-transfer role of work is a very serious obstacle to understanding the basic principles of thermal physics, and in particular serves as a nearly insuperable barrier to grasping the meaning of the first law of thermodynamics.

D. Belief that work is a state function

P-V diagrams permit a simple interpretation of the work done by a system during a process as the area under the curve describing the process. Many elementary problems involve calculations of work done during different processes linking common initial and final states, in order to illustrate and emphasize the concept that work is a process-dependent function and not a state function. It is all the more remarkable, then, that the results of our investigation show so clearly that approximately one quarter of all students in our samples are confused about this fundamental concept. This corroborates the findings of Ref. 20, which documented widespread misunderstanding of this concept among both introductory and advanced physics students when it was presented in the context of P-V diagrams.

Table I shows responses to Question #1, comparing the work done by two different processes linking initial state A and final state B. In this diagram, it is very clear that the area under the curve representing process #1 is greater than the area under the curve representing process #2, and so the work W done by the system is greater for process #1. However, 30% of the students who answered the written diagnostic in 1999, 2000, and 2001 asserted that the work done during process #1 would be equal to the work done during process #2. Of the students who were asked to provide an explanation, 19% explicitly argued that work was independent of the path. Similarly, 22% of the interview subjects claimed that $W_1 = W_2$, all of whom made an explicit argument asserting that work was independent of process, for example: ''work is a state function,'' ''no matter what route you take to get to state B from A, it's still the same amount of work,'' ''for work done take state A minus state B; the process to get there doesn't matter.''

It is evident that many students come to very directly associate thermodynamic work with properties (and even specific phrases) discussed by instructors and texts only in connection with internal energy and other state functions. This is consistent with the conclusion of Ref. 20 that students frequently have difficulty in distinguishing among work, heat, and internal energy, and in particular with their finding that many students explicitly assert the path independence of work. As they point out, it seems that overgeneralization of (poorly understood) experience with conservative forces may contribute to students' confusion about these issues.

E. Belief that heat is a state function

Among the most striking results of our investigation is that a very significant fraction of introductory students in our sample (between one third and one half) developed the idea that heat (or ''heat transfer'') is a state function, independent of process. In view of all textbooks' strenuous and oft-repeated emphasis that heat transfer is a process-dependent quantity and not a state function, this is a remarkable observation. Although several studies have noted a confusion between heat and internal energy, none have explicitly and systematically probed students regarding their understanding of the *path-dependent* property of heat transfer.[27]

Question #2 may be answered by realizing that $\Delta U_1 = \Delta U_2$ and then employing the first law of thermodynamics to obtain $Q_1 - W_1 = Q_2 - W_2$. Because the diagram shows that $W_1 > W_2$, we can conclude that $Q_1 > Q_2$. However, well over a third (38%) of the 653 students responding to Question #2, and 47% of the students in the interview sample answering the same question, asserted that the heat absorbed by the system during process #1 would be equal to that absorbed during process #2. Moreover, 21% of the students in the written sample, and 44% of those in the interview sample, offered explicit arguments regarding the path-independence of heat, for example: ''I believe that heat transfer is like energy in the fact that it is a state function and doesn't matter the path since they end at the same point''; ''transfer of heat doesn't matter on the path you take''; ''they both end up at the same PV value so ... they both have the same Q or heat transfer.'' About 150 students offered arguments similar to these either in their written responses or during the interviews.

Strong support for the idea that heat is process-independent was consistent in all four student samples. The only other explanation (aside from the correct explanation) to gain any significant support on Question #2 was one that ascribed higher Q in process #1 simply to ''higher pressure,'' without giving any consideration to the initial and final states of the two processes.

Also remarkable is that the belief in the process independence of heat was widespread even among students who clearly understood that work is not a state function, as well as among those who mistakenly believed that work also is independent of process. Of the students who incorrectly answered that $W_1 = W_2$, about half also asserted that $Q_1 = Q_2$ (1999: 40%; 2000: 51%; 2001: 53%; interview sample: 43%). However, this mistaken notion regarding heat is nearly as common among the students who realize that work is dependent of process, and who correctly answered that

$W_1 > W_2$. Of this group, more than one third also asserted that $Q_1 = Q_2$ (1999: 29%; 2000: 41%; 2001: 34%; interview sample: 50%).

This observation of students' belief in a state-function property for heat is consistent with the findings of other researchers, although as noted it goes well beyond what has previously been reported. The tendency of students to mistakenly identify heat with the state function internal energy was noted and discussed in Ref. 20 and the same observation was made by Berger and Wiesner in their interviews with advanced-level German university students in the teacher preparation program who had studied thermodynamics.[17] Manthei and Täubert[28] reported similar observations in an analysis of written responses on questions posed to advanced-level German high-school students. They, too, found a tendency to identify heat with internal energy, as well as a widespread inability to correctly identify heat as a ''process quantity'' instead of a ''state quantity.'' Similarly, a great deal of confusion was found regarding the definition of heat among entrants at a British university,[18] and Kaper and Goedhart[11] concluded that Dutch chemistry students often treat heat as a state function.

It appears that the confounding of heat with internal energy, noted in Refs. 20 and 28, extends to an explicit association of the state-function property with heat. This confusion is quite analogous to the set of mistaken associations developed by many students in connection with work, as described in Sec. V D. We must consider the possibility that students' familiarity with the equation $Q = mc\,\Delta T$ and its use in elementary calorimetry problems may contribute to their confusion regarding the nature of heat.

F. Belief that net work done and net heat transferred during a cyclic process are zero

The single most prevalent misconception encountered during our investigation was the strong belief expressed during the interviews that during a cyclic process, the net work done by the system or the net heat transferred to the system must be zero. In Ref. 20 it was noted that many students believe that the net work in a cyclic process must be zero due to the zero net change in volume. This belief often is so tenacious as to override other considerations that would imply nonzero net work.[20] In our investigation, this finding is corroborated and amplified by uncovering a parallel belief in the necessity of zero net heat transfer during a cyclic process. This belief regarding zero net heat transfer has not been documented in the literature.

Interview Question #6 asks students to consider the entire process that had been described, beginning at time A and ending at time D. They were asked whether the net work done by the gas, and the total heat transferred to the gas, are positive, negative, or zero. (''Total heat transferred'' matches the terminology of the course textbook.) Only a small minority of students realized that the net work done (35%) or that the total heat transferred (25%) would be nonzero. Less than one fifth of the students could give correct answers with satisfactory explanations to the work question (19%) or the heat question (13%). Only three students in the entire sample (9%) gave fully correct responses to both parts of Question #6, such as this answer:

> ''[S17] The total work was less than zero. I drew a diagram, pressure versus volume, and the path that I scratched out here is counterclockwise, which suggests negative work ... [The total heat transfer] is less than zero ... in order to have negative work done it needs to have less than zero heat transferred to it if it's to maintain its same initial state ... Negative work done by the gas, so if it absorbs heat here, its output is going to have to be work plus heat. So, the total heat transfer is negative because this heat coming out of the gas is greater than the heat going into it, because it includes the energy from the work and the heat going into it.''

Of the students in the interview sample, 75% either believed that the net work done by the gas, or the total heat transferred to the gas, or both, would be zero for the entire process. More than half (56%) said that both the net work done and the total heat transferred throughout the entire process would be zero. In almost every case, the reasoning was the same: Because the final position of the piston was the same as its initial position, the negative work would cancel the positive work; because the final temperature was the same as the initial temperature, the heat transferred into the system would be balanced by the heat transferred out of the system:

> ''[S1] The net work done by the gas ... is equal to zero ... The physics definition of work is like force times distance. And basically if you use the same force and you just travel around in a circle and come back to your original spot, technically you did zero work.''
>
> ''[S27] The work done by the gas on the environment is positive in the first steps where the piston goes up, but then when it goes back down it's negative. And so, since it ends up in the same place, the net work is zero.''
>
> ''[S21] The heat transferred to the gas ... is equal to zero ... The gas was heated up, but it still returned to its equilibrium temperature. So whatever energy was added to it was distributed back to the room.''

Students were asked to explain how they could be sure that the magnitude of the positive work (or heat) would exactly equal the magnitude of the negative work (or heat). In nearly every case, the students again referred to the equality of the final and initial values of the volume and temperature. Some students argued (as also was reported in Ref. 20) that because $W = \int P\,dV$ and $\Delta V = 0$, ''work equals zero.''

Interview Question #8 was another opportunity to probe students' thinking on this matter. Here students were asked to rank the absolute values of the net work done by the gas and total heat transferred to the gas, both for the process that takes place between times A and D (symbolized by $|W_1|$ and $|Q_1|$, respectively), and for a similar process with initial and final states the same as before, but characterized by higher intermediate values of the pressure and temperature. Whenever there appeared to be a discrepancy in the students' answers for Questions #6 and #8, they were asked to comment or resolve the discrepancy. (The tables reflect students' final decisions in all cases.) Table III shows the students' responses to Question #8 regarding process #1 (time A to time D) only. Exactly half answered that $|W_1| = |Q_1| = 0$, while only 16% stated correctly that $|W_1| = |Q_1| \neq 0$. Overall, 66%

claimed either that $|W_1|=0$, or that $|Q_1|=0$, or that both equal zero. The responses to Question #8 thus confirm the results from Question #6.

As will be discussed, only a minority of the students referred to a P-V diagram when answering Interview Questions #1–8. However, at the end of the interview, all students were asked to carefully draw a P-V diagram representing processes #1 and #2. More than 90% of them ultimately drew a diagram of a cyclic process. It is noteworthy that only four students realized that their diagrams implied an error in their initial response that $|W_1|=0$ or $|Q_1|=0$. (These students' final answers are reflected in the tabulated data.) Several other students expressed misgivings regarding the possible inconsistencies of their answers, but were unable to arrive at a correct resolution.

In the study of Ref. 20, students in an algebra-based course were presented with a P-V diagram that corresponded to the process described here. Although one might expect the presence of the diagram to have made the problem easier, about half of the students in that study asserted that the net work done by the gas during the process was zero, typically mentioning that there was no net change in volume. It seems clear that the ''no net change in volume'' theme plays a dominant role in student reasoning. The results of our investigation further suggest that the same could be said about the ''no net change in temperature'' theme.

G. Confusion regarding isothermal processes and the thermal reservoir

Students' responses to Interview Question #4 revealed additional aspects of their difficulties in applying the work concept, and also manifest a deep misunderstanding of the concept of thermal reservoir. This question refers to the isothermal compression that occurs between time B and time C; the question asks whether there is any net energy flow between the gas and the water reservoir during this process. Only 31% of the students answered correctly with an acceptable explanation, with acceptable being loosely defined to include explanations such as:

> ''[S6] There'd be a flow of energy from the gas to the water. Because, when you compress a gas, normally it would heat things up. And so, if everything is remaining at somewhat of an equilibrium, I'm just going to assume, because it's in such a large environment, that that kind of heat would kind of dissipate into the environment.''

Only a small minority of these acceptable explanations made an explicit reference to the unchanging internal energy of the gas or to the first law of thermodynamics. In contrast, 59% of the students said that there would be no net energy flow between gas and water. Invariably, they mentioned that the gas and water temperatures were equal and unchanging:

> ''[S2] I would think if there was energy flow between the gas and the water, the temperature of the water would heat up.''
>
> ''[S10] There is no energy flow between the gas and the water; it all stayed in the system. Since the temperature stayed the same, there is no heat flow.''

Most of the students who said that there would be no net energy transfer between the gas and the water reservoir were asked to comment explicitly on whether there could be any energy transfer to or from a gas undergoing an isothermal process. Most agreed that it would be possible, citing situations such as having ''light or energy coming out,'' having heat energy ''converted into potential energy or kinetic energy,'' ''if heat in equals heat out,'' or if there is ''expansion or contraction.'' However, none of these students believed that the process described in Question #4 fit any of their proposed circumstances.

Isothermal processes are ubiquitous in the introductory thermal physics curriculum, and invariably reference is made to a constant-temperature reservoir with which the system is in contact. The details of how the isothermal process actually takes place are very rarely discussed, with a notable exception in Chabay and Sherwood's text *Matter & Interactions*:[29] ''*As we compress the gas, the temperature in the gas starts to increase. However, this will lead to energy flowing out of the gas into the water, because whenever temperatures differ in two objects that are in thermal contact with each other, there is a transfer of energy from the hotter object to the colder object ... Energy transfer out of the gas will lower the temperature of the gas ... Quickly the temperature of the gas will fall back to the temperature of the water. The temperature of the big tub of water on the other hand will hardly change ... Therefore the entire quasistatic compression takes place essentially at the temperature of the water, and the final temperature of the gas is the same as the initial temperature of the gas*.''

It is clear that most of the students in the interview sample had never understood the details of an isothermal process as described above. They were unable to apply the first law of thermodynamics to a situation in which the isothermal compression of an ideal gas immediately implies the existence of a nonzero heat transfer out of the system.

A similar difficulty in understanding the role of a reservoir was noted by van Roon *et al.*[12] in their investigation of college chemistry students in Holland. Moreover, in a study of advanced undergraduate college science students enrolled in physical chemistry courses (at the junior–senior level), Thomas and Schwenz[14] reported that 60% of their interview sample believed that ''no heat occurs under isothermal conditions.'' Students' tendency to hold that belief also was noted in Refs. 20 and 21. However, our work is the first unambiguous finding, based on a significant sample size, of students' confusion regarding energy transfer during an isothermal process.

H. Inability to apply the first law of thermodynamics

In the investigation of Ref. 20, the majority of students examined were unable to employ the first law of thermodynamics to solve problems related to adiabatic compression. Similar difficulties in other contexts were displayed by students in the present study.

First let us consider students' responses to Question #2: ''Is Q for process #1 greater than, less than, or equal to that for process #2? Please explain your answer.'' (The fact that all relevant values of ΔU, Q and W are positive here minimizes the potential confusion regarding signs.) An example of an acceptable student explanation is the following:

> ''$\Delta U = Q - W$. For the same ΔU, the system with more work done must have more Q input so process #1 is greater.''

Students' responses to this question are shown in Table II. The percentage of students answering the written diagnostic who gave the response $Q_1 > Q_2$ to Question #2—ignoring the explanations offered—ranged from 40% to 56%, and 34% of the interview subjects gave this response as well. However, if we examine the explanations provided by the students, a rather different picture emerges. Of the students answering the written diagnostic, only 11% gave an acceptable explanation based on the first law of thermodynamics. For this analysis, explanations such as the following were considered to be acceptable:

> "Q is greater for process 1 since $Q = U + W$ and W is greater for process 1."
>
> "Q is greater for process one because it does more work, the energy to do this work comes from the Q_{in}."

Among the students in the interview sample, 19% gave a correct answer with an acceptable explanation. If we add in students who answered that $Q_1 < Q_2$ but made only a simple sign error, the proportion with acceptable explanations rises to 15% of the 1999–2001 samples, and to 22% of the interview sample.

Application of the first law of thermodynamics is needed to answer Interview Question #6ii; 13% of the interviewed students were able to answer this question correctly with a correct explanation. Although the first law also is required to give a fully correct explanation for Interview Question #4, students were not pressed to provide such an explanation during the interviews. The 31% success rate observed in answers for that question might be interpreted as an extreme upper limit on the proportion of students in our samples who were able to make any practical use of the first law of thermodynamics. Otherwise, our data consistently show that no more than about one in five students in our samples emerged from the introductory physics course with an adequate grasp of the first law of thermodynamics. This conclusion is consistent with the findings reported in Ref. 20.

I. Difficulties regarding *P-V* diagrams

It is striking that only 38% of the students in the interview sample spontaneously attempted to use a P-V diagram to aid in responding to the questions. In particular for Interview Questions #6 and #8, one might expect that sketching a simple P-V diagram would be the quickest and easiest way to find a solution. Indeed, as we noted, several students recognized that they had initially made errors on these questions when prompted by the interviewer to draw a P-V diagram. However, it is clear that most of the students were not in the habit of employing P-V diagrams when considering thermodynamics problems that did not initially provide or refer to such a diagram.

A hint of the difficulties encountered by students in employing P-V diagrams is found in the results discussed in Sec. V D. Between a third to a half of all students were unable to give a correct answer with an acceptable explanation to Question #1, a problem in which the geometrical interpretation of work might be expected to yield a relatively straightforward answer.

In discussions regarding cyclic processes, heat engines, the second law of thermodynamics, etc., the association of the area contained within the closed curve representing that process with the net work done by the system often plays a central role. However, even after successfully drawing a P-V diagram representing a cyclic process (albeit one that often had numerous errors), nearly two thirds of the students in the interview sample remained convinced that the net work done in the process they had represented was zero.

Of the students who were interviewed, 22% were successful in drawing a correct P-V diagram for process #1. An additional 28% of the students drew a closed-curve diagram that represented the isothermal segment with a straight line (or, in one instance, with a line of incorrect curvature). Nearly all of the remainder—all but two students—drew a closed-curve path, but made one or more of a large assortment of errors (for example, curved or sloping lines representing isobaric or isochoric processes, missing processes, direction errors).

The overall impression gathered from observing students draw and interpret their P-V diagrams was that these diagrams represented a resource that was severely underutilized in their problem-solving arsenal. In noting the insights achieved by several of the students when drawing their diagrams, and the near-misses by some others who failed to carry the reasoning process through to conclusion, it seemed that many students might benefit from additional practice and experience with P-V diagrams. The potential instructional benefits of P-V diagrams will be discussed further in Sec. VIII.

VI. COMMENT REGARDING RELIABILITY OF THE DATA

There is evidence that our data might actually somewhat overstate the average level of knowledge in the full class population. The discussion regarding the characterization of the interview sample makes it clear that the performance of that group is likely to be higher than the class average. Moreover, all of the written diagnostic instruments were administered either to students who were attending (optional) recitation sections, or who were present in class on the last day of the semester. In previous investigations at ISU, we have found that the average exam scores of students attending recitation sections are somewhat higher than the scores of the full class population. For the present investigation, this factor was examined by administering a question on the final exam during the Spring 2001 semester.

The final exam question (see Fig. 4) involved two different processes connecting common initial and final states (similar to the questions on the written diagnostic). As can be seen from the breakdown of student responses ($N = 407$), only 33% gave the correct answer (C) that both the work done and the heat absorbed could be different in the two processes. 37% of the students believed that the work done must be the same, while 51% thought that the heat absorbed must be the same. On the written diagnostic questions in that same class ($N = 279$), 41% of the responses represented views consistent with the correct answer on the final exam question, that is, that $W_1 \neq W_2$ and that $Q_1 \neq Q_2$. This performance is significantly better ($p = 0.03$) than the proportion of correct responses on the final exam. Moreover, only 41% of the responses on the written diagnostic claimed that the heat absorbed had to be the same for the two processes, compared to 51% on the final exam. (Performance on the work question was similar.) The performance of the full class on the

A system consisting of a quantity of ideal gas is in equilibrium state "A." It is slowly heated and as it expands, its pressure varies. It ends up in equilibrium state "B." Now suppose that the same quantity of ideal gas again starts in state "A," but undergoes a **different** thermodynamic process (i.e., follows a different path on a *P-V* diagram), only to end up again in the same state "B" as before. Consider the net work done by the system and the net heat absorbed by the system during these two different processes. Which of these statements is true?

A. The work done may be different in the two processes, but the heat absorbed must be the same.

B. The work done must be the same in the two processes, but the heat absorbed may be different.

C. The work done may be different in the two processes, and the heat absorbed may be different in the two processes.

D. Both the work done and the heat absorbed must be the same in the two processes, but are not equal to zero.

E. Both the work done and the heat absorbed by the system must be equal to zero in both processes.

***Responses* ($N = 407$):**

(A) 28% (B) 14% (C) 33%

(D) 20% (E) 3% No response: 2%

Fig. 4. Question used on final exam of Spring 2001 course, with a breakdown of students' responses.

final exam was somewhat inferior to that shown by the population that responded to the written diagnostic.

VII. DISCUSSION

Decades of research have documented substantial learning difficulties among pre-university students with regard to heat, temperature and related concepts, but the possible implications of these findings for university students have been uncertain. The work of Loverude *et al.*[20] and of the present investigation, along with work in several different countries, all suggest that a large proportion of students in introductory university physics courses emerge with an insufficient functional understanding of the fundamental principles of thermodynamics to allow problem solving in unfamiliar contexts.

It is clear that a fundamental conceptual difficulty stems from the fact that heat transfer, work and internal energy are diverse forms of the same fundamental quantity, that is, "energy," and are all expressed in the same units. Many students simply do not understand why a distinction must be made among the three quantities, or indeed that such a distinction has any fundamental significance; one of the students in the Berger and Wiesner study called this distinction "hairsplitting" [*Haarspalterei*].[17] One of the subjects in our interview sample, when invited to explain what he found particularly confusing about the heat–work–energy relationship, offered this comment: "How is it acceptable for something called 'work' to have the same units as something called 'heat' and something called 'energy'?" Another student, when pressed to explain the distinction, said: "Maybe work and heat are kind of the same thing, just a transfer of energy in both cases."

Part of this confusion stems from the ubiquitous and well-documented difficulty of learning to make a clear conceptual distinction between a quantity and the *change* or *rate of change* in that same quantity, for example: velocity and acceleration,[30] magnetic flux and the *change* in magnetic flux,[31] potential and field.[32] Many students do not learn that heat transfer and work both represent changes in a system's internal energy, and that they therefore are not properties associated with a given state of a system, but rather with the transition between two such states. This problem is exacerbated by two other distinct difficulties, both well documented: (1) the use in colloquial speech of the word "heat" or "heat energy"[18,33] (and equivalents in other languages, for example *chaleur* [French][34] or *Wärme* [German][17]) to correspond to a concept that is actually closer to what physicists would call "internal energy;" and (2) the major conceptual difficulties faced by introductory students in mastering the work concept itself in a mechanics context, let alone within the less familiar context of thermodynamics.[20] Thus, introductory students are faced with the task of learning two distinct and somewhat subtle concepts—heat and work—when their everyday familiarity with those terms tends to lead them in precisely the wrong conceptual direction.

It is ironic that the students' apparent ability to comprehend the concepts of state and state function actually may contribute to their confusion regarding process-dependent quantities such as heat and work. Students learn to become well aware that there exist quantities that are independent of process, and that energy of a state is one of these quantities. Perhaps due to their already weak grasp of the concepts of heat and work, many students improperly transfer, in their own minds, various properties of state functions either to heat, or work or both.[35] Certainly, the fact that mechanics courses frequently highlight the path-independent work done by conservative forces may contribute to this confusion, as may extensive use of the equation $Q = mc\,\Delta T$ in calorimetry problems.

Heat engines, refrigerators and an analysis based on the second law of thermodynamics crucially depend on the nonzero net heat transfer to, and the net work done by, a thermodynamic system during a cyclic process. This concept was among the most poorly understood among the students in our interview sample, and the difficulty regarding cyclic processes was directly traceable to the confusion regarding the fundamental properties of heat and work.

Another area of confusion might be traced to the limiting approximations frequently—and often tacitly—invoked in making physical arguments regarding idealized processes. Experienced physicists automatically, even unconsciously, "fill in the dots" in their own minds when describing, for instance, an isothermal process and the meaning of a thermal reservoir. They have in mind the model involving very small (and therefore negligible) temperature excursions described by Chabay and Sherwood.[29] The overwhelming majority of textbook discussions treat this and similar idealized processes only very cursorily; our data suggest that for most students, such treatments are inadequate.

VIII. IMPLICATIONS FOR INSTRUCTIONAL STRATEGIES

Loverude *et al.* have pointed out that a crucial first step to improving student learning of thermodynamics concepts lies in solidifying the student's understanding of the concept of work in the more familiar context of mechanics, with particular attention to the distinction between positive and negative work.[20] Beyond that first step, it seems clear that little progress can be made without first guiding the student to a clear understanding that work in the thermodynamic sense can alter the internal energy of a system, and that heat or heat transfer in the context of thermodynamics refers to a *change* in some system's internal energy, or equivalently that it represents a quantity of energy that is being transferred from one system to another.

As discussed in Sec. V B, most students seem comfortable with the notion of internal energy as a quantity that is characteristic of the state of the system. One might try to take advantage of this understanding by eliciting from students the distinction between the amount of energy in a system at a given moment, and a change in that quantity brought about by various distinct methods, for example, through macroscopic forces leading to changes in a system's volume, and through alterations that occur due to temperature differences without changes in the system's volume.

The instructional utility of employing multiple representations of physics concepts has been demonstrated in numerous research investigations in physics education.[36] The results of our investigation suggest that significant learning dividends might result from additional instructional focus on the creation, interpretation, and manipulation of P-V diagrams representing various thermodynamic processes. In particular, students might benefit from practice in converting between a diagrammatic representation and a physical description of a given process, especially in the context of cyclic processes.

Our results demonstrate that certain fundamental concepts and idealizations often taken for granted by instructors are very troublesome for many students (for example, the relation between temperature and kinetic energy of an ideal gas, or the meaning of thermal reservoir). The recalcitrance of these difficulties suggests that it might be particularly useful to guide students to articulate these principles themselves, and to provide their own justifications for commonly used idealizations.

Loverude[21] has described the development and testing of curricular materials based on the research reported in Ref. 20.[37] Students' learning difficulties showed a strong tendency to persist even after research-based instruction, although significant improvements were demonstrated. His report of the initial testing of their curricular materials makes it clear that the task of improving student learning in thermodynamics is challenging indeed.

IX. CONCLUSION

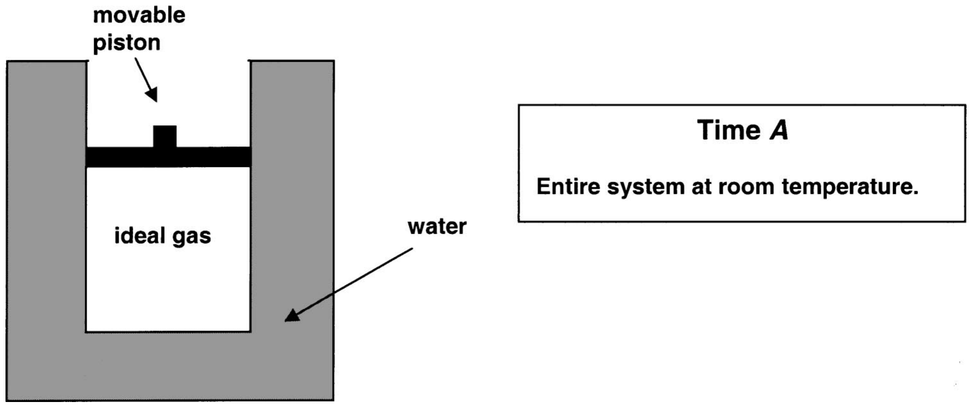
This investigation examined student learning of thermodynamics concepts in four separate offerings of the introductory calculus-based general physics course at a large public university over a period of three academic years. Several different course instructors, recitation instructors and textbooks were represented in these offerings. Results from the different population samples consistently showed that large proportions of the students in the courses emerged with a number of fundamental conceptual difficulties regarding the first law of thermodynamics, the definition and meaning of thermodynamic work, and the process-dependent nature of heat, including a belief that net heat absorbed and net work done by a system undergoing a cyclic process must be zero. Results of this investigation are in excellent agreement with those published in a recent study carried out at several other comparable institutions,[20] and are consistent with reports from several different European countries.[16–18,26,28,34] We conclude that substantial changes in instruction will be required if the level of students' mastery of thermodynamics concepts is to be significantly improved in introductory courses.

ACKNOWLEDGMENTS

I am very grateful for the cooperation, assistance, and interest of John M. Hauptman and Paul C. Canfield, and for the cooperation of the recitation instructors in Physics 222 at ISU. This material is based on work supported by the National Science Foundation under Grant No. DUE-9981140 (Co-Principal Investigator, T. J. Greenbowe), and Grant No. PHY-0406724. The thermodynamics curriculum project is a collaboration with the ISU Chemistry Education Research group directed by T. J. Greenbowe.

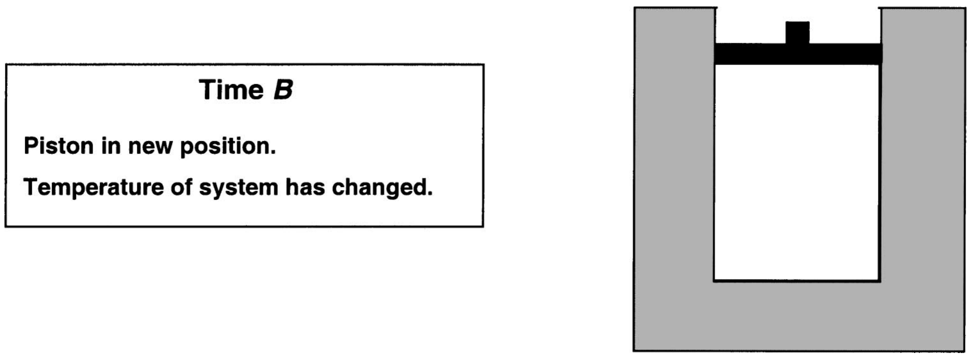
APPENDIX: INTERVIEW QUESTIONS

A fixed quantity of ideal gas is contained within a metal cylinder that is sealed with a movable, *frictionless*, insulating piston. (The piston can move up or down without the slightest resistance from friction, but no gas can enter or leave the cylinder. The piston is heavy, but there can be no heat transfer to or from the piston itself.) The cylinder is surrounded by a large container of water with high walls as shown. We are going to describe two separate processes, Process #1 and Process #2.



At initial time A, the gas, cylinder, and water have all been sitting in a room for a long period of time, and all of them are at room temperature.

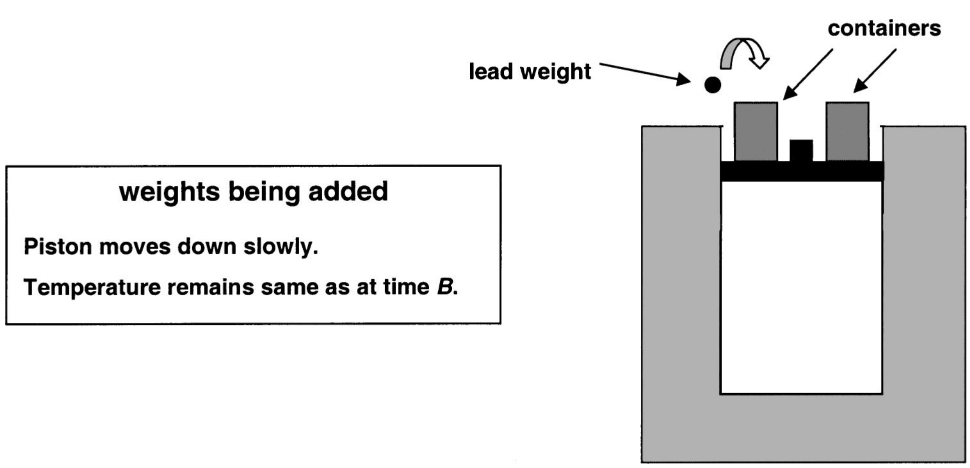
Step 1. We now begin Process #1: The water container is gradually heated, and the piston *very slowly* moves upward. At time B the heating of the water stops, and the piston stops moving when it is in the position shown in the diagram below:



Question #1: During the process that occurs from time A to time B, which of the following is true: (a) positive work is done *on* the gas *by* the environment, (b) positive work is done *by* the gas *on* the environment, (c) no *net* work is done on or by the gas.

Question #2: During the process that occurs from time A to time B, the gas absorbs x Joules of energy from the water. Which of the following is true: The total kinetic energy of all of the gas molecules (a) increases by more than x Joules; (b) increases by x Joules; (c) increases, but by less than x Joules; (d) remains unchanged; (e) decreases by less than x Joules; (f) decreases by x Joules; (g) decreases by more than x Joules.

Step 2. Now, empty containers are placed on top of the piston as shown. Small lead weights are gradually placed in the containers, one by one, and the piston is observed to move down slowly. While this happens, the temperature of the water is nearly unchanged, and the gas temperature remains practically *constant*. (That is, it remains at the temperature it reached at time B, after the water had been heated up.)



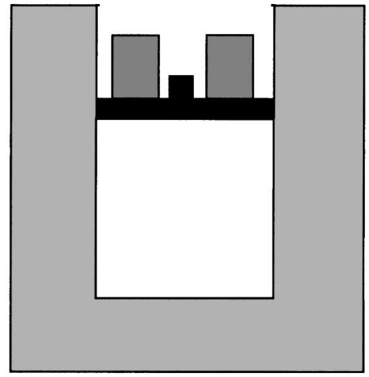
Step 3. At time C we stop adding lead weights to the container and the piston stops moving. (The weights that we have already added up until now are still in the containers.) The piston is now found to be at *exactly the same position it was at time* **A**.

Time C

Weights in containers.

Piston in same position as at time A.

Temperature same as at time B.



Question #3: During the process that occurs from time B to time C, does the total kinetic energy of all the gas molecules *increase, decrease*, or *remain unchanged*?

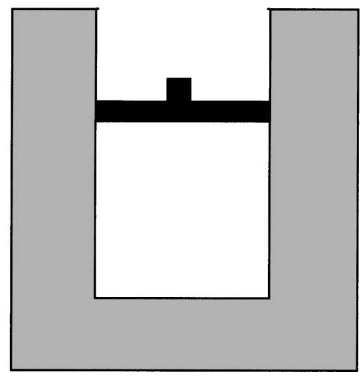
Question #4: During the process that occurs from time B to time C, is there *any* net energy flow between the gas and the water? If no, explain why not. If yes, is there a net flow of energy from gas to water, or from water to gas?

Step 4. Now, the piston is locked into place so it *cannot move*; the weights are removed from the piston. The system is left to sit in the room for many hours, and eventually the entire system cools back down to the same room temperature it had at time A. When this finally happens, it is time D.

Time D

Piston in same position as at time A.

Temperature same as at time A.



Question #5: During the process that occurs from time C to time D, the water absorbs y Joules of energy from the gas. Which of the following is true: The total kinetic energy of all of the gas molecules (a) increases by more than y Joules; (b) increases by y Joules; (c) increases, but by less than y Joules; (d) remains unchanged; (e) decreases, by less than y Joules; (f) decreases by y Joules; (g) decreases by more than y Joules.

Question #6: Consider ***the entire process*** from time A to time D. (i) Is the net work done *by* the gas on the environment during that process (a) greater than zero, (b) equal to zero, or (c) less than zero? (ii) Is the total heat transfer to the gas during that process (a) greater than zero, (b) equal to zero, or (c) less than zero?

Step 5. Now let us begin Process #2. The piston is unlocked so it is again free to move. We start from the same initial situation as shown at time A and D (i.e., same temperature and position of the piston). Just as before, we heat the water and watch as the piston rises. However, this time, we will heat the water for a *longer* period of time. As a result, the piston ends up *higher* than it was at time B.

Step 6. Now, weights are added to the piston and it begins to move down. (Temperature does not change during this process.) However, this time, *more* weights than before must be added to get the piston back to the position it had at time C.

Step 7. Again, the piston is locked and the weights are removed. After many hours, the system returns to the same temperature that it had at time A and time D (and the piston is in the same position as it was at those times). This final state occurs at time $\boldsymbol{E}$.

Question #7: Consider the total kinetic energy of all of the gas molecules at times A, D, and E; call those E_A, E_D, and E_E. Rank these in order of magnitude (greatest to least, using $>$ or $<$ signs). If two or more of these are equal, indicate that with an "$=$" sign.

Question #8: Consider the following positive quantities: $|Q_1|,|Q_2|,|W_1|,|W_2|$. These represent the absolute values of the total heat transfer to the gas during Process #1 and Process #2, and of the net work done by the gas during Processes #1 and #2. Rank these four quantities from largest to smallest. If two or more are equal, indicate with an "$=$" sign.

[a]Electronic mail: dem@iastate.edu

[1]A brief, annotated bibliography is in Lillian C. McDermott and Edward F. Redish, "Resource Letter: PER-1: Physics Education Research," Am. J. Phys. **67**, 755–767 (1999), Sec. IV A 4. A bibliography of more than 200 items can be found at ⟨http://www.physics.iastate.edu/per/index.html⟩.

[2]Michael Shayer and Hugh Wylam, "The development of the concepts of heat and temperature in 10–13 year olds," J. Res. Sci. Teach. **18**, 419–434 (1981).

[3]Sofia Kesidou and Reinders Duit, "Students' conceptions of the second law of thermodynamics—an interpretive study," J. Res. Sci. Teach. **30**, 85–106 (1993).

[4]Sofia Kesidou, Reinders Duit, and Shawn M. Glynn, "Conceptual development in physics: Students' understanding of heat," in *Learning Science in the Schools: Research Reforming Practice*, edited by Shawn M. Glynn

and Reinders Duit (Lawrence Erlbaum, Mahwah, NJ, 1995), pp. 179–198, and references therein.

[5]Gaalen Erickson and Andrée Tiberghien, ''Heat and temperature. Part A: An overview of pupils' ideas; Part B: The development of ideas with teaching,'' in *Children's Ideas in Science*, edited by Rosalind Driver, Edith Guesne, and Andrée Tiberghien (Open University Press, Milton Keynes, 1985), pp. 53–84, and references therein.

[6]Arnold B. Arons, *Teaching Introductory Physics* (Wiley, New York, 1997), Part I, p. 139; Randall D. Knight, *Five Easy Lessons: Strategies for Successful Physics Teaching* (Addison-Wesley, San Francisco, 2002), pp. 167–169.

[7]Shelley Yeo and Marjan Zadnik, ''Introductory thermal concept evaluation: Assessing students' understanding,'' Phys. Teach. **39**, 496–504 (2001); Paul G. Jasien and Graham E. Oberem, ''Understanding of elementary concepts in heat and temperature among college students and K–12 teachers,'' J. Chem. Educ. **79**, 889–895 (2002).

[8]T. R. Shultz and M. Coddington, ''Development of the concepts of energy conservation and entropy,'' J. Exp. Child Psych. **31**, 131–153 (1981); Reinders Duit and Sofia Kesidou, ''Students' understanding of basic ideas of the second law of thermodynamics,'' Res. Sci. Educ. **18**, 186–195 (1988); Ruth Ben-Zvi, ''Non-science oriented students and the second law of thermodynamics,'' Int. J. Sci. Educ. **21**, 1251–1267 (1999).

[9]J. F. Cullen, Jr., ''Concept learning and problem solving: The use of the entropy concept in college chemistry,'' Ph.D. dissertation, Cornell University, UMI, Ann Arbor, MI, 1983, UMI #8321833.

[10]M. F. Granville, ''Student misconceptions in thermodynamics,'' J. Chem. Educ. **62**, 847–848 (1985); H. Beall, ''Probing student misconceptions in thermodynamics with in-class writing,'' J. Chem. Educ. **71**, 1056–1057 (1994).

[11]Walter H. Kaper and Martin J. Goedhart, '''Forms of energy,' an intermediary language on the road to thermodynamics? Part II,'' Int. J. Sci. Educ. **24**, 119–137 (2002).

[12]P. H. van Roon, H. F. van Sprang, and A. H. Verdonk, '''Work' and 'heat': on the road towards thermodynamics,'' Int. J. Sci. Educ. **16**, 131–144 (1994).

[13]A. C. Banerjee, ''Teaching chemical equilibrium and thermodynamics to undergraduate general chemistry classes,'' J. Chem. Educ. **72**, 879–887 (1995); Roger Barlet and Géraldine Mastrot, ''L'algorithmisation-refuge, obstacle à la conceptualisation; L'exemple de la thermochimie en 1[er] cycle universitaire,'' Didaskalia **17**, 123–159 (2000).

[14]P. L. Thomas and R. W. Schwenz, ''College physical chemistry students' conceptions of equilibrium and fundamental thermodynamics,'' J. Res. Sci. Teach. **35**, 1151–1160 (1998); Peter Lynn Thomas, ''Student conceptions of equilibrium and fundamental thermodynamic concepts in college physical chemistry,'' Ph.D. dissertation, University of Northern Colorado, UMI, Ann Arbor, MI, 1997, UMI #9729078.

[15]Thomas J. Greenbowe and David E. Meltzer, ''Student learning of thermochemical concepts in the context of solution calorimetry,'' Int. J. Sci. Educ. **25**, 779–800 (2003).

[16]S. Rozier and L. Viennot, ''Students' reasonings in thermodynamics,'' Int. J. Sci. Educ. **13**, 159–170 (1991); Laurence Viennot, *Raisonner en Physique* (De Boeck Université, Brussels, 1996), pp. 118–123; *Reasoning in Physics* (Kluwer, Dordrecht, 2001), pp. 105–110.

[17]R. Berger and H. Wiesner, ''Zum Verständnis grundlegender Begriffe und Phänomene der Thermodynamik bei Studierenden,'' in *Deutsche Physikalische Gesellschaft, Fachverband Didaktik der Physik: Didaktik der Physik* (Technische Universität Berlin, Institut für Fachdidaktik Physik und Lehrerbildung, Berlin, 1997), pp. 736–741; also available on CD in *DPG 1997: Didaktik, Umwelt*; *Tagungsberichte der Fachgremien* (2N Hochschulkommunikation, Holtzheim, 1997), ISBN 3-931253-06-6.

[18]J. W. Warren, ''The teaching of the concept of heat,'' Phys. Educ. **7**, 41–44 (1972).

[19]David B. Pushkin, ''The influence of a computer-interfaced calorimetry demonstration on general physics students' conceptual views of entropy and their metaphoric explanations of the second law of thermodynamics,'' Ph.D. dissertation, Pennsylvania State University, UMI, Ann Arbor, MI, 1995, UMI #9612815.

[20]Michael E. Loverude, Christian H. Kautz, and Paula R. L. Heron, ''Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas,'' Am. J. Phys. **70**, 137–148 (2002).

[21]Michael Eric Loverude, ''Investigation of student understanding of hydrostatics and thermal physics and of the underlying concepts from mechanics,'' Ph.D. dissertation, Department of Physics, University of Washington, UMI, Ann Arbor, MI, 1999, UMI #9937617.

[22]David E. Meltzer, ''Student reasoning regarding work, heat, and the first law of thermodynamics in an introductory physics course,'' in *Proceedings of the 2001 Physics Education Research Conference*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (PERC Publishing, Rochester, NY, 2001), pp. 107–110.

[23]*1999*: David Halliday, Robert Resnick, and Jearl Walker, *Fundamentals of Physics, Extended* (Wiley, New York, 1997), 5th ed.; *2000*: Raymond A. Serway, *Principles of Physics* [custom printing] (Saunders, Fort Worth, 1998), 2nd ed.; *2001*: David Halliday, Robert Resnick, and Jearl Walker, *Fundamentals of Physics, Extended* (Wiley, New York, 2001), 6th ed.

[24]Ronald Lane Reese, *University Physics* (Brooks/Cole, Pacific Grove, 2000).

[25]See Ref. 16. A similar argument in the context of an irreversible adiabatic expansion was advanced by some of the German university students in the investigation of Berger and Wiesner (Ref. 17).

[26]H. Goldring and J. Osborne, ''Students' difficulties with energy and related concepts,'' Phys. Educ. **29**, 26–31 (1994).

[27]The conclusion of Kaper and Goedhart (Ref. 11) that students treat heat as a state function was based on interpretation of remarks made by several students during tape-recorded conversations occurring in tutorial sessions.

[28]Ursula Manthei and Paul Täubert, ''Zustandsgröße und Prozessgröße erläutert am Beispiel Energie—Arbeit, Wärme, Strahlung,'' Phys. Schule **19**, 307–317 (1981).

[29]Ruth W. Chabay and Bruce A. Sherwood, *Matter & Interactions I: Modern Mechanics* (Wiley, New York, 2002), p. 398.

[30]David E. Trowbridge and Lillian C. McDermott, ''Investigation of student understanding of the concept of acceleration in one dimension,'' Am. J. Phys. **49**, 242–253 (1981).

[31]Leith Dwyer Allen, ''An investigation into student understanding of magnetic induction,'' Ph.D. dissertation, The Ohio State University, UMI, Ann Arbor, MI, 2001, UMI #3011018, pp. 305–306.

[32]Rhett Allain, ''Investigating the relationship between student difficulties with the concept of electric potential and the concept of rate of change,'' Ph.D. dissertation, North Carolina State University, UMI, Ann Arbor, MI, 2001, UMI #3030022, Chaps. 2 and 5, and references therein.

[33]Mark W. Zemansky, ''The use and misuse of the word 'heat' in physics teaching,'' Phys. Teach. **8**, 295–300 (1970).

[34]A. Tiberghien and G. Delacôte, ''Résultats préliminaires sur la conception de la chaleur,'' in *Physics Teaching in Schools: Proceedings of the 5th Seminar of GIREP*, edited by G. Delacôte (Taylor & Francis, Ltd., London, 1978), pp. 275–282.

[35]In Ref. 22 it is shown that among students in 2000 and 2001 who responded to Question #2 by asserting that $Q_1 = Q_2$, those students who answered Question #1 correctly (that is, by responding $W_1 > W_2$) were more likely to support their incorrect answer about heat with an explicit argument that heat was independent of process, in comparison to students who had given an incorrect answer to the work question. (The latter group was more likely to offer some other explanation, or no explanation for their answer about heat.) This readiness to offer an incorrect explanation suggests the possibility that partially correct understanding (that is, regarding work) may actually be associated with an increase in the confidence with which students hold to an incorrect concept regarding heat.

[36]Lillian C. McDermott, ''A view from physics,'' in *Toward a Scientific Practice of Science Education*, edited by M. Gardner, J. G. Greeno, F. Reif, A. H. Schoenfeld, A. diSessa, and E. Stage (Lawrence Erlbaum, Hillsdale, NJ, 1990), pp. 3–30; Alan Van Heuvelen, ''Learning to think like a physicist: A review of research-based instructional strategies,'' Am. J. Phys. **59**, 891–897 (1991); Alan Van Heuvelen, ''Overview, Case Study Physics,'' *ibid.* **59**, 898–907 (1991); Ronald K. Thornton and David R. Sokoloff, ''Assessing student learning of Newton's laws: the Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula,'' *ibid.* **66**, 338–352 (1998); Alan Van Heuvelen and Xueli Zou, ''Multiple representations of work-energy processes,'' *ibid.* **69**, 184–194 (2001).

[37]Lillian C. McDermott, Peter S. Shaffer, and the Physics Education Group, *Tutorials in Introductory Physics* (Prentice–Hall, Upper Saddle River, NJ, 2002), pp. 231–235; *Tutorials in Introductory Physics, Homework* (Prentice-Hall, Upper Saddle River, NJ, 2002), pp. 173–174.

Investigation of Student Reasoning Regarding Concepts in Thermal Physics

David E. Meltzer

Decades of research have documented substantial learning difficulties among pre-university students with regard to heat, temperature, and related concepts.[1] However, it has not been clear what implications these findings might have with regard to the learning of thermodynamics. Studies reported in several European countries in recent years have indicated significant confusion among university students regarding fundamental concepts in thermal physics.[2] The recent investigation of Loverude *et al*.[3] strongly suggested that a large proportion of students in introductory university physics courses emerge with an understanding of the fundamental principles of thermodynamics that is insufficient to allow problem solving in unfamiliar contexts. In related work, the Iowa State University Physics Education Research Group has been engaged since 1999 in a research and curriculum development project aimed at improving thermodynamics instruction in the introductory university physics course. In this short report I will summarize some of the initial findings of our ongoing investigation into students' reasoning regarding concepts in thermodynamics.[4]

Our data for this initial phase of the investigation were collected during 1999-2002 and were in two primary forms: (1) a written free-response quiz that was administered to a total of 653 students in three separate offerings of the calculus-based introductory physics course; (2) one-on-one interviews that were conducted with 32 student volunteers who were enrolled in a fourth offering of the same course. All testing and interviewing was done after students had completed their study of the relevant topics. Results of all the various data sources were quite consistent with each other.

We found that students' understanding of process-dependent quantities was seriously flawed, as substantial numbers of students persistently ascribed state-function properties to both work *and* heat. Although most students seemed to acquire a reasonable grasp of the state-function concept in the context of internal energy, it was found that there was a widespread and persistent tendency to improperly over-generalize this concept to apply to both work and heat. This confusion was associated with a strong tendency to believe that the net work done and the net heat absorbed by a system undergoing a cyclic process are both zero.

The written quiz consisted of a *P-V* diagram on which curving lines represented two separate expansion processes involving a fixed quantity of ideal gas. The initial and final states of the two processes were identical, but the areas under the curve differed in the two cases. Students were asked to compare the amount of work done by the system during the two processes, and also the amount of heat transfer to the system during the same two processes. About 30% of all students asserted that the work done would be equal in the two cases, although the areas under the curve were clearly different. Similarly, 38% of all students claimed that the heat transfer to the system would be the same in both processes, although a straightforward application of the first law of thermodynamics shows that the heat transfer must be different in the two cases. (This incorrect response regarding heat was almost equally popular among students who gave the *correct* answer to the work question, as it was among those who claimed that the work done was equal in the two processes.)

During the interviews, students were shown diagrams portraying a three-step cyclic process involving a cylinder containing a quantity of ideal gas. The diagrams showed an isobaric expansion followed by an isothermal compression, followed finally by a constant-volume cooling. (The net work done by the system and the net heat transfer to the system during the complete cycle were negative.) After slowly and methodically working through and discussing this process (the typical interview lasted over one hour), 75% of the students asserted with great confidence that either the net heat transfer to the system during the complete cycle, the net work done by the system during the cycle, or both of those quantities, would have to be equal to zero. The interviews also disclosed unanticipated levels of confusion regarding the definition of thermodynamic work, as well as difficulties in recognizing the existence of heat transfer during isothermal processes involving volume changes.

Consistent results over several years of observations involving both written quizzes and oral interviews enabled us to make a high-confidence estimate that approximately 80% of students in the introductory calculus-based physics course emerged with only a very weak ability to apply the first law of thermodynamics to solving problems in unfamiliar contexts. This result was consistent with findings of Loverude *et al*.

Although it is not entirely clear *how* students arrive at their ideas regarding thermodynamics, some of the more widely shared ideas seem to have an understandable basis. It seems that a fundamental conceptual difficulty is associated with the fact that heat transfer, work, and internal energy are all expressed in the same units, and all represent either energy or *transfers of* energy. Many students simply do not understand why a distinction must be made among the three quantities, or indeed that such a distinction has any fundamental significance. One of the subjects in our interview sample, when invited to explain what he found particularly confusing about the heat-work-energy relationship, offered this comment: "How is it acceptable for something called 'work' to have the same units as something called 'heat' and something called 'energy'?"

Part of this confusion stems from the ubiquitous and well-documented difficulty students have in making a clear conceptual distinction between a quantity and the *change* or *rate of change* of that same quantity (for example, that between velocity and acceleration).[5] Many students do not learn that heat transfer and work both represent *changes* in a system's internal energy, and that they therefore are not properties associated with a given state of a system but rather with the transition between two such states. This problem is exacerbated by the use in colloquial speech of the terms "heat" or "heat energy" to correspond to a concept that is actually closer to what physicists would call "internal energy". However, our findings corroborated those of Loverude *et al*.[3] that an even more significant difficulty was that related to mastering the work concept in a *mechanics* context, let alone within the less familiar context of thermodynamics. Significant difficulties in understanding work persisted from students' studies of mechanics, and hampered their ability to master the related ideas in the context of thermodynamics.

Students do learn well that there exist quantities that are independent of process, and that (internal) energy of a system is one of these quantities. Perhaps due to their already weak grasp of the concepts of heat and work, many students improperly transfer, in their own minds, various properties of state functions either to heat, or work, or both. Certainly, the fact that mechanics courses frequently highlight the path-independent work done by conservative forces may contribute to this confusion, as may extensive use of the equation $Q = mc\Delta T$ in calorimetry problems.

Another area of confusion might be traced to the limiting approximations frequently – and often tacitly – invoked regarding idealized processes. Experienced physicists automatically "fill in the dots" when describing, for instance, an isothermal process and the meaning of a thermal reservoir. The overwhelming majority of textbook discussions treat these and similar idealized processes only very cursorily; our data suggest that for most students, such treatments are inadequate.

Implications for Instructional Strategies

Loverude *et al*. have pointed out that a crucial first step to improving student learning of thermodynamics concepts lies in solidifying the student's understanding of the concept of work in the more familiar context of mechanics, with particular attention to the distinction between positive and negative work.[3] Beyond that, it seems that little progress can be made without first guiding the student to a clear understanding that work in the thermodynamic sense can *alter* the internal energy of a system, and that heat or heat transfer in the context of thermodynamics refers to a *change* in some system's internal energy, or equivalently that it represents a quantity of energy that is being transferred from one system to another.

The instructional utility of employing multiple representations of physics concepts has been demonstrated.[6] The results of our study suggest that significant learning dividends might result from additional instructional focus on the creation, interpretation, and manipulation of *P-V* diagrams representing various thermodynamic processes. In particular, students might benefit from practice in converting between a diagrammatic representation and a physical description of a given process, especially in the context of cyclic processes.

Our results demonstrate that certain fundamental concepts and idealizations often taken for granted by instructors are very troublesome for many students (for example, the relation between temperature and kinetic energy of an ideal gas, or the meaning of thermal reservoir). The recalcitrance of these difficulties suggests that it might be particularly useful to guide students to articulate these principles themselves, and to provide their own justifications for commonly used idealizations.

It is worth noting another one of our observations that corroborated reports from other researchers. We found that students often used microscopic arguments both as a basis and as a justification for incorrect reasoning regarding thermodynamic phenomena. (This is identical to a finding reported in Ref. 3, and in other references cited in both Refs. 3 and 4.) The extent to which this faulty student reasoning was actually initiated or catalyzed by instruction involving microscopic concepts is uncertain. However, our research serves as a caution that merely incorporating a strong instructional emphasis on the microscopic, molecular viewpoint in thermal physics is unlikely, *in itself*, to dramatically impact students' understanding. Indeed, our ongoing research indicates that many key concepts emphasized in a microscopic approach are very challenging even for physics majors in their third and fourth years of study.[7]

Acknowledgments

This material is based on work supported by the National Science Foundation under Grant Number DUE-9981140 (Co-Principal Investigator: T. J. Greenbowe), and Grant Number PHY-0406724. The thermodynamics curriculum project is a collaboration with the ISU Chemistry Education Research group directed by T. J. Greenbowe.

References

1. See, for instance, Michael Shayer and Hugh Wylam, "The development of the concepts of heat and temperature in 10–13 year olds," J. Res. Sci. Teach. **18**, 419–434 (1981); Sofia Kesidou, Reinders Duit, and Shawn M. Glynn, "Conceptual development in physics: Students' understanding of heat," in *Learning Science in the Schools: Research Reforming Practice*, edited by Shawn M. Glynn and Reinders Duit (Lawrence Erlbaum, Mahwah, NJ, 1995), pp. 179–198, and references therein.
2. For instance, J. W. Warren, "The teaching of the concept of heat," Phys. Educ. **7**, 41–44 (1972); S. Rozier and L. Viennot, "Students' reasonings in thermodynamics," Int. J. Sci. Educ. **13**, 159–170 (1991); R. Berger and H. Wiesner, "Zum Verständnis grundlegender Begriffe und Phänomene der Thermodynamik bei Studierenden," in *Deutsche Physikalische Gesellschaft, Fachverband Didaktik der Physik: Didaktik der Physik* (Technische Universität Berlin, Institut für Fachdidaktik Physik und Lehrerbildung, Berlin, 1997), pp. 736–741; also available on CD in *DPG 1997: Didaktik, Umwelt; Tagungsberichte der Fachgremien* (2N Hochschulkommunikation, Holtzheim, 1997), ISBN 3-931253-06-6.
3. Michael E. Loverude, Christian H. Kautz, and Paula R. L. Heron, "Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas," Am. J. Phys. **70**, 137–148 (2002).
4. A full report may be found in: David E. Meltzer, "Investigation of students' reasoning regarding heat, work, and the first law of thermodynamics in an introductory calculus-based general physics course," Am. J. Phys. **72**, 1432–1446 (2004).
5. David E. Trowbridge and Lillian C. McDermott, "Investigation of student understanding of the concept of acceleration in one dimension," Am. J. Phys. **49**, 242–253 (1981).
6. See references cited in David E. Meltzer, "Relation between students' problem-solving performance and representational format," Am. J. Phys., **73**, 463-478 (2005).
7. David Meltzer and Warren M. Christensen, "Development of student reasoning in an upper-level thermal physics course," AAPT Announcer **34(4)**, 97 (2004).

David E. Meltzer is Assistant Professor of Physics and Astronomy at Iowa State University.

Student Learning In Upper-Level Thermal Physics: Comparisons And Contrasts With Students In Introductory Courses

David E. Meltzer

Department of Physics and Astronomy
Iowa State University

Abstract. We found that students in an upper-level thermal physics course were in general quicker than introductory students at grasping and applying fundamental concepts. Upper-level students seemed, in general, more receptive to employing qualitative reasoning using multiple representations, and capable of using it more effectively than introductory students. In addition, upper-level students were better able to utilize guided-inquiry curricular materials in the sense of reasoning with greater depth and grasping more subtle issues. However, although the overall level of preparation and ability was higher in the upper-level course, the broad range of preparation represented among the students presented various practical challenges to implementing active-learning instructional strategies. Moreover, even quite capable upper-level students would falter unexpectedly and unpredictably on various conceptual difficulties that are common among introductory students. The unpredictable and inconsistent nature of this effect demonstrated that instructors must always be prepared to detect and address such difficulties in upper-level courses.

INTRODUCTION

We have been engaged in an ongoing investigation of student learning of thermal physics in introductory courses [1-3]. In the course of this project, we have probed students' reasoning regarding heat, work, the first law of thermodynamics, calorimetry, and related topics. Based on this work, we have developed and tested preliminary versions of guided-inquiry curricular materials.

During Fall semester 2003, I taught a junior-level thermal physics course targeted at physics majors and other advanced students. In this course, many instructional methods were used that are often characterized as "interactive engagement" [4] or "active learning" [5]. Fourteen students were enrolled, mostly junior and senior physics majors along with several students majoring in chemistry or engineering. This course provided an opportunity to compare introductory and advanced students regarding learning of similar topics, using some identical curricular materials and methods. In this paper, I will discuss some of the main features of this experience.

METHODOLOGICAL ISSUES IN UPPER-LEVEL CLASSES

Students taking upper-level physics courses are certainly not representative of the overall student population enrolled in introductory general physics courses. Only a small percentage of students in the introductory course have a specific interest in physics as a major field, and most would be far more likely to take upper-level engineering courses than to enroll in advanced-level physics courses. For this reason, we must assume that observations regarding learning and transfer in advanced-level physics courses are characteristic only of a highly selected subsample of students enrolled in a typical introductory course. It is also important to remember that small class sizes (common in upper-level courses) are associated with a relatively high probability that any *one* particular class will not be fully representative of other, similar classes [6].

When evaluating students' performance in upper-level courses, two distinct factors come into play: (1) students' knowledge of material previously covered in introductory courses, and (2) students' *learning* of new

Published in *2004 Physics Education Research Conference [Sacramento, California, 4-5 August 2004]*, edited by Jeffrey Marx, Paula R. L. Heron, and Scott Franklin [American Institute of Physics Conference Proceedings **790**, 31-34 (2005)].

material during the advanced course. In order to make some determination of students' knowledge of material learned in previous courses, it is essential to administer pretests on the relevant material to the students before those topics are covered in the advanced course. This poses a logistical challenge, because extensive pretesting ordinarily requires substantial amounts of class time. Because pretests are administered before instruction on the relevant topics, they can not be graded for course credit. Moreover, even when pretest data are acquired, an important dimension of student learning is missed, that is, the degree to which students may have increased their *readiness* to learn new topics as a result of their experiences in the introductory courses.

Although students may give incorrect answers or inadequate explanations on pretest questions related to specific topics, it may well be that their exposure to those topics in their previous courses had provided them with a basis for rapidly and effectively assimilating a second round of instruction on those same topics, as is often provided in advanced courses. Determining the students' state of "readiness" for new learning is an extremely challenging problem for researchers who develop and administer diagnostic pretests. Ordinary tests assess students' knowledge at a particular point in time and do not provide a measure of the rate at which such knowledge might be *changing*, nor of its *susceptibility* to change.

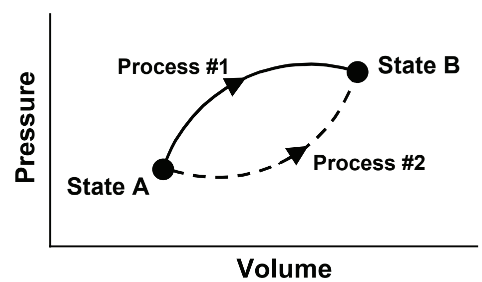
When considering "transfer" of learning from introductory courses, we are interested in two distinct (but hard to separate) factors: (1) *application* of knowledge previously learned, and (2) *synthesis* of new concepts based on knowledge elements learned or learning skills developed in previous courses. In the case of "application" of knowledge, we might try to determine the degree to which students can apply knowledge of concepts and techniques acquired in previous courses to the solution of more complex problems that make use of those same concepts or techniques. However, we have an equal or greater interest in the degree to which students can learn *new* concepts or techniques in the advanced course, and effectively apply those new concepts in problem solving. One assumes that an important product of instruction in introductory courses is the preparation of students for learning of new concepts in more advanced courses. However, it is probably much more difficult to determine students' ability to *learn* new concepts, than it is to determine their ability to *apply* previously learned concepts in new contexts.

STUDENTS' INITIAL KNOWLEDGE

On the first day of class, a small set of diagnostic questions related to calorimetry and the first law of thermodynamics was administered to provide information regarding students' initial knowledge. Overall performance on these questions was superior to the average post-instruction performance of students in the introductory physics courses reported in Refs. 1-3, although a broad range of knowledge levels was found.

During the present (Fall 2004) semester, I am again teaching the thermal physics course. On the first day of class, a larger set of diagnostic questions was administered to the students. Two (out of a total of 15) of these questions are shown in Fig. 1.

This *P-V* diagram represents a system consisting of a fixed amount of ideal gas that undergoes two ***different*** processes in going from state A to state B:



[In these questions, ***W*** represents the work done ***by*** the system during a process; ***Q*** represents the heat ***absorbed*** by the system during a process.]

1. Is W for Process #1 ***greater than***, ***less than***, or ***equal to*** that for Process #2? Explain.
2. Is Q for Process #1 ***greater than***, ***less than***, or ***equal to*** that for Process #2? Please explain your answer.

FIGURE 1. Two of the questions posed to students in both introductory and upper-level physics courses. Answers: (1) *greater than*; (2) *greater than*.

Most of the questions in this second diagnostic set had been administered (after instruction was completed) to students in the introductory calculus-based general physics course during the investigations reported in Refs. 1-3.

Among the 21 upper-level students responding to these questions, a wide range of initial knowledge levels was evident. Some students showed good ability to apply first-law concepts, while others showed little or none. On some questions, average performance was clearly superior to the post-instruction performance of students in the introductory course, while on other questions performance was virtually indistinguishable from that of students in the lower-level course.

On Question #1 shown in Figure 1 (the "work" question), about one quarter of students in both introductory and upper-level courses answered incorrectly that the work done by the system in Process #1 would be the same as that done in Process #2. By contrast, on Question #2 (the "heat" question), 38% of students in the upper-level course gave a correct or nearly correct answer with an acceptable explanation, compared to only 15-20% of students *after instruction* in the introductory courses. On questions related to cyclic processes, thermal reservoirs, and isothermal processes, performance of students in the upper-level course was comparable to the post-instruction performance of a self-selected sample of interview volunteers from the introductory course whose course grades were well above the class average [2].

COMPARISONS AND CONTRASTS WITH INTRODUCTORY STUDENTS

Students in the upper-level course demonstrated a number of important learning skills that were significantly better developed than among students in the introductory course. At the same time, even very able students in the advanced course periodically demonstrated a vulnerability to learning difficulties similar or identical to those found among students in the introductory course.

Upper-Level Students Demonstrated Superior Learning Skills

Learning skills displayed by upper-level students were superior in a number of respects to those of students in the introductory course. For example, they demonstrated an ability to make use of qualitative reasoning, multiple representations, and guided-inquiry curricular materials that was generally beyond that of the introductory students.

In covering similar material, upper-level students were quicker to generalize over specific contexts with a unifying concept. By contrast, introductory students tended to focus on pattern matching, recognizing commonalities among different problems without necessarily extracting a unifying physical theme. Despite having superior mathematical skills, upper-level students relied *less* on purely mathematical calculations and arguments than did introductory students in working identical problems. They were less likely to simply point to an equation as an explanation, and more likely to use arguments based on proportional reasoning.

Upper-level students found it easier than did introductory students to interpret the meaning of diagrams, bar charts, and other graphical material, even in novel contexts. They were more comfortable in making use of multiple representations (verbal, diagrammatic, etc.) to express their own thinking, and they showed less reliance on purely mathematical forms of reasoning. Even upper-level students with relatively less preparation demonstrated facility with multiple representations.

Upper-level students made effective use of guided-inquiry worksheets originally developed for use with introductory students. Typically, upper-level students worked through problems faster and more thoroughly, and required less guidance from instructors, than did students in the introductory course. Moreover, they were less likely to become bogged down in problem minutiae such as instructions or descriptions of apparatus, and they showed less confusion in interpreting instructions. These students worked well in groups, usually had productive discussions, and helped each other effectively. They showed a willingness to devote extra time to the resolution of confusing points.

Common Reasoning Difficulties Were Shared by Upper-Level Students

Even students receiving the highest overall grades would sometimes encounter conceptual difficulties that were the same as or similar to those observed among introductory students. The appearance of these learning difficulties among the upper-level students was intermittent and unpredictable, but recurrent (although the *same* difficulties did not generally recur). *Providing* they were addressed directly, these difficulties appeared to be resolved efficiently and thoroughly with few observable remnants.

Notable examples of conceptual difficulties encountered included the following: (1) Several students had substantial difficulty in applying the state-function property of entropy to conclude that ΔS would be equal for a free-expansion process and an isothermal process sharing identical initial and final states. In general, invoking state-function properties in contexts involving *entropy* seemed to be more difficult for most students than in the context of internal *energy* [7]. (A similar finding was recently reported by Kautz [8].) (2) Many students were slow in learning to compare engine and refrigerator efficiencies to the Carnot efficiency in order to check compliance with the second law. In addition, there were difficulties in making the correct identification of heat and work inflows *to* and outflows *from* the system in these problems. (3) When working through a guided-inquiry worksheet using diagrams that depicted a cyclic process, some students initially concluded that net work done by the system during the process had to be zero. Similar difficulties had been prevalent among students in the introductory course [2] and were evident among the upper-level students on the first-day pretest. (4) Many students displayed considerable difficulty in distinguishing between systems that had identical temperatures but different internal energies, and vice versa. (This is related to the classical confusion between heat and temperature, long recognized as a recurring learning difficulty in teaching thermodynamics to diverse student populations.)

Challenges and Difficulties

Consistent with observations made among students in introductory courses, both highly favorable and highly unfavorable reactions toward interactive-engagement techniques were displayed by upper-level students. The 10-15% unfavorable rating on evaluations matched that found in the introductory algebra-based course. Use of guided-inquiry worksheets during class (instead of in a separate recitation section) created logistical difficulties due to the broad range of speeds with which students worked. Insufficient pretesting and lack of previous relevant research made optimal course planning difficult.

SUMMARY

Students' performance on qualitative and quantitative problems throughout the course (on homework, quizzes, and exams) provided substantial evidence of effective learning in the context of this "active-learning" environment. There was some evidence of transfer of learning from previous courses, in that students seemed able to make use of (sometimes fragmentary) ideas acquired during previous instruction in the process of synthesizing an improved overall grasp of the subject. However, there was also substantial evidence suggesting that instructors must be attentive to sudden and unpredictable appearances of standard learning difficulties even among upper-level students.

When presented with unfamiliar concepts, upper-level students appeared to learn and apply them more efficiently than did introductory students. Experience with other advanced courses and a willingness to do substantial amounts of homework apparently contributed to significant learning gains. Upper-level students demonstrated, on the average, greater motivation; however, it is difficult to separate motivational factors from skill factors with respect to their relative significance in the production of observed learning gains.

The fundamental problem regarding analysis of transfer in the context of upper-level courses is the difficulty in answering this question: When learning is observed in upper-level courses, does it represent (a) transfer of *knowledge* acquired in introductory courses, (b) application of *learning skills* acquired in introductory courses, or (c) knowledge and/or skills possessed by the student all along, perhaps even before beginning introductory courses? (Or, perhaps, all three?) It is likely that extensive longitudinal investigation with diverse courses and student populations will be required to apportion the proper weights among the various relevant factors.

ACKNOWLEDGMENTS

Ngoc-Loan Nguyen has made significant contributions to this report, based on his experiences as a teaching assistant in both the introductory course and the upper-level thermal physics course. This work was supported in part by NSF DUE-#9981140 (T. J. Greenbowe, Co-P.I.) and NSF PHYS-#0406724.

REFERENCES

1. Meltzer, D. E., "Student Reasoning Regarding Work, Heat, and the First Law of Thermodynamics in an Introductory Physics Course," in *Proceedings of the 2001 Physics Education Research Conference*, edited by S. Franklin, J. Marx, and K. Cummings, Rochester, NY: PERC Publishing, 2001, pp. 107-110.
2. Meltzer, D. E., *Am. J. Phys*. **72**, 1432-1446 (2004).
3. Christensen, W. M., Nguyen, N.-L. P., and Meltzer, D. E., *AAPT Announcer* **34 (2)**, 138 (2004).
4. Hake, R. R., *Am. J. Phys*. **66**, 64-74 (1998).
5. Knight, R. D., *Five Easy Lessons*, San Francisco: Addison Wesley, 2002, pp. 46-62.
6. This is because the variance (about the population mean) for mean values of small subsamples drawn from a large population is relatively large, compared to the variance for larger subsamples.
7. Ref. 2, Sec. V B.
8. Kautz, C., *AAPT Announcer* **34 (2)**, 120 (2004).

INT. J. SCI. EDUC., 2003, VOL. 25, NO. 7, 779–800

RESEARCH REPORT

Student learning of thermochemical concepts in the context of solution calorimetry

Thomas J. Greenbowe, Department of Chemistry, Iowa State University of Science and Technology, Ames, IA 50011, USA; e-mail: tgreenbo@iastate.edu; and David E. Meltzer, Department of Physics and Astronomy, Iowa State University of Science and Technology, Ames, IA 50011, USA; e-mail: dem@iastate.edu

Student understanding of heat and thermal phenomena has been the subject of considerable investigation in the science education literature. Published studies have reported student conceptions on a variety of advanced topics, but calorimetry – one of the more elementary applications of thermochemical concepts – has apparently received little attention from science education researchers. Here we report a detailed analysis of student performance on solution calorimetry problems in an introductory university chemistry class. We include data both from written classroom exams for 207 students, and from an extensive longitudinal interview series with a single subject who was herself part of that larger class. Our findings reveal a number of learning difficulties, most of which appear to originate from failure to understand that net increases and decreases in bond energies during aqueous chemical reactions result in energy transfers out of and into, respectively, the total mass of the resultant solution.

Introduction

Students' understanding of heat and thermal phenomena has been the subject of considerable investigation in the science education literature. Most of this investigation has been in the context of pre-university students, both at the secondary and pre-secondary levels (e.g., Johnstone *et al.* 1977, Stavy and Berkovitz 1980, Shayer and Wylam 1981, Tiberghien 1983, 1985, Erickson, 1985, Linn and Songer 1991, Kesidou and Duit 1993, Kesidou *et al.* 1995, Lewis and Linn 1994, Harrison *et al.* 1999, Ben-Zvi 1999, Barker and Millar 2000). A few studies have focused on thermodynamics in the context of university-level physics instruction (e.g., Rozier and Viennot 1991, Loverude *et al.* 2002). There have also been a handful of investigations into student learning of chemical thermodynamics at the university level (Granville 1985, Beall 1994, Van Roon *et al.* 1994, Banerjee 1995, Thomas 1997, Thomas and Schwenz 1998). These investigations have reported on student conceptions regarding the first and second laws of thermodynamics, entropy and free energy, spontaneous processes, etc. However, calorimetry – one of the more elementary applications of thermochemical concepts – has apparently received very little attention from researchers in chemical education. A related study on solvation energetics, however, has recently appeared (Ebenezer and Fraser 2001).

International Journal of Science Education ISSN 0950–0963 print/ISSN 1464–5289 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/09500690210145729

Calorimetry, in the context of chemical reactions in aqueous solutions, is often the very first topic in the chemistry curriculum in which thermodynamic ideas are applied. In view of that fact, it is somewhat ironic that calorimetry has itself received so very little attention in the chemical education literature. Virtually no research data seem to have been published regarding student learning of thermodynamic concepts specifically in the context of solution calorimetry, although some preliminary data have been reported by Keller and Weeks-Galindo (1998). For this reason, in the present investigation we have set for ourselves the following research questions: What are the primary conceptual difficulties faced by college chemistry students in their initial study of calorimetry? How do these relate to other student difficulties with thermodynamic concepts previously identified in the research literature?

Previous work

The science education literature has numerous studies reporting on the difficulties students have with the concepts of heat and temperature (Erickson 1979, 1980, 1985, Tiberghien 1983, 1985, Kesidou *et al.* 1995). Cohen and Ben-Zvi (1992) suggested that misconceptions can develop because of the relatively large number of abstract concepts involved, and Linn and Songer (1991) recommended using a simplified 'heat-flow' model in middle-school instruction. In the context of thermochemistry, several investigators have reported student difficulties in understanding and distinguishing between exothermic and endothermic reactions (Johnstone *et al.* 1977, Novick and Nussbaum 1978, Thomas and Schwenz 1998, De Vos and Verdonk 1986). Boo (1998) has reported a detailed investigation of learning difficulties encountered by students in the study of chemical reaction energetics, while Barker and Millar (2000) found that A-level students demonstrated a very weak understanding of the energy changes associated with the breaking and forming of bonds in chemical reactions.

Kesidou and Duit (1993) have discussed the common student confusion between the terms 'heat' and 'temperature'. Heat is frequently viewed as an intensive quantity and temperature interpreted as degree of heat, i.e., as a measure of its intensity. However, heat is a process-dependent variable and represents a transfer of a certain amount of energy between objects or systems due to their temperature difference. Temperature, by contrast, is a measure of the average kinetic energy of molecules in a particular system. Gabel and Bunce (1994) state that:

> . . . although many of these concepts [heat and temperature] are important for understanding science . . . an in-depth understanding of them is not essential for solving many of the chemistry exercises and problems that appear in chemistry textbooks.

However, it seems that no investigations have been reported regarding the possible contribution of such conceptual understanding to college chemistry students' studies of calorimetry.

Chemical reactions and solution calorimetry

In constant pressure calorimetry experiments involving aqueous solutions, chemists view the *reaction* as the system and the total mass of the solution and the calorimeter as the surroundings. The chemical reaction that occurs, although it can exchange heat with its surroundings, is represented as an abstract entity that does not have mass. The mass of the reactants plus the mass of water, the solvent, are viewed as the total mass of the solution. It is the total mass of solution that absorbs the heat which is released by the forming of bonds during the course of a chemical reaction. Therefore, the reactants in a calorimetry experiment are viewed by chemists in two distinct ways – as the entity that releases heat, and as part of the mass that gains heat. This is a difficult concept for students to understand and apply, and it makes thermochemical experiments more difficult to comprehend than physical processes in which two objects with different temperatures are placed in contact in an insulated container. Most undergraduate students can easily understand that the hotter object in such a process transfers heat to the cooler object until thermal equilibrium is reached.

One ordinarily defines q_A as the amount of heat absorbed by object A, i.e., $q_A > 0$ if energy flows into the object, but $q_A < 0$ if energy flows out of the object. For simple physical processes, any energy that flows *out of* one object must flow into the other, so $q_{hotter} + qcooler = 0$. The formula $q = mc\Delta T$ can then be applied to the two objects simultaneously to find, for example, the final temperature. However, in solution calorimetry problems involving chemical reactions, students have difficulty making the inference that the heat 'absorbed by' the chemical reaction is equal in magnitude but opposite in sign to the heat 'absorbed by' the solution.

Most textbooks, including the one used by the students in this study (Brown *et al.* 2000), discuss the relationship of the law of conservation of energy to calorimetry experiments:

> One of the most important observations in science is that energy can be neither created nor destroyed: *energy is conserved*. Any energy that is lost by the system must be gained by the surroundings, and vice versa. (Brown *et al.* 2000: 149)
>
> If we assume that the calorimeter perfectly prevents the gain or loss of heat from the solution to its surroundings, the heat gained by the solution must be produced from the chemical reaction under study. In other words, the heat produced by the reaction, q_{rxn}, is entirely absorbed by the solution; it does not escape the calorimeter. For an 'exothermic' reaction, heat is 'lost' by the reaction and 'gained' by the solution, so the temperature of the solution rises. The opposite occurs for an endothermic reaction. The heat gained by the solution, q_{soln}, is therefore equal in magnitude and opposite in sign from q_{rxn}: $q_{soln} = -q_{rxn}$. The value of q_{soln} is readily calculated from the mass of the solution, its specific heat, and the temperature change. (Brown *et al.* 2000: 160)

Silberberg's (1996) general chemistry textbook discusses the source of the heat:

> *The energy released or absorbed during a chemical change is due to the difference in potential energy between the reactant bonds and the product bonds* . . . energy does not really 'come from' anywhere; it exists in the different energies of the bonds of the substances. In an exothermic reaction, $E_{\mathrm{p\ (bond)}}$ of the products is *less* than that of the reactants, so $\Delta E_{\mathrm{p\ (bond)}} < 0$ and the system *releases* the energy difference. (Silberberg 1996: 231, emphasis in original)

Qualitative research, think-aloud interviews, and case studies

The think-aloud interview technique has been used to elicit student understanding of chemistry and physics concepts and approaches to problem solving (Clement 1979, Champagne *et al.* 1985, Larkin and Rainard 1984, Herron and Greenbowe

1986, Nakhleh and Krajcik 1993, Bowen 1994, Welzel and Roth 1998). With respect to thermodynamics, Thomas (1997) and Thomas and Schwenz (1998) reported a study in which they interviewed 16 college students enrolled in a physical chemistry course about their understanding of equilibrium and thermodynamics. Even though the students were in an advanced chemistry course, most of them showed a lack of understanding of basic thermochemistry principles, including the meaning of 'heat' and 'temperature' (Thomas 1997: 80–81). Harrison *et al.* (1999) reported a case study of one student's understanding of heat and temperature from observations made over an eight-week period. Qualitative data collected for this study included transcripts of all classroom discussions and a student portfolio containing all written work. Through class activities which employed the *Physics by Inquiry* curriculum (McDermott 1996), the subject became better able to distinguish the meaning of the terms heat and temperature.

Our instructional experience had persuaded us that a number of serious and widespread thermochemical misconceptions are developed among college chemistry students, even those who are successful in solving algorithmic calorimetry problems. This is consistent with previous research which found that students use algorithms to help solve chemistry problems but fail to exhibit conceptual understanding (Bodner 1987, Gabel *et al.* 1987, Nurrenbern and Pickering 1987). To examine this issue, our study included *both* quantitative and qualitative problems; data sources included both student interviews and written work on students' exam papers.

Method

This study incorporates both detailed analysis of student performance on written exams for a moderately large sample of students (n = 207) and extensive longitudinal interview data from a single subject who was herself part of the same class from which that larger sample was drawn. We were able to 'calibrate' our single subject, so to speak, by comparing her performance on the various written exam questions with the performance of her classmates in the larger sample. This allowed us to make a judgment regarding the likelihood of her views being representative of a significant portion of the larger sample.

The students in this study were enrolled in an introductory chemistry course for science and engineering majors at a large mid-western university in the USA. The primary data source for the study was an analysis of students' work on two calorimetry problems for a subset of the entire class. The first problem was on the second hour examination and the second problem was on the final examination. Prior to the second hour examination, as part of the normal course work, students had the opportunity to attend three lectures on thermochemistry and calorimetry. They had the opportunity to do the assigned readings in the textbook (Brown *et al.* 2000), work homework problems, and participate in recitation and laboratory sessions on calorimetry and enthalpy.

A subset of student examination papers was selected for detailed analysis. These samples were randomly selected from the work of the entire class of students enrolled in the course (n = 541); the sample represents more than one third of the entire class (second hour exam, n = 185; final exam, n = 207). The appropriate pages from each student's examination were photocopied.

A letter was attached to about 50 students' second hour examination paper when it was returned to them, asking if they would volunteer to discuss their responses. These students had exhibited a range of problem-solving performance and conceptual understanding and none had received a grade of 'A' or 'F' on that exam. Ten students showed up for the initial interview and from this group, an individual we refer to as 'Sophia' agreed to a series of interviews. Her work and performance were compared to students from her class who solved the same calorimetry problems. Over a three-month period, observations of Sophia's work and thinking were made and two instances of instructional intervention were provided. Hence, a longitudinal case study of Sophia's understanding of calorimetry was generated.

Sophia was chosen for the case study because of her ability to clearly state her conceptions and problem solving methods. Her examination scores in the introductory chemistry course indicated she was an above-average student. Overall, we believe that she is a student who is representative of her classmates. She was asked to explain what she did on the calorimetry exam problems by thinking aloud. She gave permission for a tape recorder to be used to record her voice and she signed a voluntary informed consent form agreeing to the conditions of the interviews, including the analysis of her work on the course examinations. She regularly volunteered her opinions and willingly expressed her views during the interview sessions.

There were four interview sessions with Sophia, an average of two hours each. Sessions 1, 2, and 3 occurred between the second hour examination and the final examination; Session 4 occurred after the final examination and focused on her work on that examination. Sessions 1 and 4 involved neutral observations and interactions, while Sessions 2 and 3 involved some instructional intervention, engaging Sophia in an interchange involving 'the juxtaposition of conflicting ideas, forcing reconsideration of previous positions' (Guba and Lincoln 1989: 90). The principal interviewer was one of the authors of this paper, and neither author was the instructor for Sophia's introductory chemistry course.

A description of the calorimetry problem on the second hour examination

This problem (figure 1) was a modified version of an end-of-chapter problem from the course textbook; it involves the mixing of two aqueous solutions of known concentration and volume. The initial and final temperatures of the solutions are measured. The goal is to determine the heat of reaction, and then the molar enthalpy change of the reaction. The format of this problem appears in several general chemistry textbooks as in-chapter examples and end-of-chapter exercises (Zumdahl and Zumdahl 2000, Brown *et al.* 2000, Chang 1998).

Individuals solving this problem are expected to realize that there is a transfer of energy from the chemical reaction to the mass of the resultant solution. (It is assumed that no heat is released or absorbed by the calorimeter.) The equation $q = mc\Delta T$ is used to calculate q_{soln}, the heat absorbed by the solution, and the relation $q_{rxn} + q_{soln} = 0$ is applied to determine the heat of reaction. Since the process occurs at constant pressure, $\Delta H_{rxn} = q_{rxn}$; therefore, dividing the heat of reaction by the number of moles of the limiting reagent

Calorimetry problem on the second hour examination

In a constant-pressure calorimeter with negligible heat capacity, 50.0 mL of 2.00 M HCl and 50.0 mL of 2.00 M NH_3 were combined. The initial temperature of both solutions was 22.4°C. The temperature of the combined solutions rose to 34.8°C after mixing. Assume that the specific heat of all the solutions is 4.18 J/g-°C, and assume that all solutions have a density of 1.01 g/mL.

a. How much heat did this reaction generate in the calorimeter?

b. What is ΔH for this reaction in kJ/mol?

Calorimetry problem on the final examination

The following reaction takes place at constant pressure in an insulated calorimeter: 1.00 L of 2.00 M $Ba(NO_3)_2$ solution at 25.0°C was mixed with 1.00 L of 2.00 M Na_2SO_4 solution at 25.0°C. The final temperature of the solution after mixing was 31.2°C. Assume that all solutions had a density of 1.00 g/mL and a specific heat of 4.18 J/g-°C.

a. What is the system?

b. What are the surroundings?

c. Calculate the heat of reaction (in kJ).

d. Is the reaction endothermic or exothermic?

e. Write a balanced chemical equation for the reaction.

f. Calculate the change in enthalpy (ΔH) for the reaction with units of kJ per mole of $Ba(NO_3)_2$ that reacts.

g. If 0.500 L of 2.00 M $Ba(NO_3)_2$ solution at 25.0°C is mixed with 0.500 L of 2.00 M Na_2SO_4 solution at 25.0°C, the final temperature of this solution will be ________________ (more than, less than, or equal to) 31.2°C (within experimental error).

Figure 1. Calorimetry problems on the second hour examination and final examination.

determines the molar enthalpy change for the reaction. Specifically, we have for parts (a) and (b):

$$\text{(a)} \quad m = \rho V = (1.01\,\text{g/mL})(100.0\,\text{mL}) = 101\,\text{g}$$

$$q_{soln} = mc\Delta T = (101\,\text{g})(4.18\,\text{J/g-°C})(+12.4\,\text{°C}) = +5.24\,\text{kJ}$$

$$q_{rxn} = -q_{soln} = -5.24\,\text{kJ}$$

a. The chemical reaction

b. The solution, consisting mostly of water, and the calorimeter. (Calorimeter can be assumed to have negligible heat capacity, and so may be ignored in the calculation.)

c. $m = \rho V = (1.00\,\text{g/mL})(2 \times 10^3\,\text{ml}) = 2 \times 10^3\,\text{g}$

$q_{soln} = mc\Delta T = (2 \times 10^3\,\text{g})(4.18\,\text{J/g-°C})(+6.2°\text{C}) = +52\,\text{kJ}$

$q_{rxn} = -q_{soln} = -52\,\text{kJ}$

d. Exothermic.

e. $Ba(NO_3)_2\,(aq) + Na_2SO_4\,(aq) \rightarrow 2NaNO_3\,(aq) + BaSO_4(s)$

f. $1.00\,\text{L} \times 2.00\,\text{mol/L} = 2.00\,\text{mol}\ Ba(NO_3)_2$

$1.00\,\text{L} \times 2.00\,\text{mol/L} = 2.00\,\text{mol}\ Na_2SO_4$

$$\Delta H_{rxn} = \frac{q_{rxn}}{n_{limiting\ reagent}} = \frac{-52\,\text{kJ}}{2.00\,\text{mol}} = -26\,\text{kJ/mol}$$

g. $0.500\,\text{L} \times 2.00\,\text{mol/L} = 1.00\,\text{mol}\ Ba(NO_3)_2$

$0.500\,\text{L} \times 2.00\,\text{mol/L} = 1.00\,\text{mol}\ Na_2SO_4$

$q_{rxn} = \Delta H_{rxn} \times n_{limiting\ reagent} = -26\,\text{kJ/mol} \times 1.00\,\text{mol}\ Ba(NO_3)_2$
$= -26\,\text{kJ}$

$q_{soln} = -q_{rxn} = +26\,\text{kJ}$

$m = \rho V = (1.00\,\text{g/mL})(10^3\,\text{mL}) = 10^3\,\text{g}$

$$\Delta T = \frac{q_{soln}}{mc} = \frac{+26\,\text{kJ}}{(10^3\,\text{g})(4.18\,\text{J/g-°C})} = +6.2°\text{C}$$

$T_{final} = 25.0°\text{C} + 6.2°\text{C} = 31.2°\text{C}$

Figure 2. Solution to calorimetry problem on final examination.

(b) $0.0500\,\text{L} \times 2.00\,\text{mol/L} = 0.100\,\text{mol HCl}$

$$\Delta H_{rxn} = \frac{q_{rxn}}{n_{limiting\ reagent}} = \frac{-5.24\,\text{kJ}}{0.100\,\text{mol}} = -52.4\,\text{kJ/mol}$$

A description of the calorimetry problem on the final examination

This problem (figure 1) is similar to the one described above; a solution is shown in figure 2. Students are asked to identify the system and the surroundings, and

Table 1. Types of approaches used by students when calculating the heat of reaction on the second hour examination part (a), and the final examination part (c).

	Second hour examination ($n = 185$)	*Final examination* ($n = 207$)	
Correct or nearly correct magnitude of q_{rxn}√†	50%	40%	
		negative value 14% √	
		positive value 26% †	
Errors using formula			
Set $q = \Delta T$ (or $q = T$)	8%	5%	
Did not use $q = mc\Delta T$ *or* $q = \Delta T$	11%	9%	
Errors in value for mass			
Used mass of the reactants only	15%	21%	[Final exam:]
Used mass of one solution only	8%	5%	*negative value 13%*
Other responses	7%	15%	*positive value 41%*
No answer	2%	6%	

Notes: All values are in percent of total n for respective exam.
A √ indicates the correct response.
A † indicates the response of Sophia on that category.
'Nearly correct' means there was only a simple math error.
'Other response' means *did* use $q = mc\Delta T$, but error did not fall into other categories.

whether the reaction is exothermic or endothermic. They are also asked to calculate the heat of reaction, and then the molar enthalpy change for the reaction. Finally, students are asked to consider the final temperature for a system involving the mixing of 500 mL of each reactant, instead of 1.00 L of each: Would T_{final} be more than, less than, or equal to that observed in the original system?

Results

It is notable that none of the students in this study acknowledged the fact that since the reactions occurred under conditions of constant pressure, the heat of reaction (q_{rxn}) is equal to the enthalpy change of the reaction (ΔH_{rxn}). Also, fewer than 1% of the students stated explicitly that $q_{rxn} + q_{soln} = 0$. (One might suggest that the common practice of tolerating students' failure to explicitly state fundamental assumptions and constraints in exam solutions may be, ironically, a factor that contributes to hindering students' understanding.)

Analysis of students' responses to questions on heat of reaction

Table 1 shows the categories of responses contained in students' work on the parts of the calorimetry problems dealing with the heat of reaction, along with the percentage of the student sample corresponding to each response. Because of the way this problem was worded on the second hour exam ('How much heat did this reaction generate in the calorimeter?'), only the magnitude (and not the sign) of the students' responses was considered in the case of the second hour exam.

On the second hour exam, we counted as correct or nearly correct student answers for part (a) that had the correct magnitude for q_{rxn}, or that contained only very minor mathematical errors. Only 50% of the students were able to successfully calculate the magnitude of q_{rxn}. The major problem seems to be the use of an incorrect mass for the entity (the surroundings) that is absorbing the heat from the system (the chemical reaction).

Table 1 also includes a summary of students' responses to part (c) of the final examination problem (i.e., a very similar question about heat of reaction). Only 40% of the students were able to apply the equation $q = mc\Delta T$ with use of the correct mass to generate a correct or nearly correct magnitude for the heat of reaction, compared to 50% on the second hour exam. Overall, there was a significant decrease in performance in comparison with the second hour exam (according to a two-sample test for binomial proportions: $z = 1.99$, $p < 0.05$). Only 14% of the students provided both a correct magnitude and correct (negative) sign, while 26% provided a correct magnitude but incorrect sign. Again, the major error exhibited by students was that of using the mass of chemical reactants and not including the mass of the water, for the total mass m in the formula $q = mc\Delta T$. It is also notable that, between the second hour exam and the final exam, there was a significant increase ($z = 1.99$, $p < 0.05$) in the number of 'no answer' responses, and also in the number of 'other' responses ($z = 2.50$, $p < 0.01$) that did not correspond to any of the other listed categories. The results suggest that students' confusion on at least some calorimetry principles actually may have increased in the time between the second hour exam and the end of the course.

Taken at face value, the determination of the heat of reaction appears to be a straightforward calculation. Using the formula $q = mc\Delta T$, students need only plug in the correct values for mass, specific heat, and the change in temperature to calculate q. Students then had to recognize that they had actually found q_{soln}, and then apply the relation $q_{rxn} = -q_{soln}$. However, the students' exam responses indicate severe difficulties in a number of areas.

On the final exam question regarding heat of reaction, 20% of the sample either failed to provide any response, or failed even to realize that they would need to make use of the relation $q = mc\Delta T$. Of the remainder of the sample, about one third did not understand which physical quantity corresponded to the m. Only about one student in seven could calculate a correct value for the heat of reaction accompanied by a correct sign. Some of the students equated the heat of reaction with the change in temperature, indicating that these students were quite unable to distinguish between the terms 'heat' and 'temperature'.

Analysis of students' responses to questions on molar enthalpy

Common errors exhibited by the students on part (b) of the calorimetry problem on the second hour examination are shown in table 2.

This part of the problem asks the students to calculate ΔH for this reaction in kJ/mol; only 4% of the students provided the correct magnitude and sign for the value of ΔH_{rxn}. In this case both the sign and magnitude are required. Using the formula $q = mc\Delta T$, students need only plug in the correct values for mass, specific heat, and the change in temperature to calculate a value for q. However, most

Table 2. Responses on the second hour examination to part (b) of the calorimetry problem, calculation of the molar enthalpy change of the reaction, ΔH_{rxn} [molar].

Description of the response	*Percentage of students exhibiting this response (n = 185)*		
Correct or nearly correct magnitude for ΔH_{rxn} √	**18%**		
(Divided q_{rxn} by 0.1 mol)			
negative sign for the value of ΔH_{rxn} √	4%		
positive sign for the value of ΔH_{rxn}	14%		
Incorrect magnitude for ΔH_{rxn}	**68%**		
negative sign for the value of ΔH_{rxn}	17%		
positive sign for the value of ΔH_{rxn}	51%		
Used incorrect number of moles			
Divided q_{rxn} by 2 mol		8%	
negative sign for the value of ΔH_{rxn}			*5%*
positive sign for the value of ΔH_{rxn}			*3%*
Divided q_{rxn} by 0.2 mol		14%	
negative sign for the value of ΔH_{rxn}			*3%*
positive sign for the value of ΔH_{rxn}			*11%*
Equated enthalpy and temperature			
ΔH_{rxn}*[molar]* = ΔT		3%	
Equated molar enthalpy and heat			
ΔH_{rxn}*[molar]* = q_{rxn}		12%	
negative sign for the value of ΔH_{rxn}			*4%*
positive sign for the value of ΔH_{rxn}			*8%*
Math errors		3%	
Other responses		29%	
negative sign for the value of ΔH_{rxn}			*5%*
positive sign for the value of ΔH_{rxn}†			*24%*
No answer	**13%**		

Notes: A √ indicates the correct response.
A † indicates the response of Sophia on that category.
'Other responses' includes those using incorrect number of moles but which don't fall into specific categories listed above.

students seemed not to recognize that the value of 'q' calculated from the experimental data is q_{soln}, not q_{rxn}, and that the signs of those two quantities must differ. Beyond that, the major problem with this calculation seems to be dividing the heat of reaction by an incorrect number of moles.

Students' responses on the final exam question related to molar enthalpy (part (f)) are shown in table 3. Only 18% of the students were able to determine a correct (or nearly correct) magnitude along with a correct sign for the molar enthalpy change of the reaction, although this was a significant improvement ($p < 0.001$) over the 4% who succeeded on the second hour exam.

Table 3. Responses on the final examination to part (f) of the calorimetry problem, calculation of the molar enthalpy change of the reaction, ΔH_{rxn} [molar].

Description of the response	*Percentage of students exhibiting this response (n = 207)*			
Correct or nearly correct magnitude for ΔH_{rxn} √ † (Divided q_{rxn} by 2 mol)	**34%**			
negative sign for the value of ΔH_{rxn} √ †		18%		
positive sign for the value of ΔH_{rxn}		16%		
Incorrect magnitude for ΔH_{rxn}	**39%**			
negative sign for the value of ΔH_{rxn}		12%		
positive sign for the value of ΔH_{rxn}		27%		
Used incorrect number of moles				
Divided q_{rxn} by 4 mol			2%	
negative sign for the value of ΔH_{rxn}				*0.4%*
positive sign for the value of ΔH_{rxn}				*2%*
Equated enthalpy and temperature				
ΔH_{rxn}*[molar]* = ΔT			2%	
Equated molar enthalpy and heat				
ΔH_{rxn}*[molar]* = q_{rxn}			10%	
negative sign for the value of ΔH_{rxn}				*4%*
positive sign for the value of ΔH_{rxn}				*5%*
Other responses			25%	
negative sign for the value of ΔH_{rxn}				*7%*
positive sign for the value of ΔH_{rxn}				*17%*
No answer	**27%**			

Notes: A √ indicates the correct response.
A † indicates the response of Sophia on that category.
'Other responses' includes those using incorrect number of moles but which don't fall into specific category listed above.

Table 4 outlines the responses given by students to parts (a), (b), (d) and (g) of the calorimetry problem on the final examination. With the exception of the identification of the reaction as an 'exothermic reaction', for which 71% of the students were correct, more than 50% of the responses to these questions were incorrect. The chemical reaction was identified as the system by only 22% of the students, while only 6% of the students correctly identified the solution and the calorimeter as the surroundings. (If the students identified the mass of the resultant solution as the surroundings, they received a rating of 'correct'.)

Sophia's work on the calorimetry problems

Sophia earned six points out of eight on the calorimetry problem on the second hour exam. In trying to calculate the molar enthalpy change, Sophia divided the

Table 4. Responses on the final examination to parts (a), (b), (d) and (g) of the calorimetry problem.

Description of the response	*Students' answers*	*Percentage of students exhibiting this response* (n = 207)	
(a) What is the system?			
Correct answer	the chemical reaction √	22%	
Partially correct	the reactant(s)/reactant solution †	7%	
Incorrect answers		71%	
	Everything inside the calorimeter		14%
	The calorimeter		23%
	The solution		22%
	Calorimeter and Contents		5%
	Other		6%
	No answer		1%
(b) What are the surroundings?			
Correct answer	The solution and calorimeter √ The solution or the water	6%	
Partially correct	The calorimeter †	32%	
Incorrect answers		62%	
	Everything outside the calorimeter		31%
	The calorimeter and everything else		8%
	The air		5%
	Other		18%
	No answer		1%
(d) Is the reaction exothermic or endothermic?			
Correct answer	exothermic √ †	71%	
Incorrect answer	endothermic	29%	
(g) [Comparison of the change in temperature of the two systems]			
Correct answer	equal to √	44%	
Incorrect answer	more than †	10%	
Incorrect answer	less than	43%	
No answer		2%	

Notes: A √ indicates the correct response.
A † indicates the response of Sophia on that category.

heat of reaction by the 'moles of solution' instead of dividing by the number of moles of limiting reagent. On the final examination calorimetry problem, Sophia correctly identified the two reactants as part of the system, but she did not indicate that it is the entire chemical reaction that is considered to be the system. She identified the calorimeter as being part of the surroundings, but she did not identify the solution as being part of the surroundings. She correctly used the total mass of the solution to calculate the heat absorbed by the solution, and then correctly inferred that this must be the heat that was transferred from the system. She correctly divides the heat of reaction by the number of moles of limiting reagent involved to generate a correct value for the molar enthalpy change of the reaction. Sophia did have a negative value for ΔH_{rxn}. She did not, however, realize that the 'heat of reaction' must also have a negative sign. Sophia incorrectly stated that the system in part (g) would produce a greater change in temperature in comparison with the original problem. She does not explicitly write down the relation $q_{rxn} = -q_{soln}$, yet she succeeds in correctly solving all but the last part of this problem.

Excerpts from Sophia's interviews

In order to confirm and to elaborate on why Sophia answered some of the parts of the calorimetry problem the way she did, an interview session (Session 4) was scheduled five days after her final examination. In this session, the interviewer was trying to assess why Sophia did not identify the solution as part of the surroundings, to assess Sophia's understanding of the term 'exothermic', and to assess her understanding of the use of positive and negative signs to indicate endothermic and exothermic processes respectively.

I: Would you walk me through what you were doing and thinking on this calorimeter problem on your final exam.

Sophia: I thought that the system is the two solutions reacting and the surroundings was the calorimeter because it was at constant pressure and that the calorimeter was insulated, so anything outside the calorimeter was not going to affect the reaction . . .

I said it was exothermic because the temperature increased . . . For the change in enthalpy, I took the heat of reaction that we had found and it asked for per moles of barium nitrate. I found moles of barium nitrate by using litres and molarity. I divided the heat of reaction by those moles because I figured that the heat had to be the same so the change in enthalpy was the same as up here . . .

I: You have a negative ΔH, ΔH equals negative 25.9 kilojoules per mole . . .

Sophia: Yes, I have it negative because it is an exothermic reaction.

Sophia seems to have a good understanding of when to use positive and negative signs to indicate endothermic and exothermic process, and she elaborates a bit on the responses she gave on the final examination regarding the questions of 'what is the system' and 'what is the surroundings'. Later in this interview she also demonstrated understanding of the concept of molarity despite confusion about its application to specific heat problems. Additional excerpts from her interviews will be presented and discussed in the next section.

Students' conceptual misunderstandings uncovered by the investigation

In this section, we will summarize the specific conceptual difficulties regarding calorimetry encountered by the students in our sample, as reflected by our analysis of the data. This includes both the written exam data and the interviews conducted with the subject Sophia.

Lack of recognition that energy flow out of reactants and into solution implies a negative 'heat of reaction', which, for constant-pressure processes, has the same meaning as a negative change in enthalpy of the reactants, i.e. that $\Delta H_{rxn} < 0$

On both the second hour examination (question part b) and the final examination (question parts c and f), students had been asked to consider an exothermic reaction under constant-pressure conditions in which net energy is transferred from the chemical bonds in the reactants and products *to* the solution. The direction of energy flow can be recognized simply from the fact that the temperature of the solution increases. The conclusion should be that both the heat of reaction and the enthalpy change are negative in both cases, i.e. that both $q_{rxn} < 0$ and $\Delta H_{rxn} < 0$. However, on all three relevant questions, a large majority of the students who responded gave a positive value for their answer.

It is not clear how many of these errors in the sign of q_{rxn} and ΔH_{rxn} can be attributed to simple carelessness, and how many actually reflect a fundamental physical misunderstanding. A large majority (71%) of the students correctly identified the reaction as 'exothermic' on the final exam question, part (d). However, this may simply reflect a learned recognition that an increase in solution temperature corresponds to an exothermic reaction. (This is precisely the reasoning given by Sophia in Interview Session 4; see below.) Textbooks often make reference, rather loosely, to the heat 'released by', 'produced by', or 'evolved by' the reaction, but these terms are sometimes – not always! – assumed to refer to the absolute value of the heat of reaction – i.e., to $|q_{rxn}|$, which is defined to be a positive quantity (e.g., Zumdahl and Zumdahl 2000: 253). This obviously increases the potential confusion for the student.

Students do not necessarily give consistent answers to this type of question. On the final exam question Sophia, for example, correctly identified the reaction as exothermic and ΔH as negative; however she gave a positive value for the heat of reaction. From Interview Session 4, it is obvious that Sophia is well aware of the chain of reasoning that goes *increase in solution temperature* $\Rightarrow$ *exothermic reaction* $\Rightarrow$ $\Delta H < 0$. Here is how she explains her answer to part (f) of the final exam question during this interview:

Sophia: . . . I said it was exothermic because the temperature increased . . .

I: You have a negative ΔH, ΔH equals negative 25.9 kilojoules per mole . . .

Sophia: Yes, I have it negative because it is an exothermic reaction.

I: And you have written '$\Delta H = -q_{rxn}$'?

Sophia: Yes, I was not so sure about that. I was trying to show that it was going to be negative because it was exothermic.

In contrast to her reasoning above, she explains her answer to part (c) as follows:

Sophia: When it said to calculate heat of reaction I used the equation $q = mc\Delta T$. I found the mass by adding the two volumes, the litres, one of each solution. Then I put the values into the equation. I did the calculation on my calculator and got 51.8.

The issue of the sign of the heat of reaction – positive or negative – seems never to have entered her considerations. It appears that for Sophia, as for many other students, the fact that the 'heat of reaction' may have a positive or a negative sign – and that in fact, for constant-pressure processes, the heat of reaction is really just the same thing as the change in enthalpy ΔH – is simply an idea that has never been fully understood.

Identifying the 'heat of reaction' or 'heat generated by reaction' as simply the temperature change that results from that heat flow

Perhaps the most well-known and widely discussed student misunderstanding in the field of heat and thermodynamics is the confusion of 'heat' and 'temperature' (e.g., Kesidou *et al.* 1995). In calorimetry problems the distinction is made explicit, at least in quantitative terms, through application of the equation $q = mc\Delta T$. None the less, when asked to find the amount of heat generated by the reaction (on the second exam), and the heat of reaction (on the final exam), some students simply responded with the value of the temperature change ΔT of the solution, i.e., 12.4 °C on the second hour exam (response given by 8% of the students), and 6.2 °C on the final exam (response given by 5% of the students). In response to a question about an amount of heat – which should be measured in joules or calories – these students responded with a temperature, measured in degrees. Sophia expressed a related confusion during Session 4:

I: Good. Now, one last question, what is the difference between heat and temperature?

Sophia: Heat is energy being released or absorbed by something. Like 'q' here is the energy being released. Temperature is just a way to measure it.

I: When you use a thermometer in a calorimeter experiment, are we measuring heat or temperature when a reaction takes place?

Sophia: Heat.

The word 'heat' is properly used to represent an energy transfer into or out of a system due to a temperature difference, and it is a quantity for which a larger magnitude necessarily corresponds to a larger absolute amount of energy. The mistaken idea that temperature (an intensive quantity) is merely a measure of an amount of heat, rather than a measure of the average kinetic energy per molecule – a quantity distinctly different from heat – is clearly a misunderstanding that lingers on in many students' minds.

Not recognizing and applying the relationship between heat flow, specific heat, and temperature change (i.e., not making use of equation $\mathrm{q = mc\Delta T}$*)*

A significant number of students were simply unaware that they needed to apply the relationship $q = mc\Delta T$ in order to find the heat of reaction. About 10% of students on both the second hour exam and the final exam attempted unsuccessfully to calculate the heat of reaction without using the relevant equation.

Not recognizing that the 'm' *in the relationship* $\mathrm{q = mc\Delta T}$ *refers to the total mass of the solution contained within the calorimeter, and does not refer merely to the mass of the molecules that react to generate the heat flow*

The single most common confusion found among our student sample was that related to the meaning of 'm' – the mass – in the equation $q = mc\Delta T$. The 'ΔT'

in this case refers to the temperature change of the entire contents of the calorimeter, which is to say the total mass of the solution. The *m*, then, refers to the mass of that solution. However, about one-quarter of all students, on both exams, expressed confusion on this point. The most frequently expressed student idea (15% of students on the second exam, and 21% on the final) was that this mass refers in some fashion only to the molecules that are engaged in the chemical reaction that produces the heat. This misunderstanding led to a wide variety of incorrect numerical answers.

A significant number of students, although realizing that the *m* referred to the mass of the solution, did not realize that it was the entire mass of solution that had to be considered. As a result, 8% of students on the second hour exam and 5% on the final exam set *m* equal to half the mass of the total solution. Apparently, they were misled by the fact that there were two reacting species, each of which originally represented half of the total solution.

Not understanding that in solution calorimetry, the thermodynamic 'system' refers to the reacting molecules and their products – more precisely, to the chemical bonds (assumed to be massless) that are both made and broken (i.e., the 'reaction') – and that the 'surroundings' refers to the entire mass of material contained within the calorimeter (and, in principle at least, that which is outside the calorimeter as well)

Although it is admittedly a subtle point, the meaning of the terms 'system' and 'surroundings' in the context of calorimetry often presents students with their first opportunity to try to relate thermodynamic terminology to an actual laboratory set-up. The interpretation of these terms in the context of calorimetry was stressed during the lectures in this course. However, on the final exam, fewer than one-third of students were able to give anything close to an acceptable answer to the question 'What is the system?' Similarly, fewer than 40% of the students could properly identify the 'surroundings'.

Not understanding that the molar enthalpy change refers to the relevant quantity (i.e., ΔH_{rxn}*) divided by the number of moles of one of the reacting species, and that for constant-pressure processes,* $\Delta H_{rxn} = q_{rxn}$

Part (f) of the final exam question clearly asks for the change in enthalpy ΔH per mole of one of the reactants. Therefore, a correct response would be to divide the heat of reaction q_{rxn} (i.e., the answer to part (c)) by the number of moles of this reactant (i.e., 2.00). A large number of students answered this question incorrectly. 10% of the students simply copied their numerical answer from part (c), while 29% of the students made other types of errors in this calculation (not including those who merely carried over an incorrect answer from part (c)). Twenty seven per cent of students gave no response to this question at all.

Believing that the heat flow is produced by an energy transfer from one reactant to another, rather than from the breaking and forming of chemical bonds to the total mass of material contained within the calorimeter

This extremely interesting confusion was expressed by Sophia during Interview Session 1. On the one hand, she seems to understand that the solution is absorbing heat. On the other hand, she quite clearly is under the impression that heat is flowing *from* the solution containing one of the reacting species, *to* the solution containing the other, and is not sure about which is the source and which is the recipient. She also appears to express a confusion between a system where a chemical reaction is the source of heat, and a quite different system in which one physical object (such as hot metal) is a source of heat that flows into a surrounding liquid.

I: What are you measuring with the thermometer?

Sophia: The heat is rising in the solution because something is letting off heat but it is going into solution. There is a transfer of heat. It is going from one object to another.

I: And what is that object to the other?

Sophia: It is from one chemical to the other but I am not sure which is giving it off and which is absorbing it.

I: So, identify the chemicals that are in that solution.

Sophia: Hydrochloric acid and ammonia

I: Any other chemicals in there?

Sophia: Water. So I think water is the one absorbing the heat when the temperature is given off. I don't think water is part of the reaction. That is why we can exclude it in this problem. It is not part of the equation for finding heat.

I: So, is there water in this 101 grams?

Sophia: There is water in the 101 grams? I don't know this. Because if we had a solid . . . [Sophia looks at the chemicals on the nearby table and picks up a jar of magnesium metal], say we had the magnesium and we pour HCl(aq) on it. I would then know where one thing is going to the other. Because if the solution gains heat when you put Mg in the hydrochloric acid, then we know that the liquid solution is absorbing the heat, from the solid to the aqueous solution. But, when we have two aqueous solutions, then I don't know which is giving the heat and which one is absorbing the heat.

We were able to confirm Sophia's thinking on this issue through her explanation of the heat of reaction produced during the reaction of magnesium metal and hydrochloric acid:

I: What is this *q*?

Sophia: '*q*' is heat. Heat of the reaction. So this heat is what is given off by the magnesium and transferred to the hydrochloric acid solution. The magnesium gives or transfers heat to the 6 M HCl solution and that is why the solution gets warm. And you can see it happening because the magnesium reacts with the HCl and gives bubbles. The magnesium is where the reaction is taking place because you can see it happening!

It is very clear that Sophia does not have a concept of energy being transferred due to the breaking and forming of bonds within the reacting species; rather, she is convinced that energy flows from one of the reactants, to the other. It is difficult to say at this time just how widespread this belief may be among students in general. However, it seems likely that it forms an important component of many students' thinking, and it certainly merits additional investigation.

Belief that the total amount of heat generated depends on the concentration of the reacting solutions, rather than on the total mass of reactants

On part (g) of the final exam question, although the mass of reactants was cut in half, Sophia assumed that since the concentrations remained unchanged (at 2.0 molar), the total heat generated would also be unchanged. Here is how she expressed her thinking (using the idea of 'micro-heaters' previously introduced as a metaphor for the chemical reaction):

Sophia: So you have the same amount of concentration but you have less water. So you have the same number of micro-heaters. Just less water . . . I assumed that the heat was going to be the same.

I: . . . Why would the heat be the same up here? . . .

Sophia: Because it is the same reaction taking place. So the molarities are the same. But the only thing that changes is the volume, so the mass changes.

Implications for instruction

In the introductory chemistry curriculum, solution calorimetry problems are often the first practical application in which ideas about heat, temperature, energy changes in chemical reactions, and conservation of energy are combined. Because of the relatively simple calculations involved in calorimetry, it is tempting for both students and instructors to overlook the need for careful attention both to straightforward matters (such as the positive or negative sign of the heat of reaction), and more subtle concepts (such as the bond-forming origin of reaction heats).

The results of this investigation suggest a number of specific areas in which increased attention by instructors may yield a significant return in improved student understanding of thermochemistry and calorimetry:

Students' inattention to the sign of an energy change is a common error. This may represent a more serious misunderstanding of just how a change in a physical quantity is ordinarily defined (i.e., change is equal to final value minus initial value). In any case, consistent attention to sign conventions is important in reducing unnecessary calculation errors that are potentially wasteful of students' time and energy. Students might be advised to make the very first step in a calorimetry calculation a consideration of the sign – positive or negative – of the quantity being determined. In some cases (such as the problems described here), the sign can be determined as a matter of inspection. In other cases, a calculation will first have to be carried out.

The commonly misunderstood distinction between 'heat' and 'temperature' often first becomes an issue in the context of calorimetry. Our data support a widely reported finding: students' belief that these terms are essentially synonymous is not easily dislodged. We suggest that the realm of physical calorimetry, i.e., where physical changes only are involved, offers the best opportunity to clarify the distinction between heat and temperature. Numerous curricular approaches have been developed to achieve this goal (e.g., McDermott 1996). Once the complication of a chemical reaction is introduced, analysis of the system becomes considerably more challenging. We suggest therefore that the heat-temperature distinction is best treated before reaction energetics is introduced.

It is important to counter students' tendency to misunderstand the meaning of the mass m in the relationship $q = mc\Delta T$. Most commonly, students' errors reflected a misapprehension that the mass m referred only to the reacting species, and not to the entire quantity of material that was undergoing the temperature change. One must help students understand that in the equation $q = mc\Delta T$ the temperature change ΔT is that of the entire contents of the calorimeter, and so the mass m, the specific heat c, and the heat absorbed q must also refer to that contents.

The key to understanding energy changes in chemical systems is that reaction energies result from the breaking and forming of chemical bonds. In calorimetry, energy flows into or out of an aqueous solution as a result of bonds forming and breaking. In a physical system, by contrast, energy flows from a hotter object (such as a piece of hot metal) into a cooler object (a water solution, for instance). No changes in chemical bonds are involved. There is evidence both from the present study, and in the research literature, that confusion on this concept may be widespread among introductory students. The serious misconception expressed by Sophia that energy is transferred from one of the reacting species to the other may well underlie errors made on such questions as the meaning of 'system' and 'surroundings', and the temperature change resulting from a system that contains only half of the original quantity of reactant solutions. Because of the central importance of this issue, we will discuss it in some detail.

Martins and Cachapuz (1993) interviewed both high school and college chemistry students in Portugal to determine how they would explain the temperature increase observed in a water solution when a piece of sodium metal is placed in it. The most popular explanation was that energy was being transferred from the sodium to the water: '. . . the sodium gives out energy and the water takes in that energy . . . it becomes hotter . . .' (This is virtually the same explanation given by Sophia in the case of Mg and HCl.) In an earlier study Cachapuz and Martins (1987) had found that students often invoke a 'principle reactant' explanation in which one of the reactants plays a more important role than the others. (Similarly, Brosnan (1992) has suggested that students view chemical reactions as being caused by an active agent acting on a passive substance.) What is missing from these explanations is an appreciation of the central role of the breaking and forming of chemical bonds, and the associated absorption and release of energy, respectively.

Boo (1998) and Boo and Watson (2001) interviewed Grade 12 students in the UK to elicit their understanding of, among other things, the system in which magnesium is added to dilute hydrochloric acid and the temperature of the solution is observed to increase. They found that only a small minority (15%) of the students were able to give an explanation based on understanding that the bonds being made in the reaction are stronger than those which are being broken, and that therefore there is a net release of energy from the reaction, into the solution. A majority of the students were under the impression that bond making requires input of energy and bond breaking releases energy (i.e., the exact opposite of the chemist's view), or instead that both the processes of bond breaking and bond making required the input of energy. That this is a common belief was also noted by Ross (1993) and (in South Africa) by Ebenezer and Fraser (2001). Barker and Millar (2000) collected questionnaire data from UK students several months after they had completed the General Certificate of Secondary Education exams. They found only a very small proportion of students (about 10%) with a full or partial understanding of the basic energetics of chemical reactions, including an understanding of the role of bond breaking and formation. Several authors cited here have pointed out that confusion on this concept may be aggravated by the common notion that 'energy is stored in chemical bonds'.

A grasp of the law of conservation of energy and of the energetics associated with bond breaking and bond making plays a fundamental role in student understanding of chemical reactions and thermochemical phenomena (Boo 1998, Barker and Millar 2000, and references therein). The findings of the various investigators cited above, as well as our own – thereby representing four different countries – are rather striking in their consistency. It seems that students in widely disparate settings encounter a common set of conceptual difficulties related to the energetics of chemical reactions. We suspect that significant curricular enhancements and additional instructional time will be needed to improve student learning of these important concepts. Barker and Millar (2000), for instance, report very significant improvements in student learning of these concepts with the SAC curriculum, in which the exothermicity of bond formation is given explicit, extended attention. We have developed both tutorial worksheets and a computer animation[1] that guide students to confront very directly these conceptual difficulties, and we are in the process of assessing the effectiveness of these materials.

Acknowledgement

We are grateful for the cooperation and assistance of Prof. L. Keith Woo and Dr. Steven A. Walstrum. This material is based upon work supported by the National Science Foundation under Grant Number DUE–9981140.

Note

1. http://www.chem.iastate.edu/group/Greenbowe/sections/projectfolder/animations/index.htm

References

BANERJEE, A. C. (1995) Teaching chemical equilibrium and thermodynamics to undergraduate general chemistry classes. *Journal of Chemical Education*, 72, 879–887.

BARKER, V. and MILLAR, R. (2000) Students' reasoning about basic chemical thermodynamics and chemical bonding: what changes occur during a context-based post–16 chemistry course? *International Journal of Science Education*, 22, 1171–1200.

BEALL, H. (1994) Probing student misconceptions in thermodynamics with in-class writing. *Journal of Chemical Education*, 71, 1056–1057

BEN-ZVI, R. (1999) Non-science oriented students and the second law of thermodynamics. *International Journal of Science Education*, 21, 1251–1267.

BODNER, G. M. (1987) The role of algorithms in teaching problem solving. *Journal of Chemical Education*, 64, 513–514.

BOO, H. K. (1998) Students' understandings of chemical bonds and the energetics of chemical reactions. *Journal of Research in Science Teaching*, 35, 569–581.

BOO, H. K. and WATSON, J. R. (2001) Progression in high school students' (aged 16–18) conceptualizations about chemical reactions in solution. *Science Education*, 85, 568–585.

BOWEN, C. W. (1994) Think-aloud methods in chemistry education. *Journal of Chemical Education*, 71, 184–186.

BROWN, T. L., LEMAY, JR., H. E. and BURSTEN, B. E. (2000) *Chemistry: The central science*, eighth edition (Upper Saddle River, NJ: Prentice-Hall).

BROSNAN, T. (1992) *Explanation of change*, Tech paper No. 3 (London: University of London, London Mental Models Group, produced for EC Basic Research Actions ESPRIT II. Working Group 6237: Children's and Teachers' Explanations).

CACHAPUZ, A. F. and MARTINS, I. P. (1987) High School students' ideas about energy of chemical reactions. In J. Novak (ed.), *Proceedings of the 2nd International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Vol. III* (Cornell University, NY), pp. 60–68.

CHANG, R. (1998) *Chemistry*, sixth edition (Dubuque, IA: McGraw-Hill).

CHAMPAGNE, A. B., GUNSTONE, R. F. and KLOPFER, L. E. (1985) Instructional consequences of students' knowledge about phenomena. In L. H. West and L. Pines (eds), *Cognitive structure and conceptual change* (Orlando, FL: Academic Press).

CLEMENT, J. (1979) Mapping a student's causal conceptions from a problem-solving protocol. In J. Lockhead and J. Clement (eds), *Cognitive process instruction* (Philadelphia, PA: Franklin Institute Press), pp. 133–146.

COHEN, I. and BEN-ZVI, R. (1992) Improving student achievement in the topic of chemical energy by implementing new learning materials and strategies. *International Journal of Science Education*, 14, 147–156.

DE VOS, W. and VERDONK, A. H. (1986) A new road to reactions Part 3. Teaching the heat effect of reactions. *Journal of Chemical Education*, 63, 972–974.

EBENEZER, J. V. and FRASER, D. M. (2001) First year chemical engineering students' conceptions of energy in solution processes: phenomenographic categories for common knowledge construction. *Science Education*, 85, 509–535.

ERICKSON, G. L. (1979) Children's conceptions of heat and temperature. *Science Education*, 63, 221–230.

ERICKSON, G. L. (1980) Children's viewpoints of heat: a second look. *Science Education*, 64, 323–336.

ERICKSON, G. (1985) An overview of pupils' ideas [on heat and temperature]. In R. Driver, E. Guesne and A. Tiberghien (eds), *Children's Ideas in Science* (Philadelphia, PA: Open University Press), pp. 55–66.

GABEL, D. L. and BUNCE, D. M. (1994) Research on problem solving: Chemistry. In D. L. Gabel (ed.), *Handbook of research on science teaching and learning* (New York: Macmillan), pp. 301–326.

GABEL, D. L., SAMUEL, K. V. and HUNN, D. (1987) Understanding the particulate nature of matter. *Journal of Chemical Education*, 64, 695–697.

GRANVILLE, M. F. (1985) Student misconceptions in thermodynamics. *Journal of Chemical Education*, 62, 847–848.

GUBA, E. G. and LINCOLN, Y. S. (1989) *Fourth Generation Evaluation* (Newbury Park, CA: Sage).

HARRISON, A. G., GRAYSON, D. J. and TREAGUST, D. F. (1999) Investigating a grade 11 student's evolving conceptions of heat and temperature. *Journal of Research in Science Teaching*, 36, 55–87.

HERRON, J. D. and GREENBOWE, T. J. (1986) What can we do about Sue: a case study of competence. *Journal of Chemical Education*, 63, 528–531.

JOHNSTONE, A. H., MACDONALD, J. J. and WEBB, G. (1977) Misconceptions in school thermodynamics. *Physics Education*, 12, 248–251.

KELLER, P. C. and WEEKS-GALINDO, W. (1998) Thermochemistry misconceptions among students in a first year college chemistry lab. Presented at the Annual Meeting of the National Association for Research in Science Teaching, April 19–22 (San Diego, CA).

KESIDOU, S. and DUIT, R. (1993) Students' conceptions of the second law of thermodynamics – an interpretive study. *Journal of Research in Science Teaching*, 30, 85–106.

KESIDOU, S., DUIT, R. and GLYNN, S. M. (1995) Conceptual development in physics: Students' understanding of heat. In S. M. Glynn and R. Duit (eds), *Learning science in the schools: research reforming practice* (Mahwah, NJ: Erlbaum).

LARKIN, J. H. and RAINARD, B. (1984) A research methodology for studying how people think. *Journal of Research in Science Teaching*, 21, 235–254.

LEWIS, E. L. and LINN, M. C. (1994) Heat energy and temperature concepts of adolescents, adults, and experts: implications for curricular improvements. *Journal of Research in Science Teaching*, 31, 657–677.

LINN, M. C. and SONGER, N. B. (1991) Teaching thermodynamics to middle school students: What are appropriate cognitive demands? *Journal of Research in Science Teaching*, 28, 885–918.

LOVERUDE, M. E., KAUTZ, C. H. and HERON, P. R. L. (2002) Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas. *American Journal of Physics*, 70, 137–148.

MARTINS, I. P. and CACHAPUZ. A. (1993) Making the invisible visible: A constructivist approach to the experimental teaching of energy changes in chemical systems. In J. Novak (ed.) *Proceedings of the 3rd International Seminar on Misconceptions and Educational Strategies in Science and Mathematics* (Ithaca, NY: Misconceptions Trust).

MCDERMOTT, L. C. (1996) *Physics by Inquiry* (New York: John Wiley).

NAKHLEH, M. and KRAJCIK, J. (1993) A protocol analysis of the influence of technology on students' actions, verbal commentary, and thought processes during the performance of acid-base titrations. *Journal of Research in Science Teaching*, 30, 1149–1168.

NOVICK, S. and NUSSBAUM, J. (1978) Junior high school pupils' understanding of the particulate nature of matter: an interview study. *Science Education*, 62, 273–281.

NURRENBERN, S. C. and PICKERING, M. (1987) Concept learning versus problem solving: is there a difference? *Journal of Chemical Education*, 64, 508–510.

ROSS, K. (1993) There is no energy in food and fuels – but they do have fuel value. *School Science Review*, 75, 39–47.

ROZIER, S. and VIENNOT, L. (1991) Students' reasonings in thermodynamics. *International Journal of Science Education*, 13, 159–170.

SHAYER, M. and WYLAM, H. (1981) The development of the concepts of heat and temperature in 10–13 year olds. *Journal of Research in Science Teaching*, 18, 419–434.

SILBERBERG, M. (1996) *Chemistry: the molecular nature of matter and change* (Dubuque, IA: WCB, McGraw-Hill).

STAVY, R. and BERKOVITZ, B. (1980) Cognitive conflict as a basis for teaching quantitative aspects of the concept of temperature. *Science Education*, 64, 679–692.

THOMAS, P. L. (1997) Student conceptions of equilibrium and fundamental thermodynamic concepts in college physical chemistry. PhD Dissertation, University of Northern Colorado. UMI # 9729078 (Ann Arbor, MI: UMI).

THOMAS, P. L. and SCHWENZ, R. W. (1998) College physical chemistry students' conceptions of equilibrium and fundamental thermodynamics. *Journal of Research in Science Teaching*, 35, 1151–1160.

TIBERGHIEN, A. (1983) Critical review on the research aimed at elucidating the sense that the notions of *temperature and heat* have for students aged 10 to 16 years. In G. Delacôte, A. Tiberghien and J. Schwartz (eds), *Research on physics education, proceedings of the first international workshop*, La Londe Les Maures, France (Paris: Éditions du CNRS), pp. 75–90.

TIBERGHIEN, A. (1985) The development of ideas [on heat and temperature] with teaching. In R. Driver, E. Guesne and A. Tiberghien (eds), *Children's ideas in science* (Philadelphia, PA: Open University Press), pp. 67–84.

VAN ROON, P. H., VAN SPRANG, H. F. and VERDONK, A. H. (1994) 'Work' and 'heat': on the road towards thermodynamics. *International Journal of Science Education*, 16, 131–144.

WELZEL, M. and ROTH, W.-M. (1998) Do interviews really assess students' knowledge? *International Journal of Science Education*, 20, 25–44.

ZUMDAHL, S. S. and ZUMDAHL, S. A. (2000) *Chemistry*, fifth edition (Boston, MA: Houghton Mifflin).

V.

Multiple Representations and Learning of Physics

Differences in male/female response patterns on alternative-format versions of FCI items

Laura McCullough, Physics Department, University of Wisconsin–Stout, Menomonie, WI 54751
David E. Meltzer, Department of Physics & Astronomy, Iowa State University, Ames, IA 50011

A modified version of the FCI was created using female and daily-life contexts instead of the male and school-oriented contexts in the original. Both modified and original versions were administered in class. Differences among responses of males and females to both versions are discussed.

An important methodological issue in assessments using a diagnostic instrument is the degree to which slight changes in the instrument may result in altered student response patterns.[1] In view of the widespread use of the Force Concept Inventory (FCI) in physics assessment, exploration of possible context dependence of FCI items is of considerable interest. One issue is that of possible context dependence *in general* of FCI items: do students respond differently when FCI questions are very slightly modified, i.e. when essentially identical questions are posed in a different context? Steinberg and Sabella[2] compared responses on open-ended questions to responses on their FCI equivalents. They found substantial correlation between the results of the two versions with, nonetheless, some notable differences. Rebello and Zollman[3] presented students with four actual FCI questions without including the multiple-choice answer options. They also found, along with general overall agreement, a number of notable differences in response patterns when compared to those observed when multiple-choice responses *were* present. On the other hand, Adams and Slater[4] found that students' written explanations for their FCI responses were, for the most part, in good agreement with the answers they selected. Schecker and Gerdes[5] presented students with a number of FCI items, along with alternate versions of the same items posed in different physical contexts. They found differences in response patterns on some items. For instance, many students who incorrectly responded with an "impetus" model to FCI item 13 (forces acting on a steel ball thrown straight up) gave a correct Newtonian response when an almost identical question was asked with the ball being replaced by a vertical pistol shot.

Another potential issue is whether any possible context dependencies in response patterns are gender dependent. That is, do males and females differ *from each other* in terms of how their responses may change when question context is altered? This issue has been addressed by Rennie and Parker,[6] who suggest that females may be more successful when physics problems are posed in a "real-life" context. Dancy[7] explored this issue in the context of an animated version of the FCI. She found that on questions 3, 5, 14, and 26, females scored significantly better on the animated version than on the original version. Males scored significantly better on the animated version of questions 7, 14, and 26, but worse on item 20. Item 14 is of particular interest because both genders did better on the animated version and because of an unexpected response pattern in our own data.

In order to further explore the issue of possible context dependence, a "Gender" version of the FCI has been developed[8] in which each of the 30 items was rephrased or re-expressed in a slightly different context, or with a new or added diagram. Instead of school- and male-oriented contexts, daily-life- and female-oriented contexts were used in each case (e.g., instead of a cannon shooting a cannonball, a baby knocks a bowl off of her high-chair tray). The physics is identical; only the context has changed.

This paper appeared in *Proceedings of the Physics Education Research Conference, Rochester, New York, July 25-26, 2001*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (PERC, Rochester, New York, 2001), pp. 103-106.

METHOD

The Original FCI and the Gender FCI were administered to all students enrolled in the first semester course of the algebra-based general physics sequence at Iowa State University during Spring 2001. Only one of the two versions was given to each student, that version being randomly chosen according to the following procedure: Individual piles of question packets, 28 in each, were prepared for each recitation section. In each pile, the Original FCI and Gender FCI were placed alternately, so the sequence was Original, Gender, Original, etc. The recitation instructors were directed to distribute the packets in random order to all the students in their recitation section.

The tests were administered at the start of the recitation session during the second week of class. Students were told the tests would not affect their grade, but would give instructors a better idea of the students' physics background. Instructors were directed to allow at least 30 minutes, and to try to allow all students enough time to finish. Reports indicated good compliance. In one case, the instructor allowed the students to take the exams home and hand them in two days later. Response sheets that contained six or more blank responses were discarded. Three had to be discarded because the "Sex" box was not checked and the names were gender-indeterminate. In the end, the total sample contained 222 students.

We checked the results for every question to see whether there were any significant differences in performance on the two different versions of the exam. Because there are so many comparisons, we adopted $p = 0.01$ as the minimum level required to consider the difference significant. We used a statistical test for comparison of binomial proportions (equivalent in this case to chi-square analysis).

RESULTS

We found significant discrepancies for four test items (two for females only, two for males only), as follows:

1. Original FCI Item #14 (Gender item #24) [airplane/eagle drops object]

Female:
Original correct: 22%
Gender correct: 55%
$p = 0.002$*

2. Original FCI item #23 (Gender item #27) [rocket/person straight line path]

Female:
Original correct: 10%
Gender correct: 48%
$p = 0.0001$*

3. Original FCI Item #22 (Gender item #26) [rocket/person speed increasing]

Male:
Original correct: 47%
Gender correct: 18%
$p = 0.0003$*

4. Original FCI Item #29 (Gender item #13) [floor force on chair/book]

Male:
Original correct: 30%
Gender correct: 60%
$p = 0.0005$*

(Detailed results and text of "Gender" assessment items follow on the next two pages.)

CONCLUSION

Our data suggest that, in certain cases, slight changes in the context of a conceptual question may affect students' performance. Moreover, it appears that males and females may not be consistent with each other in their response to the contextual changes. More work is needed to better understand how changes in physics assessment instruments may depend on gender in their effect on performance.

ITEMS WITH SHIFTS FOR FEMALES:

Original FCI #14 (Gender item #24)
[plane/bird drops object]
Percentage of total responses each option:

	n	A	B	C	D*	E
Male, Original	*72*	10	8	8	74	0
Male, Gender	*65*	0	6	22	72	0
Female, Original	*41*	49	27	2	22	0
Female, Gender	*44*	9	20	16	55	0

Comment: No significant difference in correct responses for males. On Gender version, females show *drastically* decreased proportion selecting distracter A (which shows "backward" trajectory). Males show decrease for this option on Gender version as well. Net result is increase in correct responses by females from 22% to 55%.

Original FCI #23 (Gender item #27)
[rocket/person straight-line path]
Percentage of total responses each option:

	n	A	B*	C	D	E
Male, Original	*72*	7	47	22	19	4
Male, Gender	*65*	8	52	15	17	8
Female, Original	*41*	27	10	41	17	5
Female, Gender	*44*	5	48	41	5	2

Comment: No significant difference in correct responses for males. On Gender version, females show much higher proportion of correct responses (48% compared to only 10%), mostly due to decrease in number choosing A (sideways path which ignores velocity component due to applied force).

*** indicates correct answer**

ITEMS WITH SHIFTS FOR MALES:

Original FCI #22 (Gender item #26)
[rocket/person speed increasing]
Percentage of total responses each option:

	n	A	B*	C	D	E
Male, Original	*72*	32	47	1	18	1
Male, Gender	*65*	42	18	11	23	6
Female, Original	*41*	29	24	2	39	5
Female, Gender	*44*	34	25	16	18	7

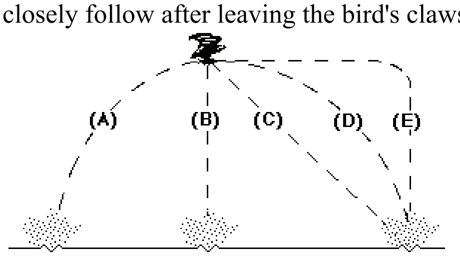
Comment: No significant difference in correct responses for females; however, number of females choosing "decreasing" speed is higher on Gender version. Males also show sharp increase in number choosing "decreasing" speed (and in those who choose "constant" speed). Net result is sharp *decrease* in correct responses by males, 47% to 18%.

Original FCI #29 (Gender item #13)
[surface force on chair/book]
Percentage of total responses each option:

	n	A	B*	C	D	E
Male, Original	*69*	33	30	4	32	0
Male, Gender	*65*	18	60	0	22	0
Female, Original	*39*	28	49	3	21	0
Female, Gender	*44*	30	57	0	11	2

Comment: No significant difference in correct responses for females. Proportion of correct responses for males *doubles* from 30% on original to 60% on Gender version. Change comes mostly from increase in proportion who now recognize presence of upward force due to surface; also, there is a decrease in number who choose "all three" forces (including air pressure).

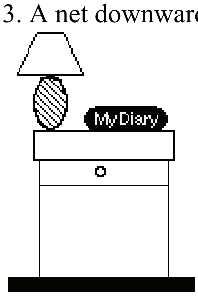
A bird is carrying a fish in its claws as it flies along in a horizontal direction above a lake. The bird accidentally drops the fish. As seen from the lakeshore, which path would the fish most closely follow after leaving the bird's claws?



Gender item #24 (Original FCI #14)

A diary is at rest on a nightstand. Which of the following force(s) is (are) acting on the diary?

1. A downward force of gravity.
2. An upward force exerted by the nightstand.
3. A net downward force exerted by the air.



(A) 1 only
(B) 1 and 2
(C) 2 and 3
(D) 1, 2, and 3
(E) none of these.
Since the book is at rest there are no forces acting on it.

Gender item #13 (Original FCI #29)

[Note: there is no diagram included on the Original FCI version of this question.]

[1]L. Enderstein and P. Spargo, "The effect of context, culture and learning on the selection of alternative options in similar situations by South African pupils," Int. J. Sci. Ed. **20**, 711 (1998).

[2]R. Steinberg and M. Sabella, "Performance on multiple-choice diagnostics and complementary exam problems," Phys. Teach. **35**, 150 (1997).

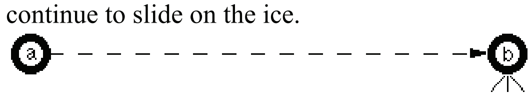
[3] N. S. Rebello and D. A. Zollman, "The effect of distracters on student performance on the Force Concept Inventory," *in review;* preprint at: http://www.phys.ksu.edu/perg/papers/

[4]J. P. Adams and T. F. Slater, "Student-supplied rationale for multiple-choice responses on the force concept inventory," *in review*; preprint at: http://www.physics.montana.edu/physed/

[5]H. Schecker and J. Gerdes, "Messung von Konzeptualisierungsfähigkeit in der Mechanik: Zur Aussagekraft des FCI," Zeitschrift für Didaktik der Naturwissenschaften **5**(1), 75-89 (1999).

USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT FOUR QUESTIONS (25 through 28).

An ice storm has knocked out power in your area and has started a fire. You have grabbed your powerful fire extinguisher and are running to help out. At point "a" you start to slip on a large patch of frictionless ice, sliding across the ice from point "a" to point "b." (Note that this diagram shows a "top view," looking down from above.) At point "b," while trying to keep upright, you accidentally turn on the fire extinguisher. The fire extinguisher produces a constant force on you in a direction at right angles to line "ab," and you slide along the ice toward point "c." When you reach point "c," you are able to turn off the extinguisher, but you continue to slide on the ice.

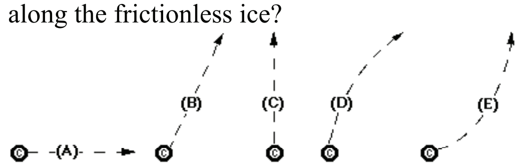


26. As you move from "b" to "c" along the ice, your **speed** is

(A) constant.
(B) continuously increasing.
(C) continuously decreasing.
(D) increasing for a while and constant thereafter.
(E) constant for a while and decreasing thereafter.

Gender item #26 (Original FCI #22)

27. At "c" the extinguisher is suddenly turned off completely. Which of the paths below will you follow beyond "c" as you continue to slide along the frictionless ice?



Gender item #27 (Original FCI #23)

[6]L. J. Rennie and L. H. Parker, "Equitable measurement of achievement in physics: high school students' responses to assessment tasks in different formats and contexts," J. Women and Minorities in Sci. Eng. **4**(2-3), 113-127 (1998).

[7]M. H. Dancy, *Investigating Animations for Assessment with an Animated Version of the Force Concept Inventory.* Ph.D. dissertation, N.C. State University (2000).

[8]L.E. McCullough and T. Foster, "A Gender Context for the Force Concept Inventory," AAPT Announcer **30**(4), 105 (2000).

Initial understanding of vector concepts among students in introductory physics courses

Ngoc-Loan Nguyen[a] and David E. Meltzer[b]
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

(Received 8 February 2002; accepted 12 March 2003)

We report the results of an investigation into physics students' understanding of vector addition, magnitude, and direction for problems presented in graphical form. A seven-item quiz, including free-response problems, was administered in all introductory general physics courses during the 2000/2001 academic year at Iowa State. Responses were obtained from 2031 students during the first week of class. We found that more than one quarter of students beginning their second semester of study in the calculus-based physics course, and more than half of those beginning the second semester of the algebra-based sequence, were unable to carry out two-dimensional vector addition. Although the total scores on the seven-item quiz were somewhat better for students in their second semester of physics in comparison to students in their first semester, many students retained significant conceptual difficulties regarding vector methods that are heavily employed throughout the physics curriculum.
[DOI: 10.1119/1.1571831]

I. INTRODUCTION

Vector concepts and calculation methods lie at the heart of the physics curriculum, underlying most topics covered in introductory courses at the university level. As Knight[1] has emphasized, the vector nature of forces, fields, and kinematical quantities requires that students have a good grasp of basic vector concepts if they are to be successful in mastering even introductory-level physics. Knight has alluded to the surprising lack of published research regarding student learning of vector concepts, and his *Vector Knowledge Test* provided an invaluable first glimpse into the pre-instruction vector knowledge of students enrolled in the calculus-based physics course. Most of the problems on the *Vector Knowledge Test* focus on algebraic aspects of vectors. Another significant investigation has been reported by Kanim,[2] who explored students' understanding of vector concepts in the context of electric forces and fields. Aguirre,[3] and Aguirre and Rankin[4] have studied students' ideas about vector kinematics, but their inquiry focused on the interrelationships among velocity, acceleration, and force rather than properties of vectors *per se*. Recently, Ortiz *et al.*[5] have reported on student learning difficulties related to basic vector operations (such as dot and cross products) as employed in introductory physics courses.

Our instructional experience has led us to believe that students' poor understanding of vector ideas posed in graphical form presents a particularly troublesome obstacle to their success in mastering physics concepts. Graphical and geometrical interpretations of vector ideas pervade the entirety of the general physics curriculum. Despite most students' previous exposure to vector concepts in mathematics courses or in high-school physics (as indicated by various surveys), and the heavy emphasis we have placed on those concepts in our own instruction, students' persistent confusion about fundamental vector notions has bedeviled our instructional efforts. We decided therefore to carry out a systematic investigation of university physics students' knowledge of basic ideas of vector addition, magnitude, and direction during the initial weeks of their physics courses. To this end, we surveyed students in both the first- and second-semester courses of the two-semester general physics sequence, both in algebra-based and calculus-based courses.

II. METHODS

We constructed a quiz containing seven vector problems posed in graphical form (see the Appendix). The problems assess whether students can correctly identify vectors with identical magnitudes and directions, and whether they can carry out vector addition in one and two dimensions. On five of the problems, students are asked to give a free response or to select multiple options from a list. On the other two (#3 and #7), they are given possible choices. On four problems students are explicitly prompted to provide explanations of their work.

This diagnostic quiz was administered to students in all introductory general physics courses taught at Iowa State University (ISU) during the 2000/2001 academic year. (We did not include in our study one-semester elementary physics courses using little or no mathematics; these courses are intended as surveys for nontechnical students.) Very minor revisions were made to the quiz between fall and spring semesters.

ISU is a large public university with a focus on engineering and technical subjects. The average ACT Mathematics score of all freshmen entering ISU in fall 2000 was 24.5, compared to the national average of 21.8 for students who completed the core college-preparatory curriculum.[6] ISU ranks 16th nationally in number of undergraduate engineering degrees awarded. It therefore seems unlikely that our results will underestimate the average performance level of physics students nationwide.

The algebra-based general physics sequence consists of Physics 111 (mostly mechanics), and Physics 112 (mostly electricity and magnetism, and optics). The calculus-based sequence is comprised of Physics 221 (mechanics, electrostatics, and dc circuits), and Physics 222 (magnetism and electromagnetism, thermal physics, optics, and modern physics). In this paper, we will use the following designations for these courses: Physics 111: A-I; Physics 112: A-II; Physics

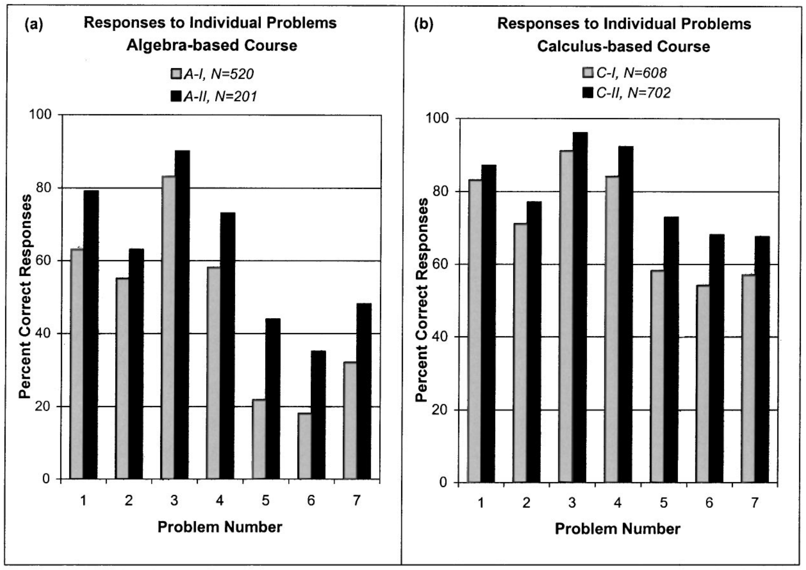


Fig. 1. Responses of students to individual problems on the vector concept diagnostic. Percent correct responses shown for students in (a) first- [A-I] and second-semester [A-II] algebra-based introductory physics; (b) first- [C-I] and second-semester [C-II] calculus-based introductory physics.

221: C-I; Physics 222: C-II, where I and II designate the first and second courses in each sequence, respectively. (That is, I is primarily a mechanics course, while II is primarily a course on electricity and magnetism. All four courses are taught during both the fall and spring semesters.) Results were obtained from a total of 2031 students, divided into the four courses as follows: Algebra-based physics: A-I, 520 total (fall: 287; spring: 233); A-II, 201 total (fall: 83; spring: 118). Calculus-based physics: C-I, 608 total (fall: 192; spring: 416); C-II, 702 total (fall: 313; spring: 389). (In the paper we refer to these courses as the ''four groups.'') Because the quiz was administered in both fall and spring offerings in all four courses during the academic year (that is, twice each in A-I, A-II, C-I, and C-II for a total of eight administrations), many students took the quiz twice, once in their fall-semester course and again in their spring-semester course. The number of repeat test-takers is not known.

We did not survey the students in this study sample with regard to their previous background in physics and mathematics. However, surveys carried out in ISU physics courses during the summer and in other years have indicated that nearly three quarters of students in the algebra-based courses, and more than 90% of those in the calculus-based courses, have studied physics in high school. In these surveys, a substantial majority of students report previous study of vectors including two-dimensional vector addition, either in their high-school physics classes or in high-school and/or college math courses. (This is the case for about two thirds of students in the algebra-based course, and about 90% of those in the calculus-based course.) These results are consistent with Knight's finding that 88% of students in the first quarter of the introductory calculus-based physics course at his institution had previous instruction on vectors. Of course, all students in the second-semester courses (that is, Physics 112 [A-II] and Physics 222 [C-II]) have had extensive exposure to vector representations and calculations in their first-semester university courses. They represent 44% of the total population sample in this study.

The quiz was administered in recitation sections (around 25 students each) during the first week of class in all four courses, before instruction on vectors took place. The quiz did not count toward a course grade and was not returned to the students. Students were asked to respond to the quiz so that instructors could get a better idea of their background knowledge in vectors. They were asked to fill in their names on the quiz to aid in record keeping. The same procedure was followed in both fall and spring semesters. Responses were obtained from the great majority of enrolled students. Responses on each problem were graded as correct or incorrect, and frequently appearing errors were noted and tabulated.

III. RESULTS

All statistical results we will cite in this paper (except for those in Sec. III C) reflect averages over the entire sample, that is, fall- and spring-semester offerings combined in the case of each of the four courses.

A. Responses to problems

Figure 1 shows the percentage of correct responses to each quiz item for students in all four courses. We now proceed to discuss the students' responses to each individual problem in more detail.

Problem #1: Vector magnitude. Performance on this problem was generally good, with a range of 63%–87% correct responses for the four different groups. However, more than one third of the students in A-I did *not* answer this question correctly, which indicates that student knowledge even on this basic vector property cannot be taken for granted. The most common error was to assume that vectors can only have equal magnitudes when they are parallel or antiparallel to each other (for example, choosing $|\vec{\mathbf{D}}|=|\vec{\mathbf{G}}|$, but not $|\vec{\mathbf{D}}|=|\vec{\mathbf{F}}|=|\vec{\mathbf{G}}|$).

Problem #2: Vector direction. A significant number of students in all classes made errors on this question (23%–45% incorrect responses). It is notable that there was very little difference in performance between students in the first- and second-semester courses, both in the algebra-based and calculus-based sequences. This small performance increment seems to suggest that, particularly on this problem, little increase in understanding occurs during the first-semester course (that is, in A-I and C-I).

The single most common incorrect response was to list both vectors $\vec{\mathbf{F}}$ *and* $\vec{\mathbf{G}}$, instead of $\vec{\mathbf{F}}$ *only*, thus reflecting confusion about the requirement that vectors with the same direction be parallel to each other. (Or, perhaps this response indicates confusion about how to *recognize* when two vectors are parallel.) This error represented 20% of all responses (almost half of all incorrect responses) in the algebra-based course, with no significant difference between A-I and A-II. However, there also were a significant number of students responding that the answer was ''none''; this category comprised 11% of all responses in the algebra-based course (one quarter of all incorrect responses in both A-I and A-II). Remarkably, those students who answered ''none'' very often asserted explicitly that all of the angles—or the ''slopes''—were different, despite the presence of the grid, which was intended to allow easy evaluation of the angles. The other option appearing with some frequency on students' responses was vector $\vec{\mathbf{C}}$, thus equating the direction of vector $\vec{\mathbf{A}}$ with that of $-\vec{\mathbf{A}}$. [It is worth noting that outside the U.S., the property we refer to as ''direction'' often is assumed to comprise two separate properties, that of ''orientation'' (line of action) and ''sense'' (loosely, ''which way it points''), see, for example, Ref. 7.]

Problem #3: Qualitative vector addition. Performance on this problem was very good for students in all courses, with correct responses in the 83%–96% correct range. However, students were not asked to provide explanations of their answer, and evidence provided by student performance on problems #4 and #5 strongly suggests that many students arrived at the correct answer for problem #3 through use of a clearly incorrect algorithm (that is, the ''split-the-difference'' algorithm to be discussed after problem #5). Because use of this algorithm reflects substantial confusion regarding vector addition, it seems probable that problem #3 does not in itself provide valid assessment of students' understanding of this vector operation.

Problem #4: One-dimensional vector addition. The students in the calculus-based courses performed very well on problem #4: C-I, 84% correct; C-II, 92% correct. However, a substantial fraction of the students in the algebra-based courses were not able to solve this problem: A-I, 58% correct; A-II, 73% correct.

In A-I, 19% of all incorrect responses consisted of a two-headed arrow as shown in Fig. 2(a); in A-II, this response was only 11% of the incorrect responses. Often this arrow was eight boxes long, but other lengths were common. Representative explanations for this response were, ''$\vec{\mathbf{R}}$ is made by connecting the end of $\vec{\mathbf{A}}$ to the end of $\vec{\mathbf{B}}$,'' and ''It is just the two vectors put together.'' Another common error in the algebra-based course (23% of all incorrect responses in A-I and A-II combined) was to show a horizontal resultant with incorrect magnitude and/or direction.

Many students produced a sloping resultant; in A-I these represented 20% of the incorrect responses, which rose to 36% in A-II. Most of these students did not show their work, but those who did typically had a diagram similar to one of those in Figs. 2(b)–2(d). Sometimes these students would explain that they were using the ''tip-to-tail'' method, or words to that effect.

Performance on problem #4 was not as good as it was on problem #3, particularly in the algebra-based courses. We suspect that, in comparison to problem #3, it may be more

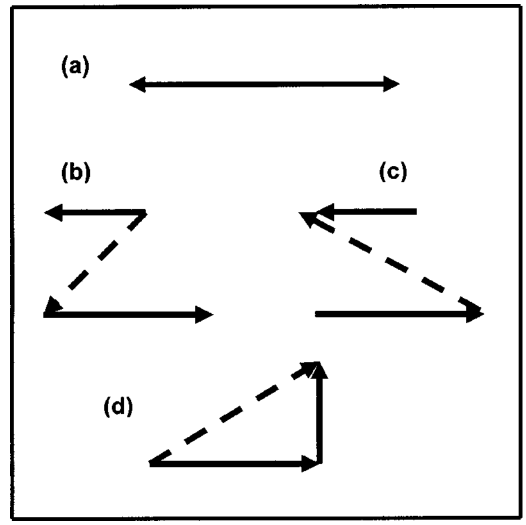


Fig. 2. Common student errors on problem #4 (addition of collinear vectors): (a) two-headed arrow; (b) tail-to-tail; (c) tip-to-tip; (d) re-orientation of top vector.

difficult to obtain a correct solution for problem #4 by using an incorrect algorithm. We will return to this issue in the discussion of problem #5.

Problem #5: Two-dimensional vector addition. The vast majority of problems in the general physics curriculum that involve vector quantities require an understanding of this basic operation. We found that most students in the calculus-

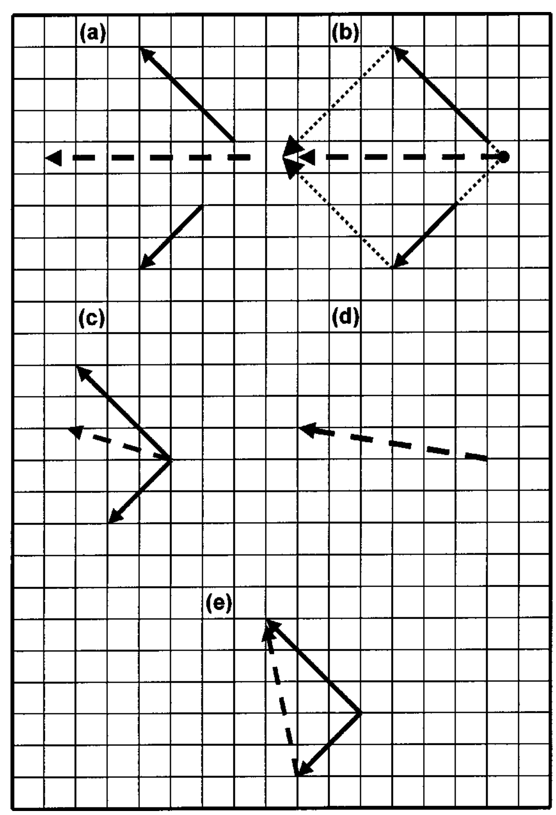


Fig. 3. Common student errors on problem #5 (addition of noncollinear vectors): (a) zero vertical component; (b) split-the-difference algorithm; (c) incorrect parallelogram addition; (d) incorrect horizontal component; (e) tip-to-tip error.

based course solved this problem correctly (58% in C-I, 73% in C-II), but only a minority of students in the algebra-based course could do so (22% in A-I, 44% in A-II).

The most common error for all four groups was to draw the resultant vector aligned along the horizontal axis (or nearly so), pointing toward the left [Fig. 3(a)]. The magnitudes of the horizontal components in this class of responses varied widely. Although some of the students who made this error were successful in determining the net horizontal component (that is, five boxes, leftward), all failed to realize that the net vertical component would be one box upward. Many students' diagrams explicitly showed the algorithm they used to obtain this result: *Join vectors* $\vec{\mathbf{A}}$ *and* $\vec{\mathbf{B}}$ *at a common vertex, and form the resultant by "splitting the difference" to obtain a net vertical component of zero* [see Fig. 3(b)]. This response was usually a clear attempt to implement a parallelogram addition rule. Some students explicitly used a very similar algorithm [see Fig. 3(c)] to obtain an apparently related error, that is, a resultant vector with the correct vertical component and pointing toward the left, but with an incorrect horizontal component. Although a particular example of this response is shown in Fig. 3(d), the magnitudes of the horizontal components represented in students' responses covered a wide range. It was not clear to us how they were able to arrive at the correct vertical component while still having an incorrect horizontal component. It seems possible that the positioning of the $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ vectors on the page—that is, one on top of the other—contributed to this outcome. It is noteworthy that in a large proportion of cases where students drew diagrams suggestive of the parallelogram addition rule, they were unsuccessful in arriving at a correct answer to this problem. Instead they produced variants of Figs. 3(b) or 3(c), or made some other error due to imprecise drawing of the parallelogram.

Most students who drew resultant vectors similar to those in Figs. 3(a) and 3(d) did not show a diagram to explain how they obtained their result. Therefore, we cannot be certain that they used the same algorithm to obtain this split-difference resultant. The proportion of the entire class that gave incorrect responses corresponding to either Fig. 3(a) or Fig. 3(d) (regardless of the horizontal component) was A-I, 42%; A-II, 29%; C-I, 21%; and C-II, 13%.

The next most common error on this problem originated from mistaken employment of a "tip-to-tip" algorithm in which the resultant vector begins at the tip of vector $\vec{\mathbf{A}}$ and ends at the tip of vector $\vec{\mathbf{B}}$ or, less often, points from the tip of $\vec{\mathbf{B}}$ to that of $\vec{\mathbf{A}}$. (This error also has been described by Knight.[1]) In this case the interpretation of students' responses was unambiguous because their diagrams explicitly showed the algorithm they had employed. There are two versions of this error: either the vectors are first brought together to a common vertex (see Fig. 3(e); this procedure actually produces the difference vector $[\vec{\mathbf{B}}-\vec{\mathbf{A}}]$), or they are left in place and the "resultant" arrow is drawn directly on the original diagram. This type of response (either version) was given by 9% of students in the algebra-based course and 6% of those in the calculus-based course, with very little difference between the I and II courses.

As was noted in connection with problems #3 and #4, the number of correct responses on problem #3 was well above that on problem #4. We now see that it was also far higher than the correct response rate on problem #5. In view of the obvious route for obtaining a correct answer to problem #3 by using the incorrect "split-the-difference" algorithm, we now believe that problem #3 is not a valid indicator of students' knowledge of vector addition.

Problem #6: Two-dimensional vector subtraction. In principle this problem could be solved with the same algorithms used for problem #5, combined with some algebraic manipulation and knowledge of how to form $-\vec{\mathbf{A}}$ from $\vec{\mathbf{A}}$. However, students probably have less practice with a specific algorithm for carrying out vector subtraction, compared with vector addition. That is, students may have memorized "place the tail of one to the tip of the other" as an addition algorithm without gaining enough understanding to extend this idea to a similar problem posed as a subtraction. One might therefore expect that performance on problem #6 would be inferior to that on problem #5, and indeed it was. However the difference was generally rather small: only 4–5% fewer correct in the calculus-based course, and 4% and 9% fewer, respectively, in A-I and A-II. Overall, error rates on problem #6 ranged from 32% incorrect in C-II, up to 82% incorrect in A-I.

In the calculus-based course (both C-I and C-II combined), 83% of the students who answered problem #5 correctly also answered problem #6 correctly. Similarly, 89% of those who answered problem #6 correctly also answered problem #5 correctly. (There was no significant difference between C-I and C-II students regarding this pattern.) This response pattern suggests that for students in the calculus-based course, problem #5 and problem #6 provide a roughly equivalent indication of students' understanding of two-dimensional vector addition.

By contrast, in the algebra-based course, only 67% of students who answered problem #5 correctly also answered problem #6 correctly. Of the students who answered problem #6 correctly, 83% also solved problem #5. (Again, there was no significant difference between A-I and A-II.) So, for students in the algebra-based course, problem #6 was indeed significantly more difficult than problem #5 ($p<0.01$ according to a z test for difference between correlated proportions[8]). In this case the two problems did not provide equivalent indications of students' knowledge, because a correct solution to problem #6 was correlated with superior performance on this two-problem subset.

There were a wide variety of incorrect responses to problem #6. Many students' explanations made it clear that they were trying to find a $\vec{\mathbf{B}}$ such that $\vec{\mathbf{R}}$ would be the "average," in some sense, of $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$. However, lacking an algorithm for this purpose, students often resorted to guessing or estimating the direction of vector $\vec{\mathbf{B}}$. A common response was to draw $\vec{\mathbf{B}}$ as a horizontal vector (vertical component=0) pointing to the right; one-quarter of all incorrect responses were of this type in both algebra-based and calculus-based courses (algebra based, 26%; calculus based, 25%). These vectors were drawn either with their tails in contact with the tail of $\vec{\mathbf{A}}$ or, more often, as isolated vectors in the blank grid space to the right of $\vec{\mathbf{A}}$ and $\vec{\mathbf{R}}$. Most students did not explain their reasoning, but some offered clear descriptions of their thinking such as "$\vec{\mathbf{R}}$ should be a combination of $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ so I tried to put it between $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$"; "The magnitude of $\vec{\mathbf{B}}$ and $\vec{\mathbf{A}}$ are equal, so the direction of the resultant is directly between the two." Overall, a large majority of students with incorrect

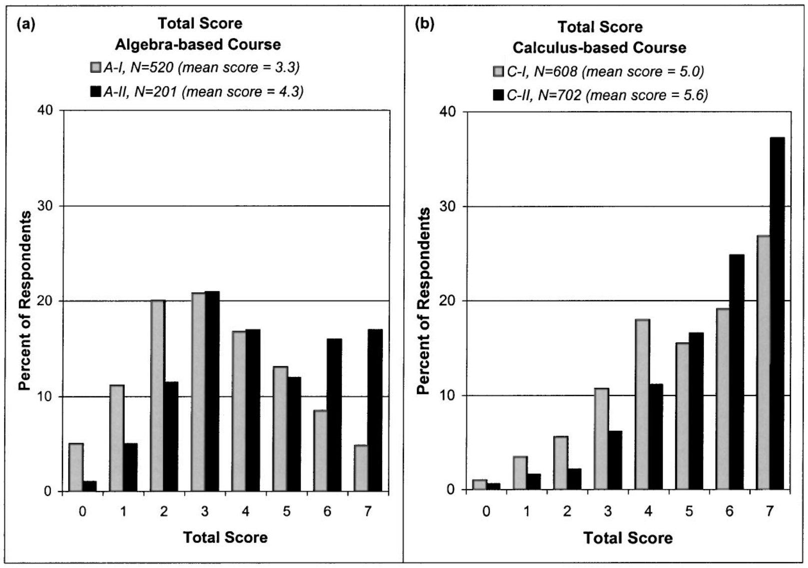


Fig. 4. Distribution of total scores on vector concept diagnostic in percent of class obtaining a particular score (score range: 0–7): (a) first- [A-I] and second-semester [A-II] algebra-based introductory physics; (b) first- [C-I] and second-semester [C-II] calculus-based introductory physics.

responses to this problem realized that $\vec{\mathbf{B}}$ would have a positive horizontal component, but were unable to determine its precise value.

Problem #7: Comparison of resultant magnitude. This problem is another application of vector addition for which students are unlikely to have memorized a specific algorithm. With no grid available, students do not have at hand as straightforward a calculation procedure as might be employed in problems #5 and #6. However, only a qualitative response is required on problem #7, while a precise quantitative answer is needed for problem #5; moreover, there are only three possible choices. This smaller selection of options may mitigate the additional challenge posed by problem #7 (if there is any). In any case, the only group for whom performance on problems #5 and #7 differed by more than 5% was students in A-I; they achieved 32% correct on problem #7 compared to only 22% correct on problem #5. However, it is interesting to note that 23% of the C-II students who successfully solved problem #5 also gave incorrect responses to problem #7. It seems that the apparently superior algorithmic skill of the C-II students did not always translate to a situation in which a grid was lacking.

Many students who chose the correct (‘‘smaller than’’) response in problem #7 gave a satisfactory explanation of their answer, often accompanied by a diagram that reflected use of the parallelogram or tip-to-tail addition rules to demonstrate that $|\vec{\mathbf{R}}_{\mathbf{A}}|<|\vec{\mathbf{R}}_{\mathbf{B}}|$. Among those students who gave incorrect answers, there was a preference for the ‘‘equal to’’ response (that is, magnitude of resultant of pair **A** is equal to that of pair **B**), very often justified by an explanation such as ‘‘the vectors in **A** and **B** are equal magnitude,’’ and sometimes accompanied by an invalid application of the Pythagorean formula to pair **B**. The ratio of ‘‘equal to’’ responses in comparison to ‘‘larger than’’ responses was almost exactly 1:1 in A-I, but in A-II the ‘‘equal to’’ response jumped in popularity to nearly a 2:1 ratio compared to ‘‘larger than.’’ In both C-I and C-II, the ‘‘equal to’’ response was the more common incorrect response by nearly a 3:2 ratio. The ‘‘larger than’’ response was justified by the larger vertex angle or the ‘‘larger area covered’’ in diagram A. Explanations such as these were typical: ‘‘**A** is larger because arrows are further apart’’; ‘‘**A** larger, the angle is greater between the vectors’’; ‘‘larger than because both vectors are farther apart than the ones in **B**.’’

B. Total score comparisons

The distribution of students’ total score on the diagnostic (maximum=7.0) is shown in Fig. 4. The scores in A-I are fairly normally distributed around a mean value of 3.3, while the A-II distribution (mean score=4.3) is somewhat bimodal. The distributions in the calculus-based course are very strongly skewed toward higher scores (although that in C-I is also somewhat bimodal). Mean scores for the calculus-based course are C-I: 5.0; C-II: 5.6. These distributions suggest that the diagnostic is a good reflection of the mean level of knowledge of students in the algebra-based courses, whereas the average level of vector knowledge of students in the calculus-based courses goes beyond that characterized by this diagnostic.

C. Differences in performance between fall- and spring-semester courses

We were surprised to find that on many of the quiz items, there appeared to be a significant difference in performance between students in the fall and spring offerings of the *very* same course (for example, the fall and spring offerings of A-I). Students enrolled in C-I during the spring semester of 2001 had higher scores on all seven quiz items than students in the fall 2000 semester of the same course. The mean scores (percent correct out of seven problems; s.d.=standard deviation) were: spring, 2001 ($N=416$): 74% correct (s.d. =25%); fall, 2000 ($N=192$): 65% correct (s.d.=27%). The difference in mean scores is statistically significant at the $p=0.0003$ level according to a two-sample t-test. A very similar fall–spring discrepancy was found for students in A-I (spring, 51%; fall, 44%; $p<0.001$). For C-II there was a smaller but still statistically significant superiority, this time however in the *fall* semester mean scores (fall, 83%; spring, 78%, $p<0.01$) while in A-II, the fall–spring difference in mean scores was very small and not statistically significant.

For C-II, on a three-item group of closely related problems (problems #5, #6, and #7), fall performance was significantly better (fall, 76%; spring, 64%; $p < 0.001$ according to a chi-square test). In A-II, a fall–spring difference on the same three-item group was again present (fall, 49%; spring, 38%), but did not quite rise to the level of statistical significance ($p = 0.12$), perhaps because of the relatively small sample size.

Although it seems clear that the discrepancy in performance between students in the fall- and spring-semester offerings of A-I and C-I is not due to chance—and the same may be true for the inverse effect observed in A-II and C-II—we do not have data that would allow us to determine the cause. Many factors might contribute (for example, students repeating courses, advanced students preferring ''off-sequence'' offerings, etc.), but at this point we can only speculate on this matter.

IV. DISCUSSION

The concepts probed in this diagnostic are among the most basic of all vector ideas. Students are assumed to have a good understanding of them throughout all but the first week or two of the introductory physics curriculum. Although a very brief (less than one lecture) discussion related to these concepts is usually provided near the beginning of the first-semester course, students often are assumed to have been exposed to vector ideas either in their mathematics courses or in high-school physics, with the further assumption that very little review is needed. The emphasis of the discussion and use of vector concepts in the college-level physics course is decidedly on the algebraic aspects and is directed toward calculational competence. As a consequence, graphical and geometrical interpretations of vector operations may be somewhat neglected.

(As a point of reference on this issue, we note that of the seven high-school physics textbooks surveyed in a recent study,[9] all but one[10] cover vector concepts to some extent, including one- and two-dimensional vector addition presented in graphical form. Most of these texts go into considerable detail. No doubt the actual extent of vector coverage in high-school physics courses varies very widely throughout the nation.)

We found that a significant proportion of students in our sample had serious conceptual confusion related to basic vector concepts represented in graphical form, even though surveys suggest that most of them had previous instruction in vectors. (More than 44% of students in our sample had taken at least one full semester of university-level mechanics.) Even in the second semester of the calculus-based physics course (that is, C-II)—in which students are assumed from the very first day to have considerable expertise with vector methods—more than one-quarter of the class could not carry out a two-dimensional vector addition. Our data from the second semester of the algebra-based course (that is, A-II) suggest that the majority of students in the first semester of this course (A-I) never successfully mastered this operation. This finding should have rather sobering implications for instructors who assume that, for example, students beginning study of electric field superposition are competent with vector addition.

On many of our quiz items, improvements in student performance from first to second semester were small or practically nonexistent, indicating that little learning of the ideas had taken place during the first-semester course. This small performance improvement was observed for both algebra-based and calculus-based courses. It seems that the bulk of students' basic geometrical understanding of vectors was brought with them to the beginning of their university physics course and was little changed by their experiences in that course, at least during the first semester.

It seemed clear that, although most students were unable to solve two or more of the problems, they did have some degree of basic knowledge which they attempted to apply to the problems they missed. For instance, there often were efforts to apply a tip-to-tail rule or a parallelogram addition rule which were unsuccessful due to imprecise execution. Frequently, students did not accurately copy the magnitude and/or the direction of the vectors they were attempting to add. Often, they were uncertain as to which ''tail'' was supposed to be in contact with which ''tip.''

Many students had an intuitive feel for how vectors should add which, it was clear, was based on their experience with forces. Although the word ''force'' is not used in the quiz, many students referred to the vectors as ''forces'' and used dynamical language to describe their thinking, such as how one vector was ''pulling'' the other in a certain direction, or how the ''pulls'' of two vectors would balance out. In many cases students were able to estimate the approximate direction of a resultant without being able to give a correct quantitative answer.

It seemed to us that many of the students' errors could perhaps be traced to a single general misunderstanding, that is, of the concept that vectors may be moved in space in order to combine them as long as their magnitudes and directions are exactly preserved. We suspect that, to some extent, this misunderstanding results in part from lack of a clear concept of how to determine operationally a vector's direction (through slope, angle, etc.)

As mentioned in Sec. I, very few reports of students' vector understanding have been published. We may make direct comparison, however, with the results reported by Knight[1] for problem 5 of his *Vector Knowledge Test*. This problem is very similar to problem #5 on our own quiz. Knight found that 43% of students in the first-quarter calculus-based course at California Polytechnic State University, San Luis Obispo, were able to answer that problem correctly. This statistic may be compared to the 58% correct response rate we observed on problem #5 in the first-semester calculus-based course (C-I) at ISU. Although the difference is statistically significant it is not particularly large, and might be accounted for by slight differences both in the test problems and in the student populations.

Another comparison we may make is to the results reported by Kanim on a problem involving net electrical force on a charge;[11] this problem is similar to our problem #7. He reports that 70% of students in a second-semester calculus-based course at the University of Illinois gave a correct response to that question, nearly identical to the 68% correct response rate to problem #7 in our second-semester calculus-based course (C-II). Kanim reports similar results on related problems among students at the University of Washington and elsewhere.

V. CONCLUSION

In previous investigations, Knight[1] and Kanim[2] have documented a variety of serious student difficulties with both algebraic and graphical aspects of vector concepts among students in introductory physics courses at several institutions similar to our own. Their results and ours consistently support a conclusion that significant additional instruction on vectors may be needed if introductory physics students are to master those concepts. We suspect that most instructors would be unsatisfied with a situation in which more than half of the students are still unable, after a full semester of study, to carry out two-dimensional vector addition (as we found to be the case in the algebra-based course).

It is clear from our findings that many students have substantial intuitive knowledge of vectors and vector superposition, obtained to some extent by study of mechanics, and yet are unable to apply their knowledge in a precise and therefore fruitful manner. They seem to lack a clear understanding of what is meant by vector direction, of how a vector may be "moved" so long as its magnitude and direction are strictly preserved, and of exactly how to carry out such moves by parallel transport. Many students are confused about the tip-to-tail and parallelogram addition rules.

One way in which vector addition may be introduced is through the use of displacement vectors, because students all have experiences that could allow understanding of how a 50-m walk to the east and subsequent 50-m walk to the north is equivalent to a 71-m walk to the northeast. Students could be guided to determine similar equivalent displacements—perhaps initially by using a grid—when the component displacements are at arbitrary angles. In order to solidify the notion of vector addition, it also would be important for students to practice applying these methods when no grid or other means for quantitative measurement is available. Many of the responses by students in our study (in particular, to problem #7) suggest that an ability to solve vector problems when a grid is available do not always translate to a similar ability in the absence of a grid. Recent interviews carried out by our group lend support to this observation.[12] We believe that curricular materials that guide students through a series of exercises in which they perform vector additions and subtractions (both with and without use of a grid) may be useful in improving their understanding of these ideas.

Further research will be needed to determine whether curricular materials based on such a strategy are effective in improving both students' performance on assessments such as the quiz used in our study, and students' ability to provide explanations of their work with precision (describing a clearly delineated calculational procedure) and accuracy (describing a *correct* calculational procedure). Additional research (such as that initiated by Oritz *et al.*[5]) is necessary to probe students' understanding of more advanced vector concepts such as scalar and vector products.

As a consequence of our findings, we have increased the amount of instructional time we devote specifically to vector concepts. We have developed some instructional materials[13] in a format similar to the problems on our diagnostic quiz, and continue development and assessment of additional materials. Our group has carried out a preliminary series of student interviews to shed additional light on student understanding of vector concepts. We are also extending our research to assess students' understanding of more advanced concepts, such as scalar and vector products, coordinate systems and rotations, etc. In addition, we are examining student understanding of vector ideas, specifically in the context of physics concepts such as superposition of forces and fields.

ACKNOWLEDGMENTS

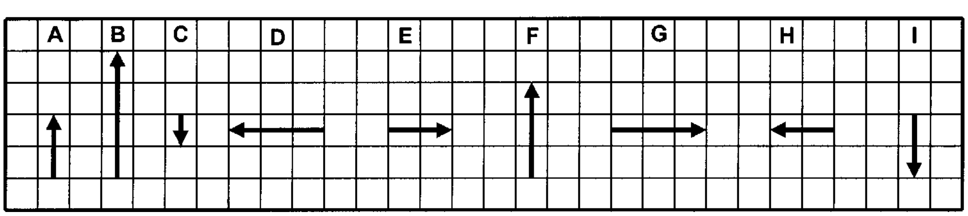
We are grateful for the assistance of Larry Engelhardt, both for his collaboration in the data analysis and for the insight he provided based on the student interviews he has recently carried out. This material is based in part upon work supported by the National Science Foundation under Grant No. REC-0206683.

APPENDIX: VECTOR CONCEPT QUIZ

Name: ____________________

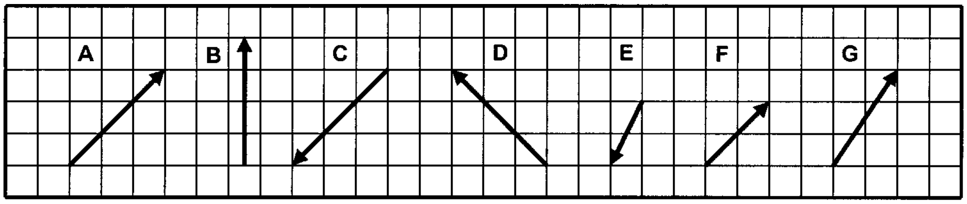
Class: ____________________ Section: ____________________

1. Consider the list below and write down **all** vectors that have the same magnitudes as each other. For instance if vectors $\vec{\mathbf{W}}$ and $\vec{\mathbf{X}}$ had the same magnitude, and the vectors $\vec{\mathbf{Y}}$, $\vec{\mathbf{Z}}$, and $\vec{\mathbf{A}}$ had the same magnitudes as each other (but different from $\vec{\mathbf{W}}$ and $\vec{\mathbf{X}}$) then you should write the following: $|W|=|X|$, $|Y|=|Z|=|A|$.



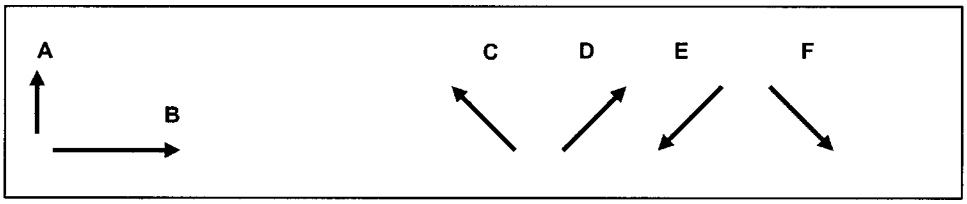
Answer ____________________

2. List all the vectors that have the same **direction** as the first vector listed, $\vec{\mathbf{A}}$. If there are none, please explain why.



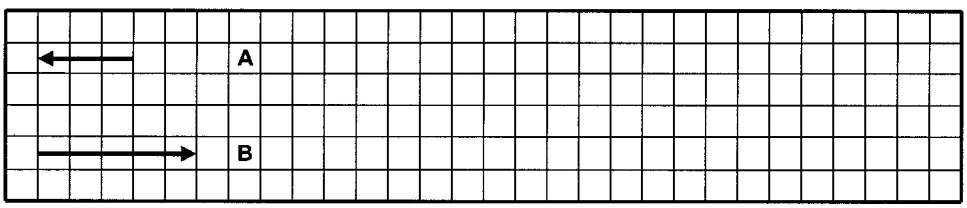
Explain ______________________________

3. Below are shown vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$. Consider $\vec{\mathbf{R}}$, the vector sum (the "resultant") of $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$, where $\vec{\mathbf{R}}=\vec{\mathbf{A}}+\vec{\mathbf{B}}$. Which of the four other vectors shown ($\vec{\mathbf{C}},\vec{\mathbf{D}},\vec{\mathbf{E}},\vec{\mathbf{F}}$) has most nearly the ***same direction*** as $\vec{\mathbf{R}}$?



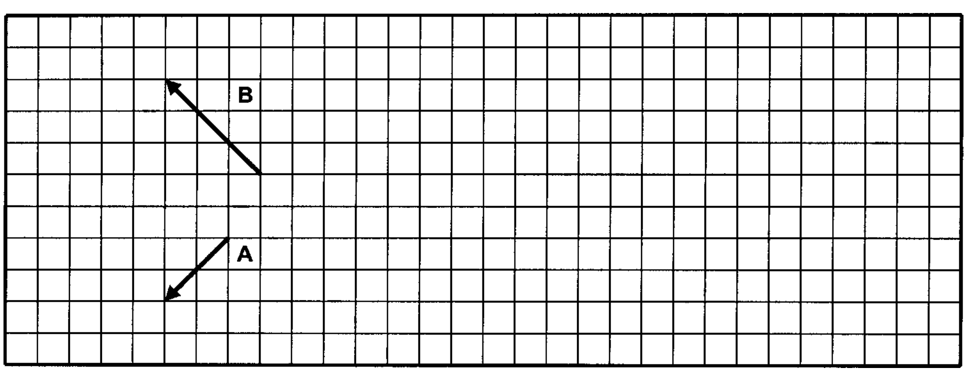
Answer ______________________________

4. In the space to the right, draw $\vec{\mathbf{R}}$ where $\vec{\mathbf{R}}=\vec{\mathbf{A}}+\vec{\mathbf{B}}$. Clearly label it as the vector $\vec{\mathbf{R}}$. Explain your work.

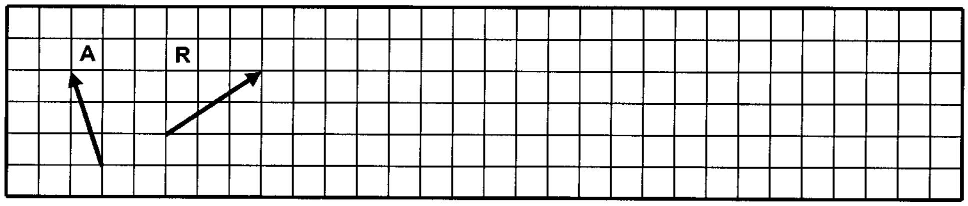


Explain ______________________________

5. In the figure below there are two vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$. Draw a vector $\vec{\mathbf{R}}$ that is the sum of the two, (i.e., $\vec{\mathbf{R}}=\vec{\mathbf{A}}+\vec{\mathbf{B}}$). Clearly label the resultant vector as $\vec{\mathbf{R}}$.



6. In the figure below, a vector $\vec{\mathbf{R}}$ is shown that is the ***net resultant*** of two other vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ (i.e., $\vec{\mathbf{R}}=\vec{\mathbf{A}}+\vec{\mathbf{B}}$). Vector $\vec{\mathbf{A}}$ is given. Find the vector $\vec{\mathbf{B}}$ that when added to $\vec{\mathbf{A}}$ produces $\vec{\mathbf{R}}$; clearly label it $\vec{\mathbf{B}}$. **DO NOT** try to combine or add $\vec{\mathbf{A}}$ and $\vec{\mathbf{R}}$ directly together! Briefly explain your answer.



Explain ______________________________________

____

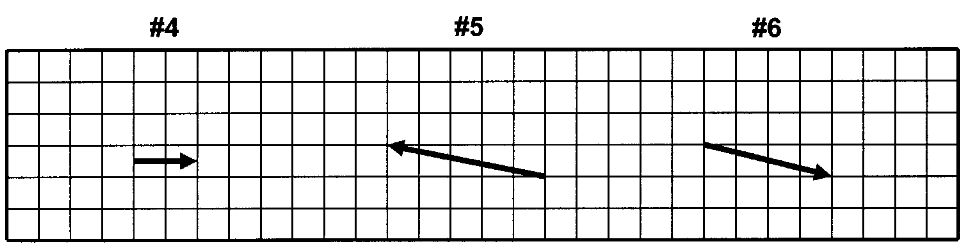
7. In the boxes below are two pairs of vectors, pair **A** and pair **B**. (All arrows have the same length.) Consider the magnitude of the ***resultant*** (the vector sum) of each pair of vectors. Is the magnitude of the resultant of pair **A** *larger than, smaller than*, or *equal to* the ***magnitude*** of the resultant of pair **B**? Write an explanation justifying this conclusion.

Explain ______________________________________

____

Problem solutions:

1. $|A|=|E|=|H|=|I|$, $|D|=|F|=|G|$
2. **F**
3. **D**



7. smaller than.

[a]Electronic mail: nguyenn@iastate.edu
[b]Electronic mail: dem@iastate.edu

[1]Randall D. Knight, "Vector knowledge of beginning physics students," Phys. Teach. **33**, 74–78 (1995).
[2]Stephen Emile Kanim, "An investigation of student difficulties in qualitative and quantitative problem solving: Examples from electric circuits and electrostatics," Ph.D. dissertation, University of Washington (UMI, Ann Arbor, MI, 1999), UMI #9936436, Chaps. 6 and 7.
[3]Jose M. Aguirre, "Student preconceptions about vector kinematics," Phys. Teach. **26**, 212–216 (1988).
[4]Jose M. Aguirre and Graham Rankin, "College students' conceptions about vector kinematics," Phys. Educ. **24**, 290–294 (1989).
[5]Luanna G. Ortiz, Paula R. L. Heron, Peter S. Shaffer, and Lillian C. McDermott, "Identifying student reasoning difficulties with the mathematical formalism of rotational mechanics," AAPT Announcer **31**(4), 103 (2001).
[6]ACT, Inc., ACT research data at ⟨www.act.org⟩.
[7]John Roche, "Introducing vectors," Phys. Educ. **32**, 339–345 (1997).
[8]J. P. Guilford, *Fundamental Statistics in Psychology and Education*, 4th ed. (McGraw-Hill, New York, 1965), pp. 188–189.
[9]T. Entwistle, D. Gentile, E. Graf, S. Hulme, J. Schoch, A. Strassenburg, C. Swartz, C. Chiaverina, R. B. Clark, T. Durkin, D. Gavenda, F. Peterson, C. Robertson, and R. Sears, "Survey of high-school physics texts," Phys. Teach. **37**, 284–296 (1999).
[10]Arthur Eisenkraft, *Active Physics* (It's About Time, Armonk, NY, 1998).
[11]See Ref. 2, p. 156, Fig. 6-8.
[12]Larry Engelhardt (private communication).
[13]David E. Meltzer and Kandiah Manivannan, "Workbook for introductory physics, additional materials: Vector exercise," CD-ROM available from DEM.

Student learning of physics concepts: efficacy of verbal and written forms of expression in comparison to other representational modes*

David E. Meltzer, Assistant Professor
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

Abstract

Physics instruction includes a variety of representational modes including diagrammatic, mathematical/symbolic, and verbal (oral and written passages employing ordinary language). Instructors attempt to assess students' understanding by observing their problem-solving performance employing this variety of representational modes. An important issue that this study investigated is the possible discrepancies in student learning abilities when using oral and written forms of expression in comparison to diagrammatic and mathematical forms. Another issue explored is the accuracy of assessment of student learning via written and oral descriptions of their reasoning, in comparison to their mathematical/symbolic problem-solving performance.

*Supported in part by National Science Foundation Grants DUE-#9354595, DUE-#9650754, DUE-#9653079, DUE-#9981140, and REC-#0206683.

This paper appeared in *Conference on Ontological, Epistemological, Linguistic and Pedagogical Considerations of Language and Science Literacy: Empowering Research and Informing Instruction*, University of Victoria, Victoria, British Columbia, Canada, September 13, 2002; <http://www.educ.uvic.ca/faculty/lyore/sciencelanguage/>.

I. Introduction

The goal of this investigation is to probe the role played by diverse representational modes in the learning of physics concepts. We explore the relationship between the *form of representation* of concepts in physics, and students' ability to learn these concepts. We are attempting to determine the specific learning difficulties that arise as students struggle to master concepts posed in different representational forms, and we plan to apply our findings to the development of improved curricular materials and instructional methods. The particular focus of this paper is to compare student thinking when using a "verbal" form of representation (written or oral) to the thinking that is manifested with other forms of representation such as mathematical, diagrammatic, and graphical.

Much previous research has shown that the use of multiple forms of representation in teaching concepts in physical science has great potential benefits, and yet poses significant challenges to students and instructors. Facility in the use of more than one representation deepens a student's understanding, but specific learning difficulties arise in the use of diverse representational modes.

By "representational mode" we mean any of the widely diverse forms in which physical concepts may be understood and communicated. For instance, problems or principles may be stated in verbal form, using words only, or purely in mathematical form, using equations and special symbols. As an example of the use of diverse representational modes, consider Coulomb's law. In Quiz #11 shown on page 16, we present four different representations of what is essentially the *identical* problem. These are posed in four distinct representational modes – verbal (#1), diagrammatic (#2), mathematical/symbolic (#3), and graphical (#4). Although to the expert these four problems are nearly identical and merely represent four different aspects of the same concept, to an introductory student they may appear very different.

What we are concerned with here are (1) common, widespread learning difficulties encountered by *many* students, and (2) the *relative degree of difficulty* of different representations in a specific context. It is often assumed by instructors that a representation that they find particularly clear and comprehensible (e.g., a graph) will also be especially clear for the average student. Research and experience shows that this is often *not* the case, but relatively little study has been devoted to this issue.

In the remainder of this paper, some preliminary results of this investigation will be presented. In Section II, I discuss some of the well-known learning difficulties that are associated with technical terms in physics that also carry meanings in "ordinary" language that diverge widely from their physics definitions. In Section III, I describe an example of a related, though somewhat distinct problem: students' alternative interpretations of words in ordinary language that have specific and precise meanings when they are employed in a technical context. In Section IV, I present results of several different probes of students' ability to interpret and respond to physics questions when posed (nearly) simultaneously in a variety of diverse representational modes.

II. Confusion due to technical terms with "everyday" meanings

It is well known that numerous technical terms in physics have everyday meanings that are very different from their "physics" definitions. The physics concepts represented by these terms are, in themselves, difficult for most students to grasp. The fact that students are burdened with alternative meanings and connotations for these words, drawn from their day-to-day experiences, significantly adds to the difficulty of learning these concepts. A few of the most prominent terms in this category are these:

force: Although the ordinary meaning of force in the sense of "push" or "pull" is in itself not misleading from the technical physics standpoint, the vector nature of forces – that is, that each force is characterized by a precise magnitude and direction – is not always appreciated by introductory students. Moreover, everyday connotations of *force* such as "energy" or "power" can be extremely misleading to students (Williams, 1999), and the mistaken impression that a force is an entity in itself – rather than an *interaction* between two objects – can make it difficult for students to grasp what is, from the physicist's standpoint, the most significant characteristic of the force concept (Touger, 1986, 1991).

power: In everyday language this word is often taken to mean "energy" (or sometimes "force"), while its precise physics meaning as *energy per time* is frequently obscured. This confusion can be particularly troublesome in the context of electricity, where the word *power* is confused not only with "energy," but often with both "current" and "voltage" (see discussion below.)

current/voltage: Most introductory students make little or no distinction between the meanings of *current* and *voltage*¸ and often confuse *power* with both of these two. All three terms are broadly conceived as connoting a form of electrical "energy," which may help explain the extremely widespread student misconception that a battery always supplies the same current regardless of the specific circuit in which it is placed. The precise physics definitions of *current* (charge flow per time), *voltage* (electric potential difference), and *power* (energy per unit time) are among the most difficult to communicate to introductory students (McDermott and Shaffer, 1992; Shaffer and McDermott, 1992).

work: The everyday notion of *work* as implying "exertion" is an impediment to grasping the physics definition, in which *displacement* of an object acted upon by a force is required in order to qualify for nonzero work. The fact that the *work* done on an object in a physics sense can be either positive or negative – depending on whether the object's kinetic energy is increased or decreased, respectively – has proven to be a particularly difficult concept to communicate to introductory students (Loverude, Kautz, and Heron, 2002). In the context of thermodynamics, difficulty in grasping the distinctions among *work*, *heat* and *internal energy* is a major obstacle to students' understanding of the first and second laws of thermodynamics (Loverude, Kautz, and Heron, 2002; Meltzer, 2001, 2002). In part, this is due to the fact that all three quantities are measured in the same (energy) units (Meltzer, 2002).

heat: In physics, heat (or "heat transfer") is a process-dependent variable and represents a *transfer* of a certain amount of energy between systems due to their temperature difference. However, among beginning science students, heat is frequently viewed as an intensive quantity –

that is, as a mass-independent *property* of an object – and temperature is interpreted as degree of heat, that is, as a measure of its intensity. Alternatively, heat is often interpreted as a specific quantity of energy *possessed* by a body (an extensive quantity), with temperature being the measure of that quantity (Zemansky, 1970; Kesidou *et al.*, 1995; Greenbowe and Meltzer, 2002). This confusion is not restricted to the English language, for terms equivalent to *heat* in other languages such as *Wärme* [German] (Berger and Wiesner, 1997) and *chaleur* [French] (Tiberghien and Delacôte, 1978) have been associated with similar pedagogical difficulties.

III. Confusion due to ambiguous meaning of words used in a technical sense: Example of word "*constant*"

There are many instances where certain words – although they are not in a strict sense technical terms – have a specific interpretation in a technical context that can easily be misunderstood by the student. For example, in physics it is extremely common to speak of "constant" values for some variable. This means that some quantitative measure for that variable has an unchanging magnitude, characterized by a specific number in some unit system. An object moving in one direction that has a "constant" acceleration is one whose speed changes by the same amount during each second. Such an object (if its mass does not vary) must be subject to a net force whose direction and magnitude do not change with time.

Williams (1999) has argued that use of the word "constant" could improve the precision of a particular statement of Newton's first law, viz.,

> Every body continues in its state of rest or of uniform speed in a straight line unless it is compelled to change that state by forces acting on it.

Williams states:

> Two alternative word choices could improve the precision of this statement:
>
> (1) replacing the adjective "uniform" {*consistent in conduct or opinion; having always the same form, manner, or degree; not varying or variable*} by "constant" {*something invariable or unchanging: as a number that has a fixed value in a given situation or universally…*} moves from a word of everyday speech with its accompanying vagueness to a familiar and more precise word in common use in mathematics; (Williams, 1999, p. 675)

However, although the word "constant" does indeed have a precise mathematical meaning, it is not necessarily the case that this meaning is the one that will be imputed to it by the typical student. This became evident during the course of a lengthy post-instruction interview with a student in an elementary physics course. This student had just completed a hands-on, inquiry-based elementary course in which kinematics and Newtonian dynamics were the central concepts discussed throughout the course.

The student was explaining her answers to a series of questions involving a sled being pushed along a frictionless, icy surface. A person wearing spiked shoes is pushing the sled. The first question was,

> Which force would keep the sled moving toward the right and speeding up at a steady rate (constant acceleration)?

Among the answer options were:

> The force is toward the **right** and is of **constant** strength (magnitude).
>
> The force is toward the **right** and is **increasing** in strength (magnitude).
>
> The force is toward the **right** and is **decreasing** in strength (magnitude).

[Emphasis in original; first statement is correct]

I repeated the question and asked the student to explain her answer:

DEM: Suppose she is speeding up at a steady rate with constant acceleration. In order for that to happen, do you need to apply a force? And if you need to apply a force, what kind of force: would it be a constant force, increasing force, decreasing force?

STUDENT: Yes you need to have a force. It can be a constant force, or it could be an increasing force.

DEM: . . . She is speeding up a steady rate with constant acceleration.

STUDENT: Constantly accelerating? Then the force has to be increasing . . . Wait a minute . . .The force could be constant, and she could still be accelerating.

DEM: Are you saying it could be both?

STUDENT: It ***could*** *be both, because if the force was increasing she would still be constantly accelerating.*

DEM: What do we mean by constant acceleration?

STUDENT: Constantly increasing speed; a constant change in velocity.

It seems evident that the student is interpreting the meaning of the word "*constant*" not as "unchanging," but rather as "persistent" or "ever-present." Its precise quantitative connotation appears to be lost on her.

IV. Multiple representations of knowledge: student understanding of "verbal" representation contrasted with understanding of other forms of representation (mathematical/symbolic; graphical; pictorial/diagrammatic)

A. "Ordinary Language" vs. Graphical Representation

A major focus of our recent work has been to explore the question of whether students' ability to learn specific physics concepts may be greater when using one form of representation, rather than another. The origin of our interest in this question was the inquiry-based elementary physics course referred to above. After the introduction of microcomputer-based laboratory tools, we found that students' ability to give correct responses to questions involving Newtonian dynamics posed in graphical form seemed to have significantly increased. However, when the questions were posed in the form of "ordinary" language, no corresponding improvement was evident (Meltzer *et al.*, 1997).

Evidence for this discrepancy was provided by students' responses to questions drawn from the "Force and Motion Conceptual Evaluation" (Thornton and Sokoloff, 1998). A set of nearly identical questions related to Newton's second law are given first in ordinary language in the form of the "Force Sled" questions (see next page), and later in the form of "Force Graph" questions (following page). The only significant difference between the questions is that the first set is posed in verbal representation, while the second uses a graphical representation. Students enrolled in this physics course had literally dozens of hours of practice, both in class and on homework assignments, with very similar questions posed in both formats.

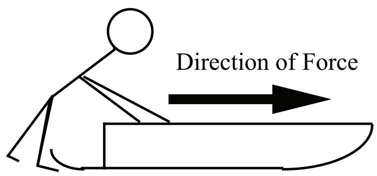
These question sets were administered post-instruction in two separate offerings of this course. A total of 18 students responded to the questions. The results are shown in the table below:

	Correct Responses, Post-instruction ($N = 18$)
Force Graph questions	56%
Force Sled questions (#1-4)	28%

In view of the great similarity of the question sets, such a large difference in correct response rates – consistent over two separate course offerings – was surprising. (A test for comparison of binomial proportions yields $p = 0.09$, marginally significant, but probably reflective of the relatively low sample size.)

A sled on ice moves in the ways described in questions 1-4 below. *Friction is so small that it can be ignored.* A person wearing spiked shoes standing on the ice can apply a force to the sled and push it along the ice. Choose the one force (**A** through **G**) which would **keep the sled moving** as described in each statement below.

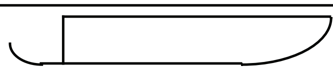
You may use a choice more than once or not at all but choose only one answer for each blank. If you think that none is correct, answer choice **J**.



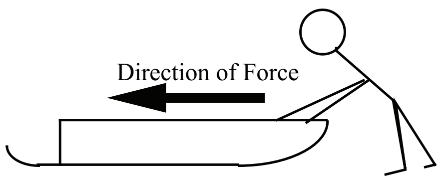
A. The force is toward the **right** and is **increasing** in strength (magnitude).

B. The force is toward the **right** and is of **constant** strength (magnitude).

C. The force is toward the **right** and is **decreasing** in strength (magnitude).



D. No applied force is needed



E. The force is toward the **left** and is **decreasing** in strength (magnitude).

F. The force is toward the **left** and is of **constant** strength (magnitude).

G. The force is toward the **left** and is **increasing** in strength (magnitude).

_______1. Which force would keep the sled moving toward the right and speeding up at a steady rate (constant acceleration)?

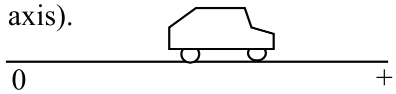
_______2. Which force would keep the sled moving toward the right at a steady (constant) velocity?

_______3. The sled is moving toward the right. Which force would slow it down at a steady rate (constant acceleration)?

_______4. Which force would keep the sled moving toward the left and speeding up at a steady rate (constant acceleration)?

"Force Sled" Questions from the Force and Motion Conceptual Evaluation (Thornton and Sokoloff, 1998).

Questions 14-21 refer to a toy car which can move to the right or left along a horizontal line (the positive part of the distance axis).



Assume that friction is so small that it can be ignored.

A force is applied to the car. Choose the one force graph (**A** through **H**) for each statement below which could allow the described motion of the car to continue.

You may use a choice more than once or not at all. If you think that none is correct, answer choice **J**

__14. The car moves toward the right (away from the origin) with a steady (constant) velocity.

__15. The car is at rest.

__16. The car moves toward the right and is speeding up at a steady rate (constant acceleration).

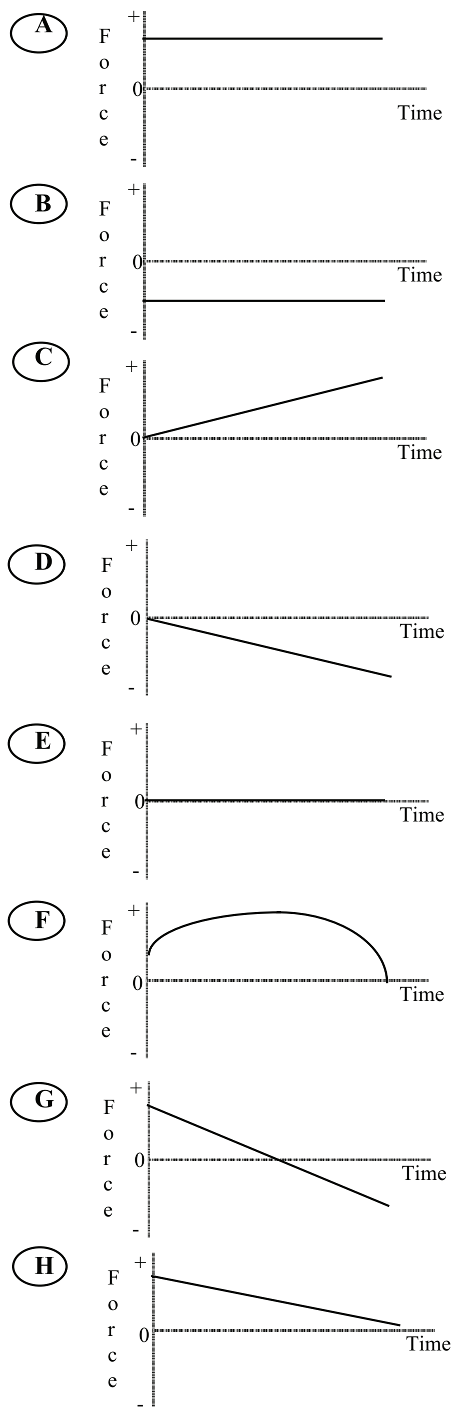
__17. The car moves toward the left (toward the origin) with a steady (constant) velocity.

__18. The car moves toward the right and is slowing down at a steady rate (constant acceleration).

__19. The car moves toward the left and is speeding up at a steady rate (constant acceleration).

__20. The car moves toward the right, speeds up and then slows down.

__21. The car was pushed toward the right and then released. Which graph describes the force after the car is released.



J None of these graphs is correct.

"Force Graph" Questions from the Force and Motion Conceptual Evaluation (Thornton and Sokoloff, 1998).

In post-instruction interviews with one of the students in this course (the same student quoted earlier in this paper), it became evident that the student did not necessarily make a connection between the methods she had learned to analyze dynamical questions by using graphical representations, and the intuitive methods she was accustomed to using in order to make decisions about what happens in everyday life. In the interview segment below, the student is asked to explain the answers she had written down when responding to the Force Sled questions shown above.

DEM: I need you to explain #3 [Force Sled Question #3]. ***["The sled is moving to the right. Which force would slow it down at a steady rate (constant acceleration)?"]***

STUDENT: [reads answer she chose] ***"The force is toward the left and is decreasing in strength."*** *. . . I was picturing the sled, and I was thinking that it would take less force once it started slowing down . . . I don't know . . .*

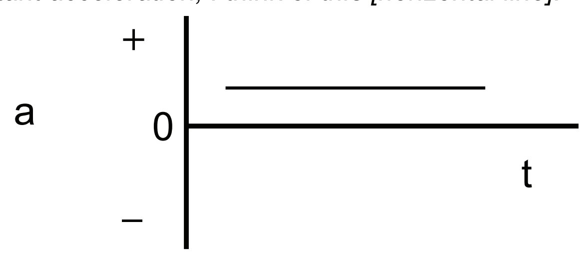
You want it to slow down at a steady rate. So since it's moving towards me and I want it to slow down, I'm actually going to have to go with it . . . and I guess I would ***increase*** *my force to slow it down, not decrease it. I don't know . . .*

DEM: Does the fact that it says "constant acceleration," does that help you to figure this out?

STUDENT: Only in so far as if the acceleration is constant, then the slope is zero . . .

DEM: The slope of what?

STUDENT: The slope of the acceleration, and so the slope of the force is going to be zero: they mirror each other. The force is going to be constant. [Draws graph to explain her reasoning.] *When I think of constant acceleration, I think of this [horizontal line].*



DEM: Now, on this one we've gone all the way around. At first you said less force was needed once it started slowing down, then you said maybe you have to increase the force. And now you're saying, "constant force."

STUDENT: Well, according to what I know, or what I think I know about graphs, I would say that the force had to remain constant because the acceleration is constant.

According to the visual image I have in my head, if a skater was coming towards me and I wanted to slow her down at a steady rate, I don't think that my force would be constant. I don't know ***why*** *I don't think that, I just think it would take less force towards the end.*

The student has apparently learned a particular algorithmic procedure for interpreting the meaning of "constant" acceleration and for relating those words to the correct response to a force question when expressed in graphical form. (That is, she says: "*When I think of constant acceleration, I think of this [horizontal line]*," and she also knows that the slope of the force and the slope of the acceleration must be the same because "*they mirror each other*.") However, it seems evident that she has not been able to make a connection between the understanding of the graphical representation of this physical situation, and her intuitive understanding of the way things actually work in the real world. Because of that lack of full understanding of the concept of Newton's second law, when a question about an object undergoing constant acceleration was posed to her in natural language form (that is, the Force Sled questions), she responded with an incorrect answer, rather than make use of the correct analysis she offered when analyzing the situation from a graphical perspective.

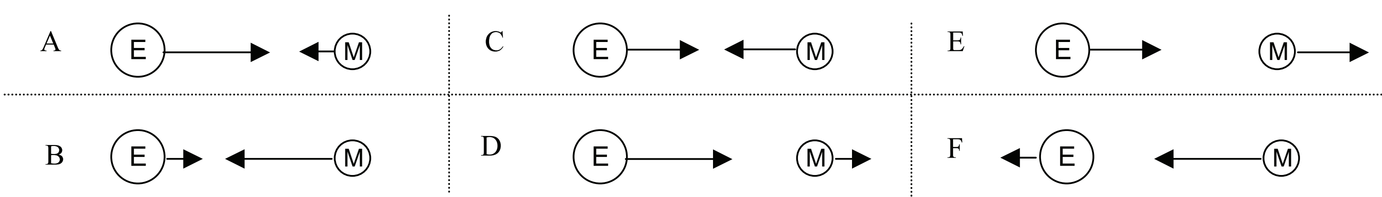
B. "Matched Sets": Similar test items posed in different representational modes

In other work, we have posed similar "matched sets" of questions to students in which other physics concepts were targeted. For example, in a question related to Newton's third law and his law of Universal Gravitation, a quiz containing the following two questions has been given over the past seven years, pre-instruction, to students taking the second semester of an algebra-based introductory physics course. (These students had all spent one full semester or more studying Newtonian mechanics, including the law of gravitation.)

#1. The mass of the sun is about 3 x 10^5 times the mass of the earth. How does the magnitude of the gravitational force exerted by the sun <u>on the earth</u> compare with the magnitude of the gravitational force exerted by the earth <u>on the sun</u>? The force exerted by the sun <u>on the earth</u> is:

A. about 9 x 10^{10} times larger
B. about 3 x 10^5 times larger
C. exactly the same
D. about 3 x 10^5 times smaller
E. about 9 x 10^{10} times smaller

#8. Which of these diagrams most closely represents the gravitational forces that the earth and moon exert ***on each other***? (Note: The mass of the earth is about 80 times larger than that of the moon.)



According to Newton's third law, the answer to both questions is that the mutual forces exerted by the interacting objects (sun and earth in Question #1, earth and moon in Question #8) are *equal* in magnitude. Therefore, the answer to both questions is "C."

These questions were both very difficult for the students, even though they all had studied the relevant concepts in their previous physics courses. In the table below, results are shown for four separate offerings of this course.

Question #1 response	**1998**	**1999**	**2001**	**2002**
A	13%	10%	8%	16%
B	68%	73%	62%	67%
C	14%	10%	23%	13%
D	5%	6%	6%	3%
E	0%	0%	0%	0%
n	*79*	*96*	*77*	*75*

Question #8 response	**1998**	**1999**	**2001**	**2002**
A	38%	47%	34%	47%
B	53%	45%	55%	43%
C	6%	6%	12%	7%
D	0%	0%	0%	3%
E	1%	0%	0%	0%
F	0%	2%	0%	1%
n	*79*	*96*	*77*	*75*

Although the rate of correct responses is consistently low, the ratio of correct responses on Question #8 to those on Question #1 is remarkably consistent from year to year:

	1998	1999	2001	2002
$\frac{correct\ responses\ on\ \#8}{correct\ responses\ on\ \#1}$	0.43	0.60	0.50	0.50

This certainly suggests that students had more difficulty, for whatever reason, with the question posed in diagrammatic form in comparison to the one posed in verbal form.

As a consistency check, Question #8 along with a question very similar to Question #1 have also been administered post-instruction in the same course. Although correct response rates were dramatically higher on both items, and most of the discrepancy was thereby erased, a small difference persisted. In 2002, the correct post-instruction response rate on Question #1 was 93%, while that on Question #8 was 86%.

Similar matched sets of test items have been administered for other physics concepts. Here I present data for one such set: Quiz #11 (see page 16), which relates to Coulomb's law of electrical force. (Correct Answers: #1, A; #2: A; #3, E; #4, E.)

It is extremely difficult to prepare such matched question sets so that each question on a set is *fully* equivalent to the others; some differences always exist with respect to some details of the information presented. (For example, a vector diagram inevitably makes available the directions of interaction forces; however, including such information in verbal or mathematical form, while possible, is much more cumbersome and would tend to unnecessarily obscure the main idea of the question.) Nonetheless, the four items on the question set shown here are substantially equivalent, and the four representations utilized (verbal, diagrammatic, mathematical/symbolic, and graphical) had all been extensively practiced by the students on quizzes, exams, and homework questions.

An extra-credit option on each test item allows students to increase their score if their response on that particular item is correct. Writing a "3" on the indicated line would increase the item score from 2.5 points for a correct response to 3.0 points. However, selecting this option and providing an *incorrect* response would result in a score of –1.0 for that item, rather than the 0.0 score that would otherwise be earned.

Results for several differing course offerings are shown in the table below. (Numbers shown are fractions of overall responses in each category; "low-conf correct" means "fraction of students who provided correct answer but indicated lower confidence by failure to select extra-credit option.")

The results show that correct response rates for items #1, #2, and #3 were nearly the same, while that for #4 – the graphical representation – was somewhat lower, perhaps due to the relatively unfamiliarity of that representation in the context of this particular question. It is also striking that the proportion of low-confidence correct responses was lower on the question posed in verbal representation than on the other three items, in each of the four years for which results have been analyzed. The overall rate of "low-confidence correct" responses was 15% on the verbal representation, compared to 22% on the other three items. This is certainly not a large discrepancy – it is only marginally statistically significant, if at all – but the fact that it was observed consistently is nonetheless remarkable and worthy of further study.

QUIZ #11	**N**	**#1 incorrect**	***#1 low-conf correct***	**#2 incorrect**	***#2 low-conf correct***	**#3 incorrect**	***#3 low-conf correct***	**#4 incorrect**	***#4 low-conf correct***
1998	71	0.04	*0.06*	0.07	*0.15*	0.10	*0.16*	0.14	*0.16*
1999	91	0.11	*0.12*	0.15	*0.17*	0.18	*0.15*	0.21	*0.15*
2000	79	0.14	*0.22*	0.11	*0.26*	0.10	*0.25*	0.11	*0.30*
2001	75	0.12	*0.21*	0.15	*0.31*	0.13	*0.35*	0.23	*0.28*
MEAN		0.10	*0.15*	0.12	*0.22*	0.13	*0.23*	0.17	*0.22*

V. Conclusion

There is little doubt that the form of representation of physics concepts may have an influence on the ways in which students learn and understand those concepts. Certain representations may pose particular learning difficulties – or, on the other hand, might be particularly fruitful – in the context of particular subject areas. It may also be the case that certain students are relatively more or less successful with particular forms of representations, and it might turn out that the relative utility of different representations varies significantly from one concept to another. The preliminary results presented in this paper suggest that these questions merit substantial addition scrutiny, and our group is continuing to investigate these issues.

Acknowledgments

This project was supported in part by National Science Foundation grants DUE-#9354595, DUE-#9650754, DUE-#9653079, DUE-#9981140, and REC-#0206683. The project on multiple representations is a collaborative effort with the Iowa State University Chemistry Education Research group, led by Prof. Thomas J. Greenbowe. T. J. Greenbowe is Co-Principal Investigator on NSF Grants DUE-#9981140 and REC-#0206683. K. Manivannan was Co-Principal Investigator on NSF Grants DUE-#9650754 and DUE-#9653079, and collaborated in the development and teaching of the inquiry-based elementary physics course referred to in this paper.

I would like to thank Leith Allen for provocative and useful conversations, and Larry Engelhardt for assistance with data analysis.

Physics 112
Quiz #11
October 6, 2000

Name: ______________________

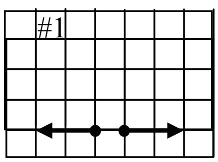
IF YOU WANT A QUESTION GRADED OUT OF THREE POINTS (–1 [MINUS ONE] FOR WRONG ANSWER!!) WRITE "3" IN SPACE PROVIDED ON EACH QUESTION.

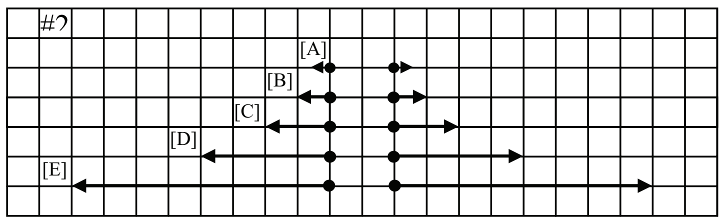
1. When two identical, isolated charges are separated by two centimeters, the magnitude of the force exerted by each charge on the other is eight newtons. If the charges are moved to a separation of eight centimeters, what will be the magnitude of that force now?

A. one-half of a newton
B. two newtons
C. eight newtons
D. thirty-two newtons
E. one hundred twenty-eight newtons

Grade out of three? Write "3" here: ______

2. Figure #1 shows two identical, isolated charges separated by a certain distance. The arrows indicate the forces exerted by each charge on the other. The same charges are shown in Figure #2. Which diagram in Figure #2 would be correct?





A.
B.
C.
D.
E.

Grade out of three? Write "3" here: ______

3. Isolated charges q_1 and q_2 are separated by distance r, and each exerts force F on the other. $q_1^{initial} = q_1^{final}$ and $q_2^{initial} = q_2^{final}$; $r^{initial} = 10m$; $r^{final} = 2m$. $\mathrm{F}^{initial} = 25\mathrm{N}$; $\mathrm{F}^{final} = ?$

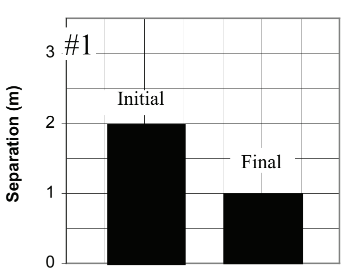
A. 1 N
B. 5 N
C. 25 N
D. 125 N
E. 625 N

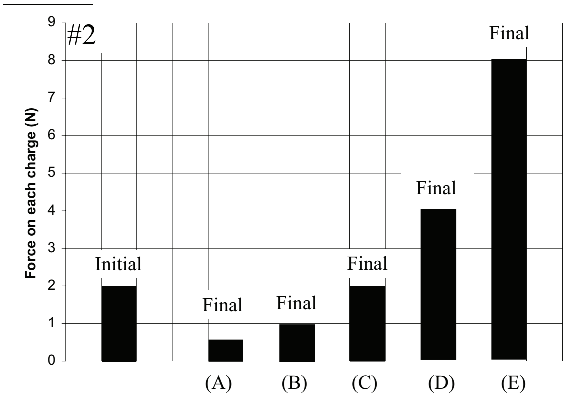
Grade out of three? Write "3" here: ______

4. Graph #1 refers to the initial and final separation between two identical, isolated charges. Graph #2 refers to the initial and final forces exerted by each charge on the other. Which bar is correct?

Grade out of three? Write "3" here: ______

A.
B.
C.
D.
E.





References

1. Berger, R. and H. Wiesner (1997). "Zum Verständnis grundlegender Begriffe und Phänomene der Thermodynamik bei Studierenden," in *Deutsche Physikalische Gesellschaft, Fachverband Didaktik der Physik: Didaktik der Physik* (Technische Universität Berlin, Institut für Fachdidaktik Physik und Lehrerbildung, Berlin, 1997), pp. 736-741.
2. Greenbowe, Thomas J. and David E. Meltzer (2002). "Student learning of thermochemical concepts in the context of solution calorimetry," *International Journal of Science Education* (in press).
3. Kesidou, Sofia, Reinders Duit, and Shawn M. Glynn (1995). "Conceptual development in physics: Students' understanding of heat," in *Learning Science in the Schools: Research Reforming Practice*, edited by Shawn M. Glynn and Reinders Duit (Lawrence Erlbaum, Mahwah, NJ, 1995), pp. 179-198, and references therein.
4. Loverude, Michael E., Christian H. Kautz, and Paula R. L. Heron (2002). "Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas," *American Journal of Physics* **70**, 137-148.
5. McDermott, Lillian C., and Peter S. Shaffer. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding. *American Journal of Physics* **60,** 994; printer's erratum **61**, 81 (1993).
6. Meltzer, David E., Kandiah Manivannan, and Tina Tassara (1997). "Effectiveness of a Microcomputer-based Laboratory Environment for Teaching Force and Motion in an Elementary Physics Course," presentation at the 1997 Summer Meeting of the American Association of Physics Teachers, Denver, Colorado, August 15, 1997 [*AAPT Announcer* **27 (2)**, 122 (1997)].
7. Meltzer, David E. (2001). "Student reasoning regarding work, heat, and the first law of thermodynamics in an introductory physics course," *Proceedings of the Physics Education Research Conference, Rochester, New York, July 25-26, 2001*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (Rochester, New York, 2001), pp. 107-110.
8. Meltzer, David E. (2002). "Investigation of students' reasoning regarding heat, work, and the first law of thermodynamics in an introductory calculus-based general physics course," submitted to *American Journal of Physics.*
9. Shaffer, Peter S. and Lillian C. McDermott (1992). "Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies," *American Journal of Physics*, **60,** 1003.
10. Tiberghien, A. and G. Delacôte (1978). "Résultats préliminaires sur la conception de la chaleur," in *Physics Teaching in Schools: Proceedings of the 5th Seminar of GIREP*, edited by G. Delacôte (Taylor & Francis Ltd, London, 1978), pp. 275-282.

11. Thornton, Ronald K. and David R. Sokoloff (1998). “Assessing student learning of Newton’s laws: the Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula,” *American Journal of Physics* **66,** 338.

12. Touger, Jerold S. (1986). “A case study in the linguistic structure of physics discourse.” Proc. Int. Conf. "Communicating Physics," ICPE and Univ. Duisburg, Duisburg, Germany, 1986; pp. 515-540.

13. Touger, Jerold S. (1991). “When words fail us,” *The Physics Teacher* **29***,* 90-95.

14. Williams, H. Thomas (1999). “Semantics in teaching introductory physics,” *American Journal of Physics* **67,** 670-680.

15. Zemansky, Mark W. (1970). “The use and misuse of the word ‘heat’ in physics teaching,” *Physics Teacher* **8**, 295-300.

Visualization Tool for 3-D Relationships and the Right-Hand Rule

Ngoc-Loan Nguyen and David E. Meltzer, Iowa State University, Ames, IA

The need to develop an understanding of spatial relationships in three dimensions is one of the major challenges faced by introductory physics students. It arises, for example, when grappling with three-dimensional coordinate systems and with the vector ("cross") product, when dealing with the concepts of torque and angular momentum, and perhaps most prominently when studying relationships involving magnetic fields and forces. A variety of so-called "right-hand rules" are important and widely used tools for working with such concepts. In this paper we describe a simple and inexpensive visualization tool that may be used to help learn and work with these important rules.

Greenslade[1] has described the evolution of the modern right-hand rule from a number of mnemonic devices that originated shortly after Oersted's discovery in 1820 of the force exerted on a compass needle by a current-carrying wire. Various physical models made of cardboard, wires, and other materials were constructed, and an assortment of visualization "rules" were developed and popularized in early textbooks. The right-hand rule in its more modern form began to appear in textbooks quite commonly beginning around 1900.

Although the right-hand rule is an important mnemonic technique, physics instructors are well aware of the difficulties accompanying its use. It is not unusual to watch students attempting to apply the right-hand rule become so fixated in their hand manipulation that they actually switch or forget which finger (or hand orientation) they initially had associated with a par-

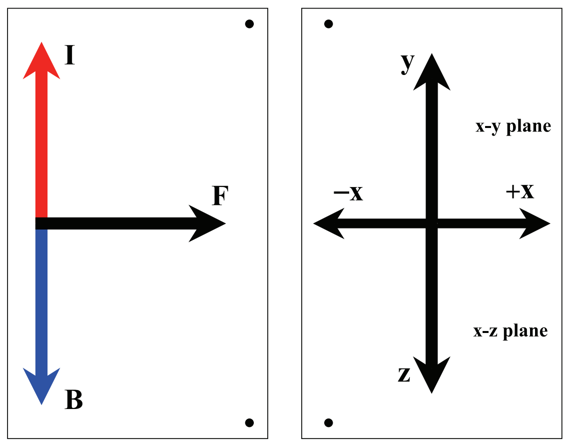


Fig. 1. Copy masters for the Current-Magnetic Field-Force ["*I-B-F*"] card and the Cartesian-axes card. These may be enlarged and copied directly onto card stock. Although different colors are used here to distinguish *I*, *B*, and *F*, monochrome cards are completely satisfactory.

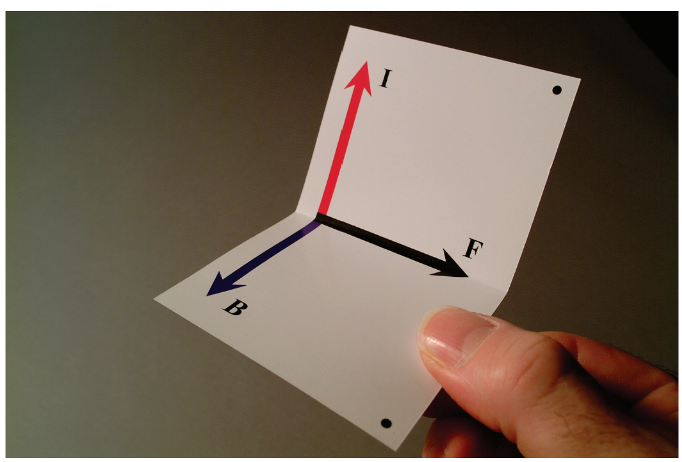


Fig. 2. The angle of the fold in the *I-B-F* card can vary between 0º and 180º.

 DOI: 10.1119/1.1869425

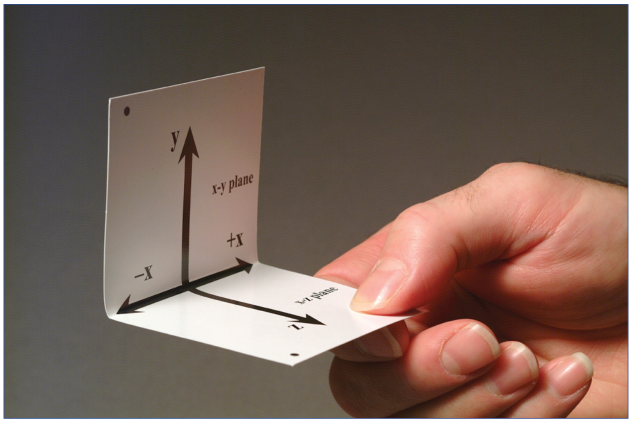


Fig. 3. Use of the Cartesian-axes card.

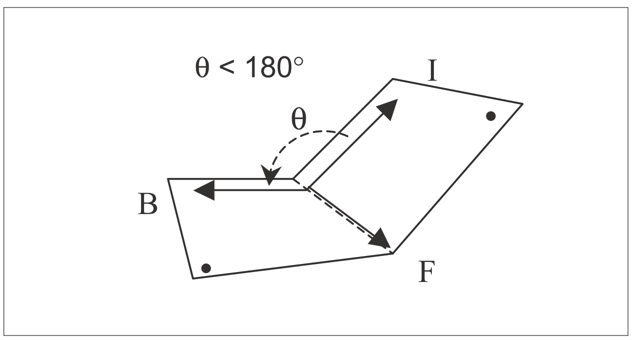


Fig. 4. The angle between the I and B vectors must be less than 180º.

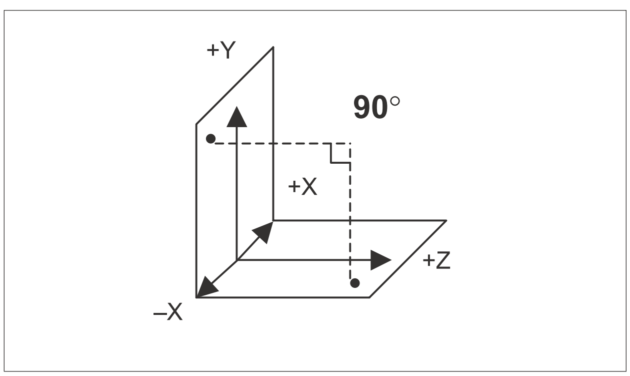


Fig. 5. The fold in the Cartesian-axes card should form a 90º angle.

ticular axis or physical quantity.[2] Another pitfall is that students attempting to apply the right-hand rule may choose an initial orientation that makes it difficult to move their body so as to obtain proper alignment of their hand and fingers with the corresponding physical entity.

Pedagogical tools to assist students in learning vector relationships have been described in the literature. Rather bulky physical models that represent three-dimensional vector relations have been described by Francis,[3] and Wunderlich et al.[4] have discussed a similar model to demonstrate Cartesian- and polar-coordinate systems. Some early devices developed to help with magnetic-field problems are discussed in the article by Greenslade.[1] Van Domelen[5] has recently described a device specifically designed to assist students with the right-hand rule. This consists of a transparent rectangular box constructed of plastic, inside of which are set three colored arrows oriented in three fixed, perpendicular directions.

For some years we have been using a simple device that is very quick and easy to construct and is very inexpensive, and yet has proved quite helpful to students in working with the right-hand rule. It consists of an ordinary 3-x-5-in index card, folded in half along its short axis, on which arrows are drawn to represent three perpendicular directions in space. These arrows may represent, for instance, *x*, *y*, and *z* coordinate axes, directions of current (I), magnetic field (**B**) and magnetic force (**F**), etc. One might also design a card on which the three arrows are simply labeled **A**, **B**, and **C** to illustrate the vector product **A** x **B** = **C**. In the simplest version of this device, one simply folds a blank card in half, draws one arrow along the fold, then draws a double-ended arrow along the long axis of the card intersecting the first arrow, labeling each arrowhead appropriately. A rather professional-looking version can be generated by copying and enlarging the accompanying figures (Fig. 1) and printing them out on card stock. The photographs (Figs. 2 and 3) illustrate the use of these cards.

Two black dots are placed on the cards to assist students in orienting the axes. For the Current-Magnetic Field-Force ["*I-B-F*"] card, the angle formed by lines connecting the dots to the centerfold should be smaller than 180º (see Fig. 4). For the Cartesian-axes card, the fold in the card should form a 90º angle at all times (Fig. 5). The cards can easily be flattened again to be stored inside a pocket of the student's notebook.

The *I-B-F* card is used to find the direction of the magnetic force on a current-carrying conductor in a magnetic field.[6] First the student can orient the current arrow I in the direction that the current flows. Then the entire card should be rotated and the fold-angle adjusted, with the current arrow staying fixed in orientation, until the magnetic-field arrow **B** is pointing along the direction of the external magnetic field. The force arrow **F** then shows the direction of the magnetic force, so long as the card doesn't become "bent backwards" and the dots are connected by an arc smaller than 180°. (If the angle exceeds 180°, the actual direction of the force will of course be opposite to the direction of the **F** arrow.)

The Cartesian-axes card can be set down ahead of time or rotated as needed in order to help the student remember the relative spatial orientation of x, y, and z axes, as well as distinguishing between the $+x$ and $-x$ directions. It can also help clarify the meaning of "x-y plane," "x-z plane," etc.

In the classroom environment, use of these cards has proved popular among most students. (We also allow their use on quizzes and exams.) We have found that the cards are most helpful when used in conjunction with other standard right-hand rule techniques. Typically, students are first asked to solve the problem utilizing the cards, and then asked to try and replicate the result with one of the standard right-hand rule mnemonics using fingers and hands. It is also useful to ask students to relate the reversal of the force direction that can occur when using the *I-B-F* card to the negative sign resulting from an angle greater than 180° between the current and magnetic-field vectors when using the equation $F = ILB \sin \theta$.

It is easy to come up with other possible uses of a folded index card to illustrate three-dimensional spatial relationships. Indeed, one might assign students the exercise of devising their own methods for illustrating such relationships with the use of the cards. We are exploring the possibility that other simple low-tech devices—perhaps somewhat more elaborate than a folded index card!—can assist students in learning physics principles in which three-dimensional vector concepts and spatial reasoning are involved.

Acknowledgment

This material is based in part upon work supported by the National Science Foundation under Grant #REC-0206683.

References

1. Thomas B. Greenslade Jr., "Ancestors of the right-hand rule," *Phys. Teach.* **18**, 669–670 (Dec. 1980).
2. "Photo of the month," *Phys. Teach*. **40**, 115 (Feb. 2002).
3. Brother Columban Francis, "Vector visual aids," *Phys. Teach*. **5**, 119–122 (March 1967).
4. Francis J. Wunderlich, Michael J. Hones, and Donald E. Shaw, "Three-dimensional vector demonstrator," *Phys. Teach.* **14**, 232–233 (April 1976).
5. David J. Van Domelen, "Artificial right-hand rule device," *Phys. Teach.* **37**, 500–501 (Nov. 1999).
6. It is interesting to compare this card to a device developed by Roget to achieve a similar objective. It was described by Noad in an early textbook and is illustrated in the article by Greenslade (Ref. 1, Fig. 3): Henry M. Noad, *A Manual of Electricity* (Lockwood and Co., London, 1859), p. 643.

PACS codes: 01.50Fa, 01.50Fb, 01.50M

Ngoc-Loan Nguyen *recently received an M.S. in physics and is working as a research assistant at Iowa State University.*
Department of Physics and Astronomy, Iowa State University, Ames, IA 50011; nguyenn@iastate.edu

David E. Meltzer *is an assistant professor of physics at Iowa State University.*
Department of Physics and Astronomy, Iowa State University, Ames, IA 50011; dem@iastate.edu

Relation between students' problem-solving performance and representational format

David E. Meltzer[a)]
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

(Received 15 August 2003; accepted 7 January 2005)

An analysis is presented of data on students' problem-solving performance on similar problems posed in diverse representations. Five years of classroom data on 400 students collected in a second-semester algebra-based general physics course are presented. Two very similar Newton's third-law questions, one posed in a verbal representation and one in a diagrammatic representation using vector diagrams, were given to students at the beginning of the course. The proportion of correct responses on the verbal question was consistently higher than on the diagrammatic question, and the pattern of incorrect responses on the two questions also differed consistently. Two additional four-question quizzes were given to students during the semester; each quiz had four very similar questions posed in the four representations: verbal, diagrammatic, mathematical/symbolic, and graphical. In general, the error rates for the four representations were very similar, but there was substantial evidence that females had a slightly higher error rate on the graphical questions relative to the other representations, whereas the evidence for male students was more ambiguous. There also was evidence that females had higher error rates on circuit-diagram problems in comparison with males, although both males and females had received identical instruction.

[DOI: 10.1119/1.1862636]

I. INTRODUCTION

This paper reports on the initial phase of an investigation into the role of diverse representations in the learning of physics concepts. The goal is to explore the relation between the form of representation of complex concepts, and students' ability to learn these concepts. Much previous research has shown that the use of multiple forms of representation in teaching concepts in physics has great potential benefit, and yet poses significant challenges to students and instructors.[1,2] Facility in the use of more than one representation deepens a student's understanding, but specific learning difficulties arise in the use of diverse representations.[3]

By representation I mean any of the widely diverse forms in which physical concepts may be understood and communicated. In Appendix A I show an example of the use of four representations for what is essentially the same problem. The representations are referred to here as verbal (V), diagrammatic (D), mathematical/symbolic (M), and graphical (G), corresponding to questions 1–4, respectively.[4] Although these questions are nearly identical and illustrate four different ways of representing the same concept, to an introductory student they might appear very different. It often is assumed by instructors that a representation which they find especially clear and comprehensible (for example, a graph) also will be especially clear for the average student. Research and experience shows that this assumption often is not correct,[3] but relatively little work has been devoted to testing it systematically. In this paper I will discuss a variety of methods of investigating how specific representations may be related to student thinking, and I will analyze classroom data to generate some preliminary hypotheses regarding this relation.

II. THE ROLE OF MULTIPLE REPRESENTATIONS IN STUDENT LEARNING OF PHYSICS

A. Outline of previous research

There is no purely abstract understanding of a physical concept—it is always expressed in some form of representation. Physical scientists employ a variety of representations as a means for understanding and working with physical systems and processes.[5–9] In many recently developed curricular materials in physics[1,2,10–16] and chemistry,[17] there has been much attention to presenting concepts with a diversity of representations. Van Heuvelen was one of the earliest to emphasize the potential benefits of this instructional strategy in physics.[1] Numerous physics educators have stressed the importance of students developing an ability to translate among different forms of representation of concepts,[1,3,18–22] and researchers in other fields have stressed similar themes.[23–27] Moreover, it has been pointed out that thorough understanding of a particular concept may *require* an ability to recognize and manipulate that concept in a variety of representations.[2,3]

It is well established that specific learning difficulties may arise with instructional use of diverse representations.[3] Student difficulties in mastering physics concepts using graphical representations have been studied in considerable detail and specificity for topics in kinematics.[18,28–30] These studies and other related work in mathematics education[31] have delineated several broad categories of conceptual difficulties with graphs. Conceptual difficulties related to diagrammatic representations of electric circuits and fields have been addressed,[32] as have those in optics.[33] Difficulties arising from linguistic ambiguities (verbal representation) also have been explored.[34] Specific representational difficulties in chemistry education, largely parallel to similar issues in physics education, also have been investigated.[35,36]

B. Research issues related to multiple representations

Beyond the investigations in the literature cited, there are few available research results that focus on problems that arise in the learning of physics concepts with multiple forms of representation. As McDermott has emphasized, there is a need to identify the specific difficulties students have with various representations.[3] I suggest that additional insight might result from investigations that explicitly compare learning in more than one form of representation. Although a number of recent investigations in science education and other fields have focused on broader issues involved in student learning with diverse representations,[37,38] there seems to have been relatively little effort to compare representations in terms of their pedagogical effectiveness in particular contexts.[39]

A closely related issue is that of students' relative performance on similar problems that make use of different representational forms.[21,26,29,40,41] In this regard, Kozma[42] and Kozma and Russell[26] have reported on the relative degree of difficulty encountered by novice students presented with a chemistry problem posed in various representations. Among physics and chemistry educators, there has been speculation regarding the role that students' individual learning styles might play,[43] and the possible relevance of gender differences[35,40] and spatial ability.[44]

The present investigation focuses on specific issues arising when multiple representations are utilized in undergraduate physics instruction. Ultimately, the issues we plan to investigate include the following:

(1) What subject-specific learning difficulties can be identified with various forms of representation of particular concepts in the introductory physics curriculum?
(2) What generalizations might be possible regarding the relative degree of difficulty of various representations in learning particular concepts? That is, given an average class engaging in a typical sequence of instructional activities, do some forms of commonly used representations engender a disproportionately large number of learning difficulties?
(3) Do individual students perform consistently well or poorly with particular forms of representation with widely varying types of subject matter?
(4) Are there any consistent correlations between students' relative performance on questions posed in different representations and parameters such as major, gender, age, and learning style?

Preliminary results regarding these issues will be presented in this paper. The analysis and discussion are based on five years of classroom data, generated during the initial stages of an investigation into these issues. Ultimately, our goal is to investigate the relative effectiveness of various representations in learning; however, the initial data discussed in this paper will focus on student performance. Although these objectives are presumably closely related, it must be kept in mind that they are not identical, and that the connection between the two in the context of multiple representations must be explicitly investigated.

III. COMPARISON OF STUDENT PERFORMANCE: VERBAL VERSUS DIAGRAMMATIC VERSION OF NEWTON'S THIRD-LAW QUESTION

A. Description of questions

Two very similar questions related to Newton's third law were used to probe possible differences in students' interpretation of and performance on questions posed in different representational formats. The two questions are shown in Fig. 1(a); they were part of an 11-item quiz on gravitation, and they are numbered here according to their position on the original quiz. Question 1 is posed in a verbal (V) representation. Question 8 is posed in a diagrammatic (D) representation, making use of vector diagrams.

The quiz containing these questions was administered on the second day of class in a second-semester, algebra-based general physics course at Iowa State University. This quiz was administered in courses offered during five consecutive years, 1998–2002, during the fall semester. All students had completed the equivalent of a one-semester course focusing on mechanics, and had previous instruction related to Newton's laws with vector representations. Most took a traditional first-semester course.

The quiz did not count for a grade; students were told that it was given to help assess their level of preparation on topics that would be needed in subsequent class discussions. I will refer to this quiz as the gravitation pretest, because a second version of the same quiz was administered to the students after instruction had taken place.

B. Results

The responses to the gravitation pretest are shown in Table I.[45] Responses varied from year to year, with the percentage of correct responses ranging from 10% to 23% on question 1 (overall average: 16% correct, $N=408$) and 6% to 12% on question 8 (overall average: 9% correct). This low proportion of correct responses to a Newton's third-law question is consistent with previous research on traditional courses regarding students' belief that unequal masses in an interacting pair exert forces of unequal magnitude. It is related to a general view referred to as the "dominance principle."[46] There are two interesting and consistent discrepancies between the responses to the two questions: the significantly lower correct-response rate on the diagrammatic question ($p=0.03$ according to a two-sample t-test), and the far greater popularity on this question of a response that could be interpreted as a "larger mass exerts a smaller force" conception (response A on question 8, responses D and E on question 1).

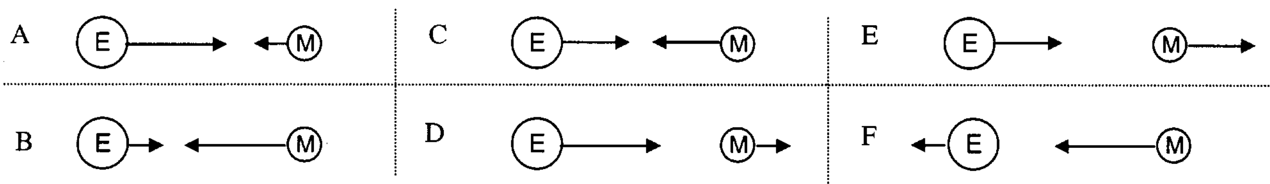
The first row of Table II shows the ratio of the number of correct responses on question 8 to that on question 1. It is particularly striking that although the proportion of correct responses (response C on both questions) varied substantially from year to year, the ratio of correct responses on one question relative to the other in a particular year is nearly constant. The range is 0.45–0.60 (the overall average is 0.53), a 33% variation that contrasts with the more than 200% year-to-year variation in the correct-response rate itself. These

(a)

#1. The mass of the sun is about 3×10^5 times the mass of the earth. How does the magnitude of the gravitational force exerted by the sun on the earth compare with the magnitude of the gravitational force exerted by the earth on the sun? The force exerted by the sun on the earth is:

A. about 9×10^{10} times larger
B. about 3×10^5 times larger
C. exactly the same
D. about 3×10^5 times smaller
E. about 9×10^{10} times smaller

#8. Which of these diagrams most closely represents the gravitational forces that the earth and moon exert ***on each other***? (Note: The mass of the earth is about 80 times larger than that of the moon.)



(b)

#1. A 5-kg lead sphere is hanging 12 m from a 500-kg lead sphere. How does the gravitational force exerted by the 5-kg sphere on the 500-kg sphere compare with the magnitude of the gravitational force exerted by the 500-kg sphere on the 5-kg sphere? The forced exerted by the 5-kg sphere on the 500-kg sphere is:

A. 100 times larger
B. 10 times larger
C. exactly the same
D. 10 times smaller
E. 100 times smaller

Fig. 1. Questions on the gravitation quiz: (a) gravitation pretest questions 1 (verbal representation) and 8 (diagrammatic representation); (b) gravitation posttest question 1. The posttest version of question 8 was unchanged from the pretest.

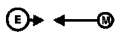
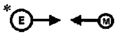
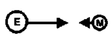
questions also were given once (in spring 2000) in the second-semester calculus-based general physics course. Although the correct-response rate was far higher on both questions in this course (62% on V, 38% on D), the ratio of the correct responses on D compared to V was consistent with the results from the algebra-based course (see the final column of Table II).

The proportion of students giving the response corresponding to "larger mass exerts a smaller force" (response A) on the D question also is consistently far higher than on the V question, as shown by the second row in Table II. Overall, this response accounted for only 5% of all responses to the V question, but 41% of those to the D question. On the gravitation pretest, those who correctly answered C on the V question were divided on their responses to the D question: 41% answered it correctly (response C), but nearly all others gave either response A (larger mass exerts a smaller force) or B (larger mass exerts a larger force), in almost equal numbers. This equally divided response pattern paralleled the behavior of the majority who had answered the V question incorrectly. Of all incorrect responses on the D question, 45% were A and 53% were B.

A posttest version of the gravitation quiz was administered approximately one week after the pretest. The posttest version of question 1 is shown in Fig. 1(b); question 8 was unchanged from the pretest. The posttest was a graded quiz. The instruction that occurred between the pre- and posttests was based on interactive-engagement methods[16] and was used to lead in to a discussion of electrical forces and fields.

The overall error rate on the posttest (N=400) dropped to 6% on V (range: 5%–8%), but only to 20% on D (range: 14%–25%). Even after substantial improvement in the overall correct-response rate, the significantly higher error rate on the D question persisted. Again, the errors on the D version of the question were split between the "larger mass exerts a smaller force" response A (25% of incorrect responses) and the more popular "larger mass exerts a larger force" response B (75% of incorrect responses). This preference for B

Table I. Responses to questions 1 and 8 on the gravitation pretest. For question 1, ‘‘larger’’ refers to responses A and B, ‘‘the same’’ refers to response C, and ‘‘smaller’’ refers to responses D and E. An asterisk (*) denotes the correct answer. The rate of correct responses fluctuates significantly from year to year, but the ratio of correct responses (on question 1 versus question 8) is nearly constant.

	1998	**1999**	**2000**	**2001**	**2002**
N	*78*	*96*	*83*	*77*	*74*
1. force by the sun is					
larger [A or B]	81%	83%	76%	70%	84%
*the same [C]	14%	10%	20%	23%	14%
smaller [D or E]	5%	6%	4%	6%	3%
8. earth/moon force					
 [B]	54%	45%	45%	55%	43%
*  [C]	6%	6%	12%	12%	7%
 [A]	38%	47%	41%	34%	46%
other	1%	2%	2%	0%	4%

contrasted with the much more even split observed on the pretest.[47] A large majority (81%) of the incorrect responses on the *V* posttest question were for response E, corresponding to the smaller mass exerting the smaller force. Therefore, among students who responded incorrectly, the preference for a response consistent with the dominance principle (larger mass exerts a larger force) was unchanged from the pretest.

In 2002, a pair of questions nearly identical to questions 1 and 8 in Fig. 1 was placed on the final exam of the course (see Fig. 2). These questions[48] changed the context to electrostatics, one of the major topics covered in the course. On the *D* question, students were required to explain their answer. The error rate on these questions was 9% on *V* and 14% on *D* ($N=70$). Again the errors on *D* were split almost evenly between responses A and B. Most of the written explanations for these incorrect responses were clearly consistent with a belief that the larger-magnitude charge exerts the greater-magnitude force, including 80% of the explanations given by those who had chosen response A for this question, that is, the diagram consistent with the smaller force being exerted by the larger charge. An example of an explanation given to justify choice A is that ‘‘*Opposite charges attract. Since q_1 is the greater charge it will exert a greater force*.’’

This explanation is consistent with the hypothesis that the large proportion of responses observed for the A option (smaller mass exerts a larger force) on question 8 of the gravitation quiz was due to students’ confusion about whether the arrow in such diagrams represents the force exerted *on* or the force exerted *by* the object.

There also were several students who gave a correct response on the *V* question, but an incorrect response on the *D* question, and whose explanations were consistent with the dominance principle. This pattern is consistent with the observation that almost 60% of those who gave the correct response to the *V* question on the gravitation pretest from 1998 to 2002 did not correctly answer the *D* question, but instead gave an A or B response consistent either with the dominance principle or its opposite.

In 2002, 64% of the students who made errors on either the gravitation posttest or the final exam questions made representation-related errors on one or the other, but not on both tests. A representation-related error refers either to a correct answer on only one of the two (*D* and *V*) questions in the pair, or incorrect but inconsistent answers on both questions, such as B on 1 and A on 8. This observation is consistent with results regarding the consistency of students’ responses, as will be discussed further in Sec. IV.

Table II. Comparison of responses on gravitation pretest: diagrammatic (*D*, question 8) versus verbal (*V*, question 1). First row: ratio of number of correct (C) responses on *D* to number of correct (C) responses on *V*; fluctuations are in a relatively narrow range. Second row: ratio of number of ‘‘smaller than’’ (A) responses on *D* to number of ‘‘smaller than’’ (D and E) responses on *V*; ratios are much greater than one, implying a consistent response discrepancy. Data for algebra-based second-semester general physics course (1998–2002) are shown. The final column shows data for a calculus-based second-semester general physics course (spring 2000), which are in good agreement with those for the algebra-based course.

Ratio of	1998	1999	2000	2001	2002	Calculus-based course (2000) $N=240$
correct on *D*/ correct on *V*	0.45	0.60	0.59	0.50	0.50	0.61
‘‘smaller’’ on *D*/ ‘‘smaller’’ on *V*	8	8	11	5	18	26

IV. MULTI-REPRESENTATIONAL QUIZZES: COMPARISON OF RESPONSES ON DIVERSE REPRESENTATIONS

A. Background

Two additional quizzes were designed to incorporate questions posed in the four representations described in the Introduction. (Note that in this context, ‘‘graphical’’ refers to bar charts and not to line graphs.)

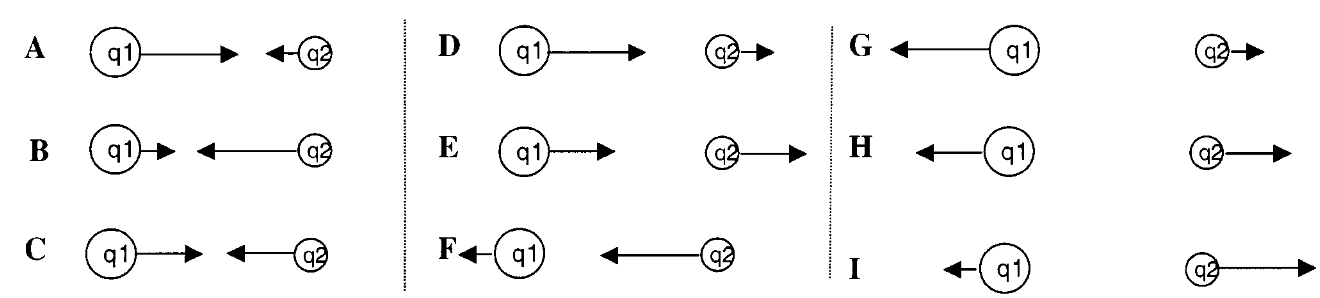
The first quiz (Appendix A, Coulomb quiz) required students to find the magnitude of the electrostatic force between two interacting charges, given the initial force and the initial and final separation distances. This quiz was administered midsemester and counted toward students’ grades. The second quiz (Appendix B, circuits quiz) involved a comparison of two different two-resistor direct-current circuits, one series and one parallel. The two circuits utilize batteries of the same voltage, but the individual resistances are different. Students were required to determine whether the current through a specified resistor in the parallel circuit is greater than, equal to, or less than the current flowing through a specified resistor in the series circuit. This quiz also was administered midsemester, during 1998–2002.

The intention was to make the four questions on each quiz as nearly equal in difficulty to each other as possible. For example, the separation ratios in the Coulomb quiz (larger separation distance divided by smaller separation distance) are all small integers (2, 4, and 5), and all five answer options correspond to the same set of choices, that is, the force

1. Particle A has a charge that is ten times the magnitude of the charge on particle B. How does the magnitude of the electrical force exerted by charge A on the smaller charge B compare with the magnitude of the electrical force exerted by charge B on charge A? The force exerted by charge A on the smaller charge B is:

A. 100 times larger
B. 10 times larger
C. exactly the same
D. 10 times smaller
E. 100 times smaller

2. In the figure below, particle q1 has a charge of +10 C, and particle q2 has a charge of –2 C.



(A) *[3 points]* Which of these diagrams most closely represents the electrical forces that the two charges exert ***on each other?***

(B) *[2 points]* Explain your answer to part (A).

Fig. 2. Electrostatic version of Newton's third-law questions; administered as part of 2002 final exam.

increases or decreases by a factor equal to the separation ratio or the separation ratio squared, or no change. It is important to emphasize that by the time these quizzes were administered, the students had had extensive exposure to and practice with various questions and problems utilizing all four representations on many quizzes, exams, and homework assignments.

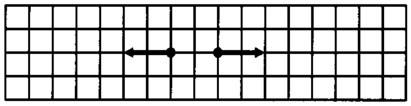
B. Common errors on Coulomb quiz and circuits quiz

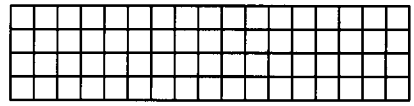
On the Coulomb quiz, the most common error by far was the assumption that the electrical force was proportional to $1/r$, instead of $1/r^2$. This error corresponded to the response sequence B, B, D, D on questions 1–4, respectively. The proportion of all incorrect responses represented by this error was 74%, 62%, 51%, and 50%, respectively. Very few of the incorrect responses corresponded to the "no change" answer with the exception of question 2. On this question (the D version), the "no change" response C represented 16% of all incorrect responses. Interview data and informal discussions with students indicated that they sometimes overlooked the fact that in this question, the separation between the charges has been changed in the diagram on the right.

In 2001 non-multiple-choice variants of the D and M questions on the Coulomb quiz were given as part of a follow-up quiz (see Fig. 3). On this quiz, students were required to explain their answers to the D question. The nearly identical error rates on these questions (28% and 25% on D and M, respectively, disregarding explanations; $N=75$) were approximately double those on the earlier multiple-choice quiz (15% and 13%, respectively). The "$1/r$" error continued to represent the majority of incorrect responses, which was consistent with students' written explanations and algebraic work. The proportion of incorrect responses represented by this error on the follow-up quiz (76% for D, 58% for M) was comparable to that observed on the initial quiz in 2001 (64% for D, and 80% for M).

It appeared that many students who had not made the $1/r$ error on the original quiz did make this error on the follow-up quiz on one or another of the two questions. There was no clear pattern which would suggest that their error was due specifically to the form of representation. The number of

3. The diagram on the left shows two isolated particles with equal magnitude charges, along with the electrical forces acting on those particles due to their mutual interactions. The ***same*** charges are to be repositioned in the diagram on the right, this time separated from each other by ***half*** the distance that separated them in the diagram on the left. Complete the figure in the right diagram to represent the new positions and forces.





Explain how you decided on the lengths of the arrows:

4. Isolated particles with charges q_1 and q_2 ($q_1 = q_2$) are separated by distance r, and initially experience mutual interaction forces with magnitudes $F_1^{initial} = F_2^{initial} = 40$ N; $r^{initial} = 10$ m. The particles are repositioned so $r^{final} = 2\, r^{initial}$.

F_1^{final} = ________

[no partial credit; answer must be within 15% of correct answer; −1 deduction for missing or incorrect units]

Fig. 3. Non-multiple-choice versions of diagrammatic and mathematical questions on the Coulomb quiz, administered as part of a follow-up quiz in 2001; numbered according to their position on the quiz.

students who switched from correct on D (on the initial quiz) to incorrect (on the follow-up quiz) was exactly the same as the number who switched from correct to incorrect on M, and the proportion who moved in the other direction—from incorrect to correct—was almost identical in the two representations. Of the students who made errors on the follow-up quiz, only 28% made consistent errors on both D and M questions (for example, making the $1/r$ error on both), while most (62%) made errors on only one of the two questions.

On the circuits quiz (Appendix B), the most common incorrect response corresponded to greater current flowing through the resistor in the series circuit (it has the smaller of the two resistances in three of the four questions), instead of the one in the parallel circuit. The proportion of all incorrect responses represented by this error was 88%, 89%, 79%, and 67%, respectively, on questions 1–4. The "equal currents" response (response B in all cases) represented 8%–15% of the incorrect responses on questions 1–3, but 30% on question 4. This difference might be due to the fact that in contrast to questions 1–3, the parallel and series resistors whose currents are being compared in question 4 are shown to be of equal resistance (instead of the parallel resistance being greater). This response pattern might imply the existence of a nonrepresentational artifact in the data.

The diagrams, algebraic work, and other notations written on students' papers were scrutinized carefully to ascertain why some students made an error on one or two questions, and yet did not do so on other questions on the same quiz. No pattern could be determined—the errors appear to occur almost randomly. This finding was consistent with observations made of students' work on all instruments employed in this study. In a further attempt to probe for any possible representation-related learning difficulties, students' responses to the quiz questions were subjected to considerable additional statistical analysis as will be described in the following.

C. Error rates

One question of interest is whether, on average, students find particular representations more difficult than others. The error rates for each question on the Coulomb and circuits quizzes are shown in Table III. There were no blank responses. "Any Error" refers to students who made errors on one or more of the questions on a given quiz, with the following exception: Students who gave four incorrect answers that were clearly consistent with each other were not counted in the "Any Error" statistic. Such a set of responses was, for instance, B, B, D, D on the Coulomb quiz, because each of these corresponded to an answer that assumed $F \propto 1/r$ (instead of $F \propto 1/r^2$). Such a set of consistent responses gives no evidence of any confusion related strictly to the representation.

The error rates are low; 31% is the highest rate observed on any of the quiz questions in any one year, and the year-to-year fluctuations are substantial. The error rates on the circuits quiz are much higher than those on the Coulomb quiz. However, the mean error rates of different representations on the same quiz differed only slightly. Moreover, the relative ranking of the four representations with respect to error rate varied from year to year, and varied between the two quizzes in the same year. No one representation yielded the highest error rate consistently for all five years on either quiz.

Statistical comparisons were made between representations using a paired two-sample t-test[49] in which the error rates on, for instance, the V question on the Coulomb quiz were compared to those for the D question on the same quiz, for the sample of five pairs of error rates, one pair for each year. Of the 12 possible comparisons, that is, V versus D, V versus M, V versus G, D versus M, D versus G, and M versus G (all six on each quiz), only one difference between

Table III. Error rates on multi-representational quizzes, in percent; the proportion of all students giving incorrect responses to each of four quiz questions. ''Verbal'' corresponds to question 1 on both Coulomb quiz and circuits quiz; ''Diagrammatic'' corresponds to question 2, Coulomb quiz and question 3, circuits quiz; ''Mathematical'' corresponds to question 3, Coulomb quiz and question 2, circuits quiz; ''Graphical'' corresponds to question 4 on both Coulomb quiz and circuits quiz. ''Any Error'' corresponds to students who made an error on one or more of the quiz items, not including students who gave four incorrect responses that were clearly consistent with each other (see text). Error rates in the ''Average'' row were calculated from cumulated total errors (1998–2002) divided by the 5-year total number of students.

All students		*N*	Verbal	Diagrammatic	Mathematical	Graphical	Any Error
Coulomb quiz	1998	*71*	4	7	10	14	24
	1999	*91*	11	15	18	21	30
	2000	*79*	14	11	10	11	24
	2001	*75*	12	15	13	23	35
	2002	*67*	15	16	24	19	33
	Average		**11**	**13**	**15**	**18**	**29**
Circuits quiz	1998	*68*	24	18	28	31	49
	1999	*88*	22	18	22	31	53
	2000	*68*	15	19	15	18	31
	2001	*75*	19	24	24	24	48
	2002	*63*	22	13	13	19	32
	Average		**20**	**19**	**20**	**25**	**43**

the means was statistically significant at the $p=0.05$ level according to a two-tailed test. This difference was on the Coulomb quiz, D versus G ($p=0.03$).

The discrepancy that appears to be most consistent is that between the error rates on G and those on V, D, and M. The overall error rates on G, on both quizzes, are 5% higher than the combined V-D-M mean error rates on the respective quiz, while the differences among the mean error rates on V, D, and M are all $\leqslant 4\%$. This will be discussed further in Sec. V below.

D. Confidence levels

I attempted to assess students' confidence in their use of the various representations. Each question had an extra-credit option that allowed students with high confidence in the correctness of their response to gain additional points for a correct answer (see Appendices A and B). If this option is chosen, a correct answer is credited with 3.0 points instead of the 2.5 points it would be worth normally. However, there is a substantial penalty for an incorrect response. Instead of an incorrect answer being worth zero points, it is worth -1.0 points; that is, a deduction is taken from the student's total score. I analyzed students' responses on the extra-credit option to gauge their confidence with the various representations.

Students who gave a correct response but did not choose the extra-credit option are defined as giving a ''low-confidence correct'' response. This response suggests that although the student is able to find a correct answer, they lack full confidence in the correctness of their response. In Table IV, low-confidence correct responses are tabulated for each question on each quiz.

On both quizzes, the proportion of low-confidence correct responses on the V question is lower than that on the three other questions on the same quiz. The differences are not large, and so I tested the significance of the differences between low-confidence correct response rates on the V questions and those on the D, M, and G questions by employing a paired t-test. Each sample consisted of the five pairs (one for each year) of the error rates on the V question, and either the D, M, and G question, respectively, for a total of six comparisons (three for each quiz). The difference between the means was found significant at the $p\leqslant 0.01$ level (one-tailed test) for the V-D and V-G comparison on the Coulomb quiz, and $p\leqslant 0.05$ for the V-M and V-G comparison

Table IV. Correct but low-confidence responses: the proportion of students giving correct response but not choosing extra-credit option.

	1998–2002	Verbal	Diagrammatic	Mathematical	Graphical
Coulomb quiz	Number correct	340	333	326	315
	Low-confidence correct	17%	24%	22%	24%
Circuits quiz	Number correct	289	295	288	272
	Low-confidence correct	33%	37%	41%	45%

Table V. Consistency of responses: the students who took both quizzes and made one, two, or three errors on at least one quiz. A ''repeat'' error refers to an error on both quizzes for questions in a particular form of representation; ''≤50% repeat errors'' indicates that half or fewer of all incorrectly used representations (combined for both quizzes) were part of a repeat-error pair (see text). (Students who gave four incorrect but consistent responses on a single quiz as defined in the text were not counted as having made any errors on that quiz for the purposes of this tabulation.)

	N	Errors on one quiz only (*no* repeat errors)	Errors on both quizzes but *no* repeat errors	Errors on both quizzes, but ≤50% repeat errors	Errors on both quizzes, >50% repeat errors
2000	*23*	78%	9%	9%	4%
2001	*44*	73%	7%	14%	7%
2002	*26*	77%	12%	8%	4%

on the circuits quiz. Corresponding values for the remaining comparisons were $p = 0.10$ (V-M on the Coulomb quiz), and $p = 0.12$ (V-D on the circuits quiz). These results suggest that students had slightly greater confidence when responding correctly to questions posed in the V (''words only'') representation on these two quizzes. In comparison, among students responding *incorrectly*, lower-than-average confidence was associated with D and M responses on the circuits quiz.

E. Consistency of students' error

To explore whether a given student consistently made errors with the same form of representation, a subset of the data was examined in more detail. For the years 2000, 2001, and 2002, a tabulation was made of students who took both quizzes and made one, two, or three errors on at least one quiz. When students made four errors, there is no direct evidence as to whether they have—or have not—made a representation-related error (in contrast to a physics error).

Table VI. (a) Error rates on multi-representational quizzes, in percent; male students only. (b) Error rates on multi-representational quizzes, in percent; female students only.

		N	Verbal	Diagrammatic	Mathematical	Graphical	Any Error
			(a)				
Males							
Coulomb quiz	1998	*27*	7	7	7	11	26
	1999	*36*	6	11	11	11	14
	2000	*32*	13	16	9	13	22
	2001	*30*	10	10	10	10	31
	2002	*30*	17	10	30	20	30
	Average		**10**	**11**	**14**	**13**	**24**
Circuits quiz	1998	*27*	26	11	33	33	52
	1999	*35*	9	14	14	29	49
	2000	*29*	14	14	14	21	31
	2001	*28*	18	21	21	14	43
	2002	*28*	14	11	14	11	29
	Average		**16**	**14**	**19**	**22**	**41**
			(b)				
Females							
Coulomb quiz	1998	*44*	2	7	11	16	23
	1999	*55*	15	18	22	27	40
	2000	*47*	15	9	11	11	26
	2001	*45*	13	18	16	31	38
	2002	*37*	14	22	19	19	35
	Average		**12**	**14**	**16**	**21**	**32**
Circuits quiz	1998	*41*	22	22	24	29	46
	1999	*53*	30	21	26	32	57
	2000	*39*	15	23	15	15	31
	2001	*47*	19	26	26	30	51
	2002	*35*	29	14	11	26	34
	Average		**23**	**21**	**21**	**27**	**45**

Therefore, students who made four errors on either quiz (a very small proportion of students overall) are not counted in this tabulation. In contrast, students who gave four incorrect but consistent responses on a particular quiz were not counted as having made any errors on that quiz for the purposes of this analysis. These data are shown in Table V. A ''repeat'' error refers to an error on both quizzes for questions in a particular representation. If students made errors on V, D, and M on one quiz and D, M, and G on the other, 50% of their errors (two $[D,M]$ out of four $[V,D,M,G]$) are considered to be repeats. The statement ''$\leqslant$50% repeat errors'' in Table V indicates that half or fewer of all incorrectly used representations were part of a repeat-error pair.

The results of the three years are very consistent: most students made errors on one quiz only. Of those who made errors on both quizzes, most did not repeat the same error. That is, they did not make two errors using the same representation. If they did repeat an error, half or fewer of their representation errors were repeated. These data do not support the hypothesis that students tend to err consistently in one or another representation.

V. GENDER-RELATED DIFFERENCES

In Table VI, error rate data are shown for male, Table VI (a), and female, Table VI (b), students. This breakdown allows us to test for possible gender-related differences. We see that the mean error rates (average values, all years combined) for the female students are higher than those of the males, on all questions on both quizzes. In most cases, the male-female difference is relatively small. To gauge the statistical significance of the differences, a paired t-test was carried out separately for each question on each quiz, where each sample consisted of five pairs of values (male error rate, female error rate), one pair for each year.[49] This test also was done for the ''Any Error'' rate. Of these ten cases, the only difference in the mean error rate significant at the $p=0.05$ level with a two-tailed test was the D question on the circuits quiz (male: 14%, female: 21%, $p=0.008$). Due to the low statistical power of a test with a sample of only five pairs, and in view of the consistency of the observed male–female error rate difference, it may be more appropriate to use a $p \leqslant 0.10$ criterion and apply a one-tailed test. Two additional cases met that criterion: Coulomb quiz, G question (male: 13%, female: 21%, $p=0.08$), and Coulomb quiz, any error (male: 24%, female: 32%, $p=0.09$).

A noticeable contrast between the Table VI and Table III data is that the difference among the male students between the G error rate on the Coulomb quiz (13%) and the mean combined V-D-M error rate on the same quiz (12%) is much smaller than the corresponding difference in the ''all students'' sample (Table III). In contrast, a sizeable difference still exists for the female students (G: 21%; V-D-M: 14%). This observation suggests that the larger error rate on G (relative to V-D-M) in Table III is primarily due to the female students. It is not as clear whether this pattern may be true for the circuits quiz as well, for here a discrepancy is still present for males (G: 22%, V-D-M: 16%), as well as for females (G: 27%, V-D-M: 22%).

To examine this question more closely, I did three statistical tests. To probe the statistical significance of the observation that the G error rates are higher than V, D, or M error rates on the same quiz during the same year, I employed a Wilcoxon sign rank test.[50] This is a nonparametric test that does not depend on the shape of the distribution of sample values, and thus is less sensitive to deviations from normality in the data sample. In this test I considered all pairwise comparisons between the G error rate and the V, D, and M error rates, respectively, on a given quiz for a given year. This procedure yielded 15 comparisons on each quiz (three for each year), both for males and females. For instance, for male students on the Coulomb quiz, the G-V, G-D, and G-M pairs for 2000 were (0.13, 0.13), (0.13, 0.16), and (0.13, 0.09). For female students during the same year, the pairs were (0.11, 0.15), (0.11, 0.09), and (0.11, 0.11). The four samples and their resulting p values (for a two-tailed test) are Coulomb-male, $p>0.10$; Coulomb-female, $p<0.01$; Circuits-male, $p>0.10$; and Circuits-female, $p<0.02$; each sample consisted of 15 pairs of values. These results suggest that the error rates for females might be higher on G questions than on V-D-M questions.

A paired two-sample t-test was used to make a full set of 12 interrepresentation comparisons, separately for males and females. There were six on each quiz, that is, V versus D, V versus M, V versus G, D versus M, D versus G, and M versus G. Each sample consisted of five pairs of values, one for each year. No interrepresentation differences were found to be significant at the $p=0.05$ level using a two-tailed test. Several comparisons were significant at the $p \leqslant 0.10$ level using a one-tailed test; all p values corresponding to the one-tailed test are shown in Table VII.

Table VII. p values for statistical tests (one-tailed test) of the significance of differences between mean error rates on questions from the same quiz posed in different representations. The paired t-test and the test for correlated proportions are described in the text. These p values represent the probability that differences in mean error rates equal to or larger than those actually observed (but with the same sign) would occur in an ensemble of paired random samples of the same size, drawn from an infinitely large population in which the true difference in mean error rates is zero.

		Coulomb quiz		Circuits quiz	
		Paired t-test	Correlated proportions	Paired t-test	Correlated proportions
Females	G versus V	0.04	0.001	0.12	⋯
	G versus D	0.05	0.02	0.10	0.05
	G versus M	0.07	0.04	0.03	0.07
	V versus D	0.15	⋯	0.34	⋯
	V versus M	0.08	0.08	0.29	⋯
	D versus M	0.26	⋯	0.42	⋯
Males	G versus V	0.04	0.23	0.14	⋯
	G versus D	0.20	⋯	0.12	⋯
	G versus M	0.40	⋯	0.31	⋯
	V versus D	0.43	⋯	0.32	⋯
	V versus M	0.17	⋯	0.04	0.18
	D versus M	0.29	⋯	0.15	⋯

To examine these possibly significant comparisons more closely, a test for the difference between correlated proportions was applied.[51] With this method a test statistic z is calculated by comparing, for instance, the number of students (all five years) who were correct on the G question but incorrect on the V question (C_{GV}) to those who were incorrect on the G question but correct on the V question (C_{VG}). After applying a continuity correction,[52] we have $z=(|C_{GV}-C_{VG}|-1)/(C_{GV}+C_{VG})^{0.5}$. The calculated p values resulting from this statistic are shown in Table VII for those pairs that met the $p \leqslant 0.10$ criterion on the t-test.

Even with this wealth of statistical data, the conclusions remain ambiguous. However, the various results support the hypothesis that there is a discrepancy between the male and female students regarding the relative error rates on G questions in comparison to V-D-M questions, at least on the Coulomb quiz. On this quiz, the female students did more poorly on G questions in comparison to V-D-M questions, whereas the male students did not, or at least not as much. There also was support (noted above) for the hypothesis that female students perform more poorly on the diagrammatic question on the circuits quiz, in comparison to male students. Because the male and female students in this study received identical instruction, these results are potentially significant.

VI. DISCUSSION

A. Newton's third-law questions

The analysis of the gravitation quiz data leaves no doubt that there is a systematic discrepancy among students in this sample between their interpretation of the verbal and diagrammatic versions of the Newton's third-law question. Although the correct-response rate on the pretest version of the two questions varied substantially from year to year, the rate of correct responses on the diagrammatic version was never greater than 60% of that on the verbal version. A substantial majority (59%) of students who correctly answered the verbal version gave an incorrect response on the diagrammatic version. In the latter context they were influenced by the dominance principle that had not, apparently, determined their response to the verbal version. Written explanations on the electrostatic version of these questions on the 2002 final exam are consistent with this interpretation, although they do not directly support it.[53] (It is notable, however, that of the students who correctly answered the diagrammatic version of this question on the pretest, only 23% gave an incorrect response to the verbal version on the same test.)

Over the five years of this study, 59% of students who answered the Newton's third-law pretest question with a correct "equal-force" response on the verbal representation gave an "unequal-force" response on the diagrammatic representation. Yet the total number of such students is relatively small in comparison to the size of the full sample since only 16% of all students gave a correct response on the verbal pretest question. This discrepancy in response rates demonstrates how sharply divergent students' responses may be in different contexts[54]—even when the context is merely a different representation accompanied by slightly different wording.[55] However, this particular divergence is not representative of a large fraction of the student population. In contrast, the error corresponding to the "larger mass exerts a larger (smaller) force" response (described below) is one that characterizes a sizeable fraction—perhaps more than a third—of this population.

It was observed that response A on the diagrammatic question 8 of the gravitation quiz—what we call an "antidominance principle" response (larger mass exerts a smaller force)—represents more than 40% of responses to this question, while the corresponding D and E responses on the verbal question 1 represent only 5% of all responses to this question. The implication is that many students have an incorrect understanding of vector arrow conventions, that is, the arrow whose tail is attached to an object represents the force that is exerted on that object, not by it. This implication is strongly supported by the written explanations offered by students on the 2002 final exam questions.[56]

These observations are intriguing and important, and yet leave unanswered questions. What is still unclear is the precise nature of students' thinking that leads some to answer that the gravitational forces exerted by the sun and earth on each other are of equal magnitude, and yet moments later to select a vector diagram in which the interaction forces of earth and moon are clearly not the same. Similarly, the details of students' thinking regarding the representation of forces exerted on or by an object are not well understood. It is possible that confusion related to the specific words or phrases used in the gravitation questions has contributed to the differences observed in students' responses, independent of confusion introduced by the diagrammatic representation. Our experience suggests that extensive interviewing will be required to clarify these matters.

B. Multi-representational quizzes

The mean error rates on the Coulomb and circuits quizzes were consistently low (below 30% on each question), and year-to-year variations were high (up to 400%). These facts imply that statistical conclusions from this data set will have limited reliability. In particular, it would not be reasonable to generalize conclusions from these data to problem sets of significantly greater difficulty without further investigation. Most students in this data sample did not make errors on the test questions; therefore, one could argue that the interrepresentational competence of a substantial fraction of the population sample was not directly probed by these instruments. More difficult test questions (including non-multiple-choice items) that could probe a larger fraction of the population sample might yield conclusions that are different than, and even contradictory to, those discussed here.

Most students in this sample did not show a pattern of consistent representation-related errors on the multi-representational quizzes. The specific physics errors made by students were quite consistent; as discussed in Sec. IV, a large proportion of incorrect responses were concentrated on just one conceptual error on each quiz. However, the typical student made errors on only one or two questions (or none), and gave correct answers on the other questions. They typically did not make an error with the same representation on both quizzes, and this pattern of no repeat errors was consistent with results on the Newton's third-law questions discussed in Sec. III. The precise trigger that led a student to make a "standard" physics error when using one particular representation on a particular quiz—and not with any other representations, nor on a follow-up quiz—is unclear, and appeared to be almost random, both for individual students and for the students as a whole. On the Coulomb questions in 2001, for example, the number of students getting a D question incorrect later in the semester (after they had already answered it correctly earlier in the semester) was exactly matched by the number of students displaying the same pattern with the M questions. (See Sec. IV B).

There is evidence for slightly higher confidence rates on the verbal questions. This finding might surprise some, because many physics instructors would find the verbal version of the quiz questions to be awkward to interpret and analyze, in comparison to the D, M, and G versions based on very familiar and long practiced representations. This result suggests that the instructor's view of the ease or difficulty of a particular representation in a particular context might not match the views of a large proportion of students. The results of previous investigations regarding student understanding of kinematics diagrams[18,28–30] are consistent with this inference.

C. Gender differences

On the multi-representational quizzes, there is evidence that student performance on the G questions was slightly inferior to that on the V, D, and M questions. However, this evidence is strong only for female students on the Coulomb quiz. The poorer performance on G questions might be ascribed to less familiarity and practice with this representation. However, the instruction for both females and males was identical, and the relatively poorer performance by females on the G questions, at least on the Coulomb quiz, suggests a genuine performance discrepancy between the genders in the larger population. Whether this discrepancy may be due to different degrees of previous experience with G representations or some other cause is a matter for speculation. Similarly, the substantial evidence for poorer performance by females on the circuit-diagram question (D question; female error rate $=21\%$; male error rate $=14\%$) cannot be explained based on available information. The slightly higher error rates by females overall, in comparison to males, are not statistically significant for the most part.[57]

VII. CONCLUSION

We can summarize the results of this investigation as follows: (1) Some students did give inconsistent answers to the same question when it was asked using different representations; however, there was no clear evidence of a consistent pattern of representation-related errors among individual students. (2) Specific difficulties were noted when using vector representations in the context of Newton's third law. Many students apparently lacked an understanding of how to use vector arrows to distinguish forces acting *on* an object from forces exerted *by* that object. An apparently different difficulty was reflected by a smaller, though still substantial, number of students who gave a correct ‘‘equal-force’’ answer to a verbal question but an incorrect ‘‘unequal-force’’ answer to a very similar question using vector diagrams. (3) There was substantial evidence that females had a slightly higher error rate on graphical (bar chart) questions in comparison to verbal, diagrammatic, and mathematical questions, whereas the evidence for male students was more ambiguous. (4) Some evidence of possible gender-related differences was identified. Specifically, a possible difficulty related to electric circuit diagrams has been identified for females in comparison to males.

Although the observed error rate differences among the different representations were quite small or statistically insignificant in general, this result was in the context of a course that emphasized the use of multiple representations in all class activities. In addition, the overall error rates were quite low and suggest that the questions were too simple to probe possible representation-related difficulties among the majority of the students. What results might be found for students in a more traditional course which focuses on mathematical representations is an open question, as is the question of what results might be observed if significantly more challenging problems were posed.

However, this preliminary investigation has yielded at least one dramatic example of how student performance on very similar physics problems posed in different representations might yield strikingly different results (gravitation quiz, questions 1 and 8).[58] This ‘‘existence proof’’ serves as a caution that potential interrepresentational discrepancies in student performance must be carefully considered in the design and analysis of classroom exams and diagnostic test instruments. (This idea is already implicit in the work of many other authors cited in this paper.) For instance, if students are observed to make errors on Coulomb's law questions using a vector representation, representational confusion would be signaled by correct answers on closely related conceptual questions using other representations.

The evidence provided here for possible gender-related discrepancies in interrepresentational performance suggests that substantial additional investigation of this possibility is warranted, with a view toward possible implementation of appropriately modified instructional strategies. Many unanswered questions regarding the details of students' reasoning when using diverse representations must await more extensive data from interviews and analysis of students' written explanations.

ACKNOWLEDGMENTS

I am indebted to Leith Allen for many fruitful conversations and valuable insights regarding this work, and in particular for emphasizing the significance of the ‘‘larger mass exerts a smaller force’’ response discrepancy, and for designing the electrostatic version of the Newton's third-law problem discussed in Sec. III. She also carried out a series of interviews that added perspective to the analysis presented here, and carefully reviewed the manuscript. Larry Engelhardt carried out a series of interviews that shed additional light on the issues examined in this paper. Jack Dostal contributed to the analysis of the data from the gravitation quiz. This material is based in part on work supported by the National Science Foundation under Grant No. REC-0206683; this project is in collaboration with Thomas J. Greenbowe, co-principal investigator.

APPENDIX A

Coulomb quiz. Designations of representations, and correct answers: 1, Verbal, answer: A; 2, Diagrammatic, answer: A; 3, Mathematical, answer: E; 4, Graphical, answer: E.

Physics 112
Quiz #11
October 6, 2000

Name: ______________________

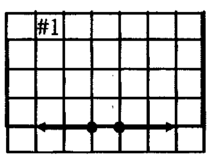
IF YOU WANT A QUESTION GRADED OUT OF THREE POINTS (–1 [MINUS ONE] FOR WRONG ANSWER!!) WRITE "3" IN SPACE PROVIDED ON EACH QUESTION.

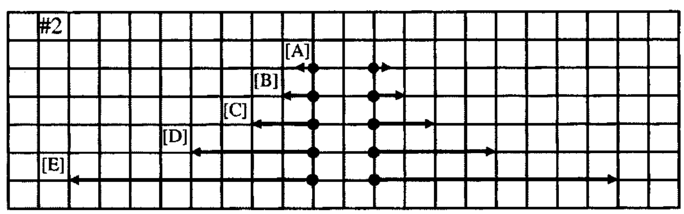
1. When two identical, isolated charges are separated by two centimeters, the magnitude of the force exerted by each charge on the other is eight newtons. If the charges are moved to a separation of eight centimeters, what will be the magnitude of that force now?

A. one-half of a newton
B. two newtons
C. eight newtons
D. thirty-two newtons
E. one hundred twenty-eight newtons

Grade out of three? Write "3" here: _____

2. Figure #1 shows two identical, isolated charges separated by a certain distance. The arrows indicate the forces exerted by each charge on the other. The same charges are shown in Figure #2. Which diagram in Figure #2 would be correct?





A.
B.
C.
D.
E.

Grade out of three? Write "3" here: _____

3. Isolated charges q_1 and q_2 are separated by distance r, and each exerts force F on the other. $q_1^{initial} = q_1^{final}$ and $q_2^{initial} = q_2^{final}$; $r^{initial} = 10m$; $r^{final} = 2m$. $F^{initial} = 25N$; $F^{final} = ?$

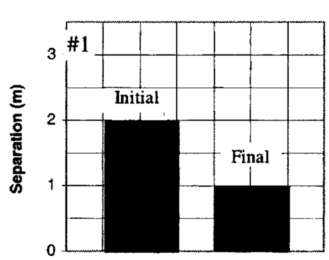
A. 1 N
B. 5 N
C. 25 N
D. 125 N
E. 625 N

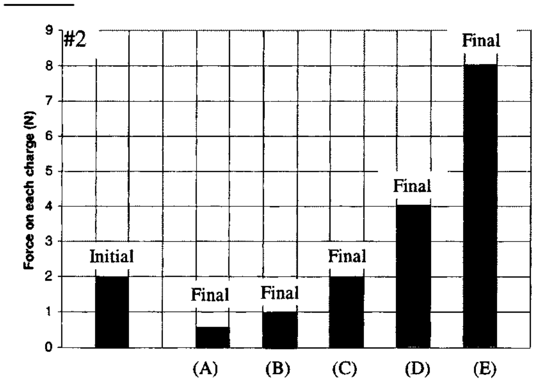
Grade out of three? Write "3" here: _____

4. Graph #1 refers to the initial and final separation between two identical, isolated charges. Graph #2 refers to the initial and final forces exerted by each charge on the other. Which bar is correct?

Grade out of three? Write "3" here: _____

A.
B.
C.
D.
E.





APPENDIX B

Circuits quiz. Designations of representations, and correct answers: 1, Verbal, answer: A; 2, Mathematical, answer: A; 3, Diagrammatic, answer: A; 4, Graphical, answer: C.

Physics 112
Quiz #16
October 27, 2000

Name: ______________________

IF YOU WANT A QUESTION GRADED OUT OF THREE POINTS (–1 [MINUS ONE] FOR WRONG ANSWER!!) WRITE "3" IN SPACE PROVIDED ON EACH QUESTION.

1. In a parallel circuit, a three-ohm resistor and a six-ohm resistor are connected to a battery. In a series circuit, a four-ohm and an eight-ohm resistor are connected to a battery that has the ***same*** voltage as the battery in the parallel circuit. What will be the ratio of the current through the six-ohm resistor to the current through the four-ohm resistor? Current through six-ohm resistor divided by current through four-ohm resistor is:

   A. greater than one
   B. equal to one
   C. less than one
   D. equal to negative one
   E. cannot determine without knowing the battery voltage

   Grade out of 3? Write "3" here: _____

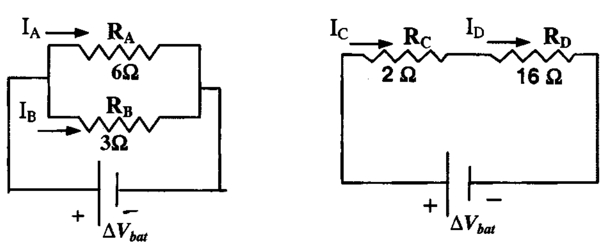
2. Parallel circuit: $R_A = 6\ \Omega$; $R_B = 9\ \Omega$.
   Series circuit: $R_C = 7\ \Omega$; $R_D = 3\ \Omega$.
   $\Delta V_{bat}(series) = \Delta V_{bat}(parallel)$

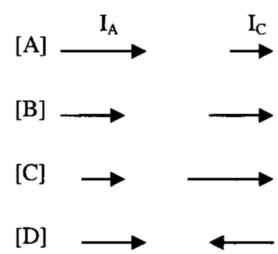
   A. $\frac{I_B}{I_C} > 1$ B. $\frac{I_B}{I_C} = 1$ C. $\frac{I_B}{I_C} < 1$ D. $\frac{I_B}{I_C} = -1$ E. need ΔV_{bat}

   Grade out of 3? Write "3" here: _____

3. The arrows represent the magnitude and direction of the current through resistors A and C. Choose the correct diagram.

   A.
   B.
   C.
   D.
   E. need to know ΔV_{bat}



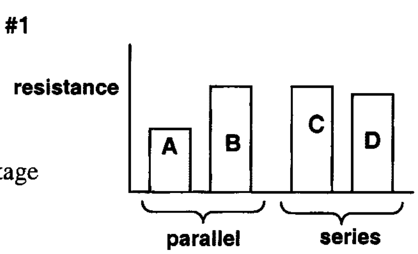


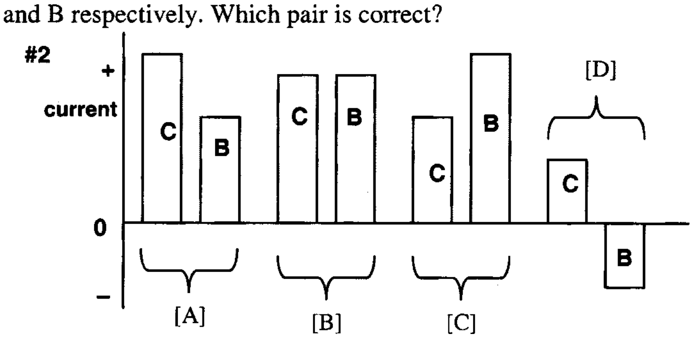
[E] (need to know ΔV_{bat})

Grade out of 3? Write "3" here: _____

4. Graph #1 represents the relative resistances of resistors A, B, C, and D. Resistors A and B are connected in a parallel circuit. Resistors C and D are connected in a series circuit. The battery voltage in both circuits is the same. Graph #2 represents the currents in resistors C and B respectively. Which pair is correct?

A.
B.
C.
D.
E. need to know voltage





Grade out of 3? Write "3" here _____

[a]Electronic mail: dem@iastate.edu

[1]Alan Van Heuvelen, ''Learning to think like a physicist: A review of research-based instructional strategies,'' Am. J. Phys. **59**, 891–897 (1991); ''Overview, Case Study Physics,'' *ibid.* **59**, 898–907 (1991).

[2]David Hestenes, ''Modeling methodology for physics teachers,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 935–957 (1997)], pt. 2.

[3]Lillian C. McDermott, ''A view from physics,'' in *Toward a Scientific Practice of Science Education*, edited by M. Gardner, J. G. Greeno, F. Reif, A. H. Schoenfeld, A. diSessa, and E. Stage (L. Erlbaum, Hillsdale, NJ, 1990), pp. 3–30.

[4]In this investigation we will concentrate on the representations V, D, M, and G. In this paper G refers to bar charts and not to line graphs. ''Graphical'' is used in a broad sense to refer to bar charts because they often are grouped together with line graphs, but there are very significant differences between the two representations that would have to be considered in future work. We restrict ourselves primarily to these four representations for practical and logistical reasons. There are certainly other pedagogically significant representations, for example, pictorial representations, computer animations and simulations, haptic (sense of touch) and kinesthetic interfaces and representations, video recordings of actual physical processes, and actual physical objects and systems using laboratory equipment. All of these are under investigation by many research groups (including ours), but they lack the relative standardization (due to long-term use in instruction) and ease and flexibility of implementation that characterize V, D, M, and G. Historically, V, D, M, and G have been ubiquitous in scientific work.

[5]M. T. H Chi, P. J. Feltovich, and R. Glaser, ''Categorization and representation of physics problems by experts and novices,'' Cogn. Sci. **5**, 121–152 (1981); Yuichiro Anzai, ''Learning and use of representations for physics expertise,'' in *Toward a General Theory of Expertise*, edited by K. Anders Ericsson and Jacqui Smith (Cambridge U. P., Cambridge, 1991), pp. 64–92.

[6]Jill H. Larkin, ''The role of problem representation in physics,'' in *Mental Models*, edited by Dedre Gentner and Albert L. Stevens (L. Erlbaum, Hillsdale, NJ, 1983), pp. 75–98.

[7]David P. Maloney, ''Research on problem solving: Physics,'' in *Handbook of Research on Science Teaching and Learning*, edited by Dorothy L. Gabel (Macmillan, New York, 1993), pp. 327–354.

[8]R. Kleinman, H. Griffin, and N. K. Kerner, ''Images in chemistry,'' J. Chem. Educ. **64**, 766–770 (1987); Robert Kozma, Elaine Chin, Joel Russell, and Nancy Marx, ''The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning,'' J. Learn. Sci. **9**, 105–143 (2000).

[9]Frederick Reif, ''Millikan Lecture 1994: Understanding and teaching important scientific thought processes,'' Am. J. Phys. **63**, 17–32 (1995).

[10]Alan Van Heuvelen, *ALPS Kit: Active Learning Problem Sheets, Mechanics; Electricity and Magnetism* (Hayden-McNeil, Plymouth, MI, 1990); Ruth W. Chabay and Bruce A. Sherwood, *Matter & Interactions I, II* (J. Wiley, New York, 2002); Frederick Reif, *Understanding Basic Mechanics* (J. Wiley, New York, 1995); Lillian C. McDermott and the Physics Education Group, *Physics by Inquiry* (J. Wiley, New York, 1996); Randall D. Knight, *Physics for Scientists and Engineers: A Strategic Approach* (Pearson Addison-Wesley, San Francisco, 2004); Eric Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997); Gregor M. Novak, Evelyn T. Patterson, Andrew D. Gavrin, and Wolfgang Christian, *Just-In-Time Teaching: Blending Active Learning with Web Technology* (Prentice Hall, Upper Saddle River, NJ, 1999); Thomas L. O'Kuma, David P. Maloney, and Curtis Hieggelke, eds., *Ranking Task Exercises in Physics* (Prentice Hall, Upper Saddle River, NJ, 2000); Wolfgang Christian and Mario Belloni, *Physlets: Teaching Physics with Interactive Curricular Material* (Prentice Hall, Upper Saddle River, NJ, 2001); Lillian C. McDermott, Peter S. Shaffer, and the Physics Education Group, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 2002-2003).

[11]Fred Goldberg, ''Constructing physics understanding in a computer-supported learning environment,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 903–911 (1997)], pt. 2.

[12]R. R. Hake, ''Promoting student crossover to the Newtonian world,'' Am. J. Phys. **55**, 878–884 (1987); Richard R. Hake, ''Socratic pedagogy in the introductory physics laboratory,'' Phys. Teach. **30**, 546–552 (1992); Ronald K. Thornton and David R. Sokoloff, ''Learning motion concepts using real-time microcomputer-based laboratory tools,'' Am. J. Phys. **58**, 858–867 (1990); Priscilla W. Laws, ''Calculus-based physics without lectures,'' Phys. Today **44(12)**, 24–31 (1991); ''Millikan Lecture 1996: Promoting active learning based on physics education research in introductory physics courses,'' Am. J. Phys. **65**, 14–21 (1997).

[13]Patricia Heller and Mark Hollabaugh, ''Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups,'' Am. J. Phys. **60**, 637–644 (1992).

[14]Robert J. Beichner, ''The impact of video motion analysis on kinematics graph interpretation skills,'' Am. J. Phys. **64**, 1272–1277 (1996).

[15]Robert J. Dufresne, William J. Gerace, and William J. Leonard, ''Solving physics problems with multiple representations,'' Phys. Teach. **35**, 270–275 (1997); Lawrence T. Escalada and Dean A. Zollman, ''An investigation on the effects of using interactive digital video in a physics classroom on student learning and attitudes,'' J. Res. Sci. Teach. **34**, 467–489 (1997); Melissa Dancy, Wolfgang Christian, and Mario Belloni, ''Teaching with Physlets: Examples from optics,'' Phys. Teach. **40**, 494–499 (2002); Anne J. Cox, Mario Belloni, Melissa Dancy, and Wolfgang Christian, ''Teaching thermodynamics with Physlets in introductory physics,'' Phys. Educ. **38**, 433–440 (2003).

[16]David E. Meltzer and Kandiah Manivannan, ''Promoting interactivity in physics lecture classes,'' Phys. Teach. **34**, 72–78 (1996); ''Transforming the lecture-hall environment: The fully interactive physics lecture,'' Am. J. Phys. **70**, 639–654 (2002).

[17]K. A. Burke, Thomas J. Greenbowe, and Mark A. Windschitl, ''Developing and using conceptual computer animations for chemistry instruction,'' J. Chem. Educ. **75**, 1658–1661 (1998); Thomas J. Greenbowe, ''An interactive multimedia software program for exploring electrochemical cells,'' J. Chem. Educ. **71**, 555–557 (1994); Michael J. Sanger and Thomas J. Greenbowe, ''Addressing student misconceptions concerning electron flow in aqueous solutions with instruction including computer animations and conceptual change strategies,'' Int. J. Sci. Educ. **22**, 521–537 (2000); Hsin-kai Wu, Joseph S. Krajcik, and Elliot Soloway, ''Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom,'' J. Res. Sci. Teach. **38**, 821–842 (2001).

[18]Robert J. Beichner, ''Testing student interpretation of kinematics graphs,'' Am. J. Phys. **62**, 750–762 (1994).

[19]John Clement, ''Observed methods for generating analogies in scientific problem solving,'' Cogn. Sci. **12**, 563–586 (1988).

[20]Rolf Plötzner, *The Integrative Use of Qualitative and Quantitative Knowledge in Physics Problem Solving* (Peter Lang, Frankfurt am Main, 1994), pp. 33–46.

[21]Ronald K. Thornton and David R. Sokoloff, ''Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula,'' Am. J. Phys. **66**, 338–352 (1998).

[22]Alan Van Heuvelen and Xueli Zou, ''Multiple representations of work-energy processes,'' Am. J. Phys. **69**, 184–194 (2001); Xueli Zou, ''The role of work-energy bar charts as a physical representation in problem solving,'' in *Proceedings of the 2001 Physics Education Research Conference*, edited by Scott Franklin, Jeffrey Marx, and Karen Cummings (PERC, Rochester, NY, 2001), pp. 135–138.

[23]Allan Paivio, *Imagery and Verbal Processes* (Holt, Rinehart and Winston, New York, 1971).

[24]Claude Janvier, ed., *Problems of Representation in the Teaching and Learning of Mathematics* (L. Erlbaum, Hillsdale, NJ, 1987).

[25]Richard Lesh, Tom Post, and Merlyn Behr, ''Representations and translations among representations in mathematics learning and problem solving,'' in *Problems of Representation in the Teaching and Learning of Mathematics*, edited by Claude Janvier (L. Erlbaum, Hillsdale, NJ, 1987), pp. 33–40; Paul White and Michael Mitchelmore, ''Conceptual knowledge in introductory calculus,'' J. Res. Math. Educ. **27**, 79–95 (1996); Peter C.-H. Cheng, ''Unlocking conceptual learning in mathematics and science with effective representational systems,'' Comput. Educ. **33**, 109–130 (1999); Shaaron Ainsworth, ''The functions of multiple representations,'' *ibid.* **33**, 131–152 (1999).

[26]Robert B. Kozma and Joel Russell, ''Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena,'' J. Res. Sci. Teach. **34**, 949–968 (1997).

[27]Donald R. Jones and David A. Schkade, ''Choosing and translating be-

tween problem representations,'' Organ. Behav. Human Decision Process. **61**, 214–223 (1995).

[28]Janice R. Mokros and Robert F. Tinker, ''The impact of microcomputer-based labs on children's ability to interpret graphs,'' J. Res. Sci. Teach. **24**, 369–383 (1987); Lillian C. McDermott, Mark L. Rosenquist, and Emily H. Van Zee, ''Student difficulties in connecting graphs and physics: Examples from kinematics,'' Am. J. Phys. **55**, 503–513 (1987); Fred M. Goldberg and John H. Anderson, ''Student difficulties with graphical representation of negative values of velocity,'' Phys. Teach. **27**, 254–260 (1989).

[29]Craig A. Berg and Philip Smith, ''Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments,'' Sci. Educ. **78**, 527–554 (1994).

[30]Italo Testa, Gabriella Mouroy, and Elena Sassi, ''Students' reading images in kinematics: The case of real-time graphs,'' Int. J. Sci. Educ. **24**, 235–256 (2002).

[31]For example, see: G. J. Hitch, M. C. Beveridge, S. E. Avons, and A. T. Hickman, ''Effects of reference domain in children's comprehension of coordinate graphs,'' in *The Acquisition of Symbolic Skills*, edited by Don Rogers and John A. Sloboda (Plenum, New York, 1982), pp. 551–560.

[32]Norman H. Fredette and John J. Clement, ''Student misconceptions of an electric circuit: What do they mean?,'' J. Coll. Sci. Teach. **10**, 280–285 (1981); Samuel Johsua, ''Students' interpretation of simple electrical diagrams,'' Eur. J. Sci. Educ. **6**, 271–275 (1984); Lillian C. McDermott and Peter S. Shaffer, ''Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding,'' Am. J. Phys. **60**, 994–1003 (1992); erratum, **61**, 81 (1993); Peter S. Shaffer and Lillian C. McDermott, ''Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies,'' *ibid.* **60**, 1003–1013 (1992); S. Törnkvist, K.-A. Pettersson, and G. Tranströmer, ''Confusion by representation: On student's comprehension of the electric field concept,'' *ibid.* **61**, 335–338 (1993); Randal Robert Harrington, ''An investigation of student understanding of electric concepts in the introductory university physics course,'' Ph.D. dissertation, University of Washington (UMI, Ann Arbor, MI, 1995), UMI #9537324, Chap. 5; Stephen Emile Kanim, ''An investigation of student difficulties in qualitative and quantitative problem solving: Examples from electric circuits and electrostatics,'' Ph.D. dissertation, University of Washington (UMI, Ann Arbor, MI, 1999), UMI #9936436, Chaps. 4–7; Leith Dwyer Allen, ''An investigation into student understanding of magnetic induction,'' Ph.D. dissertation, The Ohio State University (UMI, Ann Arbor, MI, 2001), UMI #3011018, Chap. 6; Rasil Warnakulasooriya and Lei Bao, ''Towards a model-based diagnostic instrument in electricity and magnetism–an example,'' *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, August 7-8, 2002]*, edited by Scott Franklin, Karen Cummings, and Jeffrey Marx (PERC, New York, 2002), pp. 83–86.

[33]J. Ramadas, ''Use of ray diagrams in optics,'' School Sci. **10**, 1–8 (1982); Fred M. Goldberg and Lillian C. McDermott, ''Student difficulties in understanding image formation by a plane mirror,'' Phys. Teach. **24**, 472–481 (1986); ''An investigation of student understanding of the real image formed by a converging lens or concave mirror,'' Am. J. Phys. **55**, 108–119 (1987); P. Colin and L. Viennot, ''Using two models in optics: Students' difficulties and suggestions for teaching,'' *ibid.* **69**, S36–S44 (2001); Philippe Colin, Françoise Chauvet, and Laurence Viennot, ''Reading images in optics: students' difficulties and teachers' views,'' Int. J. Sci. Educ. **24**, 313–332 (2002).

[34]Glenda Jacobs, ''Word usage misconceptions among first-year university physics students,'' Int. J. Sci. Educ. **11**, 395–399 (1989); P. Kenealy, ''A syntactic source of a common 'misconception' about acceleration,'' in *Proceedings of the Second International Seminar: Misconceptions and Educational Strategies in Science and Mathematics III* (Cornell Univ., Ithaca, NY, 1987), pp. 278–292; Jerold S. Touger, ''When words fail us,'' Phys. Teach. **29**, 90–95 (1991); H. Thomas Williams, ''Semantics in teaching introductory physics,'' Am. J. Phys. **67**, 670–680 (1999).

[35]Patricia F. Keig and Peter A. Rubba, ''Translations of representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge,'' J. Res. Sci. Teach. **30**, 883–903 (1993).

[36]W. L. Yarroch, ''Student understanding of chemical equation balancing,'' J. Res. Sci. Teach. **22**, 449–459 (1985).

[37]Andrew Elby, ''What students' learning of representations tells us about constructivism,'' J. Math. Behav. **19**, 481–502 (2000).

[38]Jiajie Zhang, ''The nature of external representations in problem solving,'' Cogn. Sci. **21**, 179–217 (1997); Maarten W. van Someren, Peter Reimann, Henry P. A. Boshuizen, and Ton de Jong, editors, *Learning with Multiple Representations* (Pergamon, Amsterdam, 1998); Jeff Zacks and Barbara Tversky, ''Bars and lines: A study of graphic communication,'' Mem. Cognit. **27**, 1073–1079 (1999); Bruce L. Sherin, ''How students invent representations of motion: A genetic account,'' J. Math. Behav. **19**, 399–441 (2000); Andrea A. diSessa and Bruce L. Sherin, ''Meta-representation: An introduction,'' *ibid.* **19**, 385–398 (2000); Roser Pintó and Jaume Ametller, ''Students' difficulties in reading images. Comparing results from four national research groups,'' Int. J. Sci. Educ. **24**, 333–341 (2002); Tae-Sun Kim and Beom-Ki Kim, ''Secondary students' cognitive processes for line graphs and their components,'' in *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, August 7–8, 2002]*, edited by Scott Franklin, Karen Cummings, and Jeffrey Marx (PERC, New York, 2002), pp. 91–94.

[39]A comparison of this type was made by Fernando Hitt, ''Difficulties in the articulation of different representations linked to the concept of function,'' J. Math. Behav. **17**, 123–134 (1998).

[40]Melissa Hayes Dancy, ''Investigating animations for assessment with an animated version of the Force Concept Inventory,'' Ph.D. dissertation, North Carolina State University (UMI, Ann Arbor, MI, 2000), UMI #9982749.

[41]David E. Meltzer, ''Comparative effectiveness of conceptual learning with various representational modes,'' AAPT Announcer **26(4)**, 46 (1996); ''Effectiveness of instruction on force and motion in an elementary physics course based on guided inquiry,'' *ibid.* **28(2)**, 125 (1998); Antti Savinainen and Jouni Viiri, ''A case study evaluating students' representational coherence of Newton's first and second laws,'' in *2003 Physics Education Research Conference [Madison, Wisconsin, August 6–7, 2003]*, edited by Jeffrey Marx, Scott Franklin, and Karen Cummings [AIP Conf. Proc. **720**, 77–80 (2004)].

[42]Robert B. Kozma, ''The use of multiple representations and the social construction of understanding in chemistry,'' in *Innovations in Science and Mathematics Education*, edited by Michael J. Jacobson and Robert B. Kozma (L. Erlbaum, Mahwah, NJ, 2000), pp. 11–46.

[43]Teresa Larkin-Hein, ''Learning styles in introductory physics: Enhancing student motivation, interest, & learning,'' in Proceedings of the International Conference on Engineering and Computer Education, São Paulo, Brazil (2000), ⟨http://nw08.american.edu/~tlarkin/larkin.htm⟩.

[44]Maria Kozhevnikov, Mary Hegarty, and Richard Mayer, ''Spatial abilities in problem solving in kinematics,'' in *Diagrammatic Representation and Reasoning*, edited by Michael Anderson, Bernd Meyer, and Patrick Olivier (Springer, London, 2002), pp. 155–171; Eun-Mi Yang, Thomas Andre, and Thomas J. Greenbowe, ''Spatial ability and the impact of visualization/animation on learning electrochemistry,'' Int. J. Sci. Educ. **25**, 329–349 (2003).

[45]A preliminary analysis of some of these data has been published previously. David E. Meltzer, ''Issues related to data analysis and quantitative methods in PER,'' in *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, August 7-8, 2002]*, edited by Scott Franklin, Karen Cummings, and Jeffrey Marx (PERC, New York, 2002), pp. 21–24.

[46]The ''dominance principle'' (a term used by Halloun and Hestenes) refers to students' tendency to attribute larger-magnitude forces to one or the other object in an interacting pair, based on an ostensibly ''dominant'' property such as greater mass, velocity, or charge. See David P. Maloney, ''Rule-governed approaches to physics—Newton's third law,'' Phys. Educ. **19**, 37–42 (1984); Ibrahim Abou Halloun and David Hestenes, ''Common-sense concepts about motion,'' Am. J. Phys. **53**, 1056–1065 (1985); Lei Bao, Kirsten Hogg, and Dean Zollman, ''Model analysis of fine structures of student models: An example with Newton's third law,'' *ibid.* **70**, 766–778 (2002).

[47]This result suggests that some students' expertise in using vector representations may have increased faster than did their understanding of Newton's third law, because response B is an accurate representation of an answer based on the dominance principle.

[48]Question #2 in this set was designed by Leith Allen, private communication (2002).

[49]J. P. Guilford, *Fundamental Statistics in Psychology and Education*, 4th ed. (McGraw-Hill, New York, 1965), p. 184. This test considers each pair of values to be an independent measurement of the difference between the paired quantities. It is the appropriate test here because there are many

year-to-year variations (in student demographics, course logistics, etc.) but in each individual year, there is no *a priori* reason to expect differences between the paired quantities.

[50]Reference 49, p. 255.

[51]Reference 49, pp. 188–189.

[52]David J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures* 2nd ed. (Chapman & Hall/CRC, Boca Raton, 2000), p. 498.

[53]We have tried to further test this interpretation with interview data [Leith Allen and Larry Engelhardt, private communication (2003)]. Approximately 15 students were interviewed in all; they had volunteered in response to a general solicitation. None of the students interviewed showed any clear evidence of the representation-related difficulties identified in this paper. Our experience (and that of others) has been that most students who volunteer for interviews are well above the average in terms of course performance. It seems that the relatively simple nature of the questions used in this investigation (indicated by the low error rates) was an inadequate challenge for the interview volunteers. It will probably be necessary to target potential interviewees in the future, soliciting students who have already shown (on quizzes or exams) evidence of the learning difficulties being investigated.

[54]Lei Bao and Edward F. Redish, ''Concentration analysis: A quantitative assessment of student states,'' Am. J. Phys. **69**, S49–S53 (2001). Also see Ref. 45.

[55]Although the *V* and *D* versions of the gravitation question (and related Coulomb's law question) include similar options regarding force magnitudes, the *D* version obviously portrays directional information as well. This directional information is an additional bit of complexity which probably contributes to overall confusion, although it is not clear how (or whether) it might make it more difficult for a student to pick out an ''equal magnitudes'' option.

[56]This convention—that the tail of the arrow representing a force exerted on an object is attached to the object—is certainly not universal. However, in the context of question 8, the attractive nature of the gravitational force guarantees that the force exerted on an object must point toward the other object in the interacting pair. This fact makes the assignment of force vector arrows in question 8 unambiguous; regardless of the convention for locating the tails of the arrows, the arrow corresponding to the force exerted on the moon must point toward the earth. Therefore, it is not merely a confusion about notation or vector conventions that leads to the error identified here. [It is notable that not a single student chose either response *G* or *H* on the electrostatic final-exam question (Fig. 2); these responses would be acceptable representations of a dominance-principle answer, or the correct answer, respectively, if one ignored tail location.] This observation leaves open the question of whether the students' confusion was primarily with the tail location, the meaning of the arrow direction itself, the meaning of ''attractive force,'' or some amalgam of these (and possibly other) issues.

[57]Most gender-related differences in this study seem to be smaller than the differences documented to exist between traditional instruction and interactive-engagement instruction; see, for instance, Richard R. Hake, ''Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,'' Am. J. Phys. **66**, 64–74 (1998), ⟨http://www.physics.indiana.edu/~sdi/⟩. Marshall has recently reported on a study that suggests the existence of gender differences in interpretation of electric circuit diagrams: Jill Marshall, ''Gender differences in representations of electric circuits,'' AAPT Announcer **34(4)**, 96 (2004).

[58]However, one must also consider the possibility that specific differences in the way the questions were worded also may have contributed significantly to the discrepancies in responses that were observed.

VI.

Methodological Issues in PER

The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores

David E. Meltzer
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011

(Received 27 July 2001; accepted 23 August 2002)

There have been many investigations into the factors that underlie variations in individual student performance in college physics courses. Numerous studies report a positive correlation between students' mathematical skills and their exam grades in college physics. However, few studies have examined students' learning gain resulting from physics instruction, particularly with regard to qualitative, conceptual understanding. We report on the results of our investigation into some of the factors, including mathematical skill, that might be associated with variations in students' ability to achieve conceptual learning gains in a physics course that employs interactive-engagement methods. It was found that students' normalized learning gains are not significantly correlated with their pretest scores on a physics concept test. In contrast, in three of the four sample populations studied it was found that there is a significant correlation between normalized learning gain and students' preinstruction mathematics skill. In two of the samples, both males and females independently exhibited the correlation between learning gain and mathematics skill. These results suggest that students' initial level of physics concept knowledge might be largely unrelated to their ability to make learning gains in an interactive-engagement course; students' preinstruction algebra skills might be associated with their facility at acquiring physics conceptual knowledge in such a course; and between-class differences in normalized learning gain may reflect not only differences in instructional method, but student population differences ("hidden variables") as well.

[DOI: 10.1119/1.1514215]

I. INTRODUCTION

A primary goal of research in physics education is to identify potential and actual obstacles to student learning, and then to address these obstacles in a way that leads to more effective learning. These obstacles include factors that originate during instruction—such as instructional method—as well as those that relate to students' preinstruction preparation. Previous studies have examined various preinstruction factors that may or may not be related to students' performance in physics, with mathematics skill being the most common factor. However, in almost all of these studies, the measures of performance adopted were student grades on course exams that emphasized quantitative problem solving. Only in a few cases was students' conceptual knowledge assessed through the use of qualitative problems. And with only a handful of exceptions, there was no attempt to directly measure the gain in student understanding that resulted from instruction.

This paper examines students' mathematics skills and their initial physics conceptual knowledge as factors that may underlie variations in student learning. Learning gain is assessed through pre- and post-testing using a qualitative test of physics conceptual knowledge. One objective of the present study is to determine whether individual students' learning gains are correlated with their initial level of conceptual knowledge as measured by pretest scores on the physics concept test. Another objective is to determine whether those learning gains are correlated with the students' mathematics skills, as determined by preinstruction testing with a college entrance exam or an algebra/trigonometry skills exam.

In Secs. II and III, I review the results and limitations of previous studies on the relation of students' pre-instruction preparation to their performance in physics courses. In Sec. IV I describe a widely adopted measure of student learning called "normalized learning gain" and explain why it is an appropriate measure for the objectives of this study. In Sec. V various factors that may be related to learning gain are discussed, and the motivation of the present study is presented. The context, methods, and results of the present study are described in Secs. VI, VII, and VIII, respectively, and the results are discussed in Sec. IX. The limitations of this study are outlined in Sec. X, and implications for instruction are examined in Sec. XI. The methodological implications of this study for physics education research are addressed in Sec. XII, and Sec. XIII briefly summarizes the main results.

II. PREVIOUS RESEARCH ON THE RELATION OF VARIOUS FACTORS TO STUDENTS' PERFORMANCE IN PHYSICS COURSES

A. Students' mathematical preparation

Many studies appear to show that mathematical ability (mathematical aptitude or accumulated procedural knowledge) is positively correlated to success in traditional introductory physics courses that emphasize quantitative problem solving. Most of these studies have involved college physics students; some have examined the preparation that these students received in high school. Some studies have found a positive correlation between physics course grades and scores on the mathematics part of college entrance exams.[1,2] Many investigators have found positive correlations between grades in college physics and a mathematics skills pretest administered at or near the very beginning of the course. Typically, these pretests involve algebra and trigonometry, although most investigators do not provide samples of their tests.[3–8]

The correlation between mathematics skill and physics performance has not been observed to hold consistently. Reported correlation coefficients vary widely and are not statistically significant for all groups tested. For example, one study found that the overall correlation between grades and an algebra pretest was not significant for males ($r = +0.10$), while for females the correlation was highly significant ($r = +0.48$).[8]

All the studies cited have focused on student performance either on a single physics course exam or on a mean grade from several such exams. In contrast, Hake *et al.*[9] and Thoresen and Gross[10] have reported preliminary investigations of student learning gains in physics courses, determined by both preinstruction and post-instruction testing. They found that students with the highest learning gains in physics had scored higher on a mathematics skills test than students with the lowest learning gains.

Several investigators have found positive correlations between grades earned by students in their college physics courses and their previous experience and/or grades in either high-school, college mathematics courses, or high-school physics courses.[11,12] However, the overall weight of the literature on factors related to college students' performance in introductory physics is that the measurable impact on performance is substantially larger for mathematics skills as determined by preinstruction testing, than it is from any measure derived simply from students' experience or lack of it in previous physics or mathematics courses.

B. Students' reasoning skills and other factors

Another factor that has been studied extensively is the possible relation between precourse measures of students' reasoning ability and their college physics grades. Significant correlations between these variables have been reported by numerous investigators.[2,4–6,8,13] However, the reported correlations are not significant for all groups, and in most cases the reports do not provide samples of the specific questions used to assess reasoning ability. Recently, Clement[14] has reported a positive correlation between a pretest measure of reasoning ability and learning gain in a high-school physics course.

Other factors that have been found significant to one degree or another are students' achievement expectations,[15] homework grades,[6] high-school GPA,[11,12] college GPA,[16] and a variety of cognitive and emotional factors.[17] A large number of significant preparation and demographic factors were identified by Sadler and Tai.[12] Two studies[4,7] found that students' performance on a pretest of physics conceptual knowledge had a significant positive correlation with course grades.

III. LIMITATIONS OF PREVIOUS RESEARCH

Almost all of the investigations discussed in Sec. II used students' scores (or grades derived from those scores) on physics course exams as a performance measure. It is very likely that in most cases, all or most of the exam questions would be described as traditional quantitative physics problems, although in most cases the nature of the questions was not discussed explicitly. There is by now a large body of literature[18–24] that demonstrates convincingly that good performance on such problems does not necessarily indicate good understanding of the physics concepts involved. Performance on such traditional problems may not even be highly *correlated* with conceptual understanding.[24] The author's conclusion is that virtually all previously published studies on the relationship between mathematics preparation and physics course performance leave open the question of how, and whether, such preparation may be related to conceptual understanding of physics.

Although various factors—such as mathematics preparation—may be correlated with students' performance on physics exams, this correlation is not direct evidence that there is a causal relationship between the two. To our knowledge, no studies directly test for such a relation. Therefore, it would be improper to conclude from previous studies that, for instance, requiring students to practice and improve their mathematics skills before beginning college physics would necessarily improve their performance in these courses.

Another important limitation of previous research is its failure to examine student learning. A student's performance on a course exam is an indication of the student's knowledge state at the time of the exam, and is not necessarily related to what the student has learned in a particular course. Hence, it is necessary to have some measure of student learning, in contrast to a measure that merely quantifies students' knowledge. One way to provide such a measure is to test students both at the beginning and at (or near) the end of a course to assess how much they may have learned. In this way we can obtain a measure of students' learning gain, which is the quantity that, in principle, is most susceptible to change by actions of the instructor and students during the course. Students' performance on course exams may or may not be correlated with learning gain, and the relationship between performance and learning gain is, at best, an indirect one. Nearly all previous studies have failed to directly investigate the possible relationship of mathematics (and other) preparation to students' learning gain in a college physics course.

IV. NORMALIZED LEARNING GAIN: A KEY MEASURE OF STUDENT LEARNING

The question of how to measure learning gain is not simple and is subject to many methodological difficulties.[25] Because the maximum on a diagnostic instrument is 100%, it is common to observe a strong negative correlation between students' absolute gain scores (posttest minus pretest score) and their pretest scores: higher pretest scores tend to result in smaller absolute gains, all else being equal. For example, in Hake's study of 62 introductory physics courses, absolute gain scores on the Force Concept Inventory (FCI) were significantly (negatively) correlated with pretest score ($r = -0.49$).[20] An alternative is to normalize the gain score to account for the variance in pretest scores. Such a measure is g, the normalized gain, which is the absolute gain divided by the maximum possible gain:

$$g = \frac{\text{post-test score} - \text{pretest score}}{\text{maximum possible score} - \text{pretest score}}.$$

Hake found that $\langle g \rangle$, the mean normalized gain, on the FCI for a given course was almost completely uncorrelated (r

$=+0.02$) with the mean pretest score of the students in the course.[26] Therefore, the normalized gain seems to be relatively independent of pretest score. This independence leads us to expect that if a diverse set of classes has a wide range of pretest scores but all other learning conditions are similar, the values of normalized learning gain measured in the different classes would not differ significantly. This pretest independence of the normalized gain also suggests that a measurement of the difference in $\langle g \rangle$ between two classes having very different pretest scores would be reproduced even by a somewhat different test instrument which results in a shifting of pretest scores.

Empirical evidence for this hypothesis is provided by an analysis of the data from Table II of Ref. 21. Students' knowledge of mechanics concepts was tested with two different diagnostic instruments, the FCI, and the Force and Motion Conceptual Evaluation (FMCE).[22] The pretest scores and absolute gain scores yielded by the two instruments were significantly different, but the normalized gains were statistically indistinguishable. The most persuasive empirical support for use of $\langle g \rangle$ as a valid and reliable measure is that $\langle g \rangle$ has now been measured for tens of thousands of students in many hundreds of classes worldwide with extremely consistent results for classes at a broad range of institutions with widely varying student demographic characteristics (including pretest scores).[27]

V. FACTORS THAT MAY BE RELATED TO NORMALIZED LEARNING GAIN

An obvious question is, What are the factors that are related to g? Is g related to instructional method, or to individual characteristics of the students and their pre-instruction knowledge state?

Hake's original investigation[20] focused on $\langle g \rangle$ for mechanics courses as determined by pre- and post-testing of the FCI. He distinguished two separate groups of courses: (1) those taught with interactive-engagement (IE) methods, and (2) traditional courses that make little or no use of IE methods. Many studies have been published that broadly confirm Hake's major findings,[27] which are that normalized learning gain $\langle g \rangle$ as measured by the FCI in introductory mechanics courses is (1) largely independent of class mean pretest score; (2) virtually independent of the instructor when traditional instructional methods are used; and (3) tends to be significantly higher (by a factor of about two or more) when IE methods are used in comparison with traditional instructional methods. The issue of what *other* factors may be related to variations in g, besides instructional method, has, with few exceptions, not been addressed.

Another way of investigating the factors that are related to g is to examine the g scores of *individual* students to see if the characteristics of individual students may be related to their own learning gains. Hake *et al.*[9] found indications that students' mathematics skills and spatial visualization abilities might be related to their normalized learning gain, and similar results were reported in Ref. 10. Research on high-school students has led Clement to suggest[14] that reasoning ability may be an independent factor. Preliminary data reported in Ref. 28 strongly suggest that there may be a certain amount of variation in $\langle g \rangle$ that can be ascribed to pretest scores (that is, students' initial degree of physics conceptual knowledge). However, in a separate study,[21] the correlation between $\langle g \rangle$ and pretest scores was very low: $r=-0.06$ on FCI; $r=+0.16$ on FMCE.

The objective of the present study is to aid in building a model of the factors that significantly affect students' learning success in physics. To this end, we examine individual students' normalized learning gain scores using a qualitative test of physics conceptual knowledge; students are tested both before and after instruction. We hope to determine (1) whether individual learning gains are correlated with students' initial level of conceptual knowledge as measured by pretest scores on the same physics concept test, and (2) if those learning gains are correlated with the students' mathematics skills, as determined by pre-instruction testing with a college entrance exam or an algebra/trigonometry skills exam.

VI. CONTEXT OF THIS STUDY

This investigation was carried out in the second semester of a two-semester algebra-based general physics sequence. The data reported here originate in four courses taught by the author: two at Southeastern Louisiana University (SLU) in Fall 1997 and Spring 1998, and two courses taught at Iowa State University (ISU) in Fall 1998 and Fall 1999. The number of students in each course ranged from 65 to 92. The focus of the course was electricity and magnetism, including DC circuits. The SLU course consisted of three 50-minute meetings each week held in the lecture room. (A separate lab course was optional and was not taught by the lecture course instructor; there was no recitation session.) At ISU, in addition to three weekly 50-minute meetings in the lecture room, there is one 50-minute recitation session each week. (There is also a separate required lab in which the lecture instructor has only limited involvement.) These courses made much use of IE instructional methods and employed a variant of Mazur's Peer Instruction.[24,29] The primary curricular material was the *Workbook for Introductory Physics*.[29] Instruction in the recitation sessions at ISU was modeled closely on the University of Washington tutorials,[23] although most of the material used came from the *Workbook for Introductory Physics*.

VII. METHODS

Students' conceptual knowledge was assessed by the administration of a physics concept diagnostic test on the first and last days of class; only students who took both pre- and post-tests are part of the sample. Students' preinstruction mathematics skill was assessed by their score either on the ACT Mathematics Test or on an algebra–trigonometry skills test. A variety of statistical tests were then performed to assess the relation (if any) between students' individual normalized learning gain, and their preinstruction scores on both the physics concept test and the mathematics skills test.

The diagnostic instrument was the Conceptual Survey in Electricity (CSE). This 33-item multiple-choice test surveys knowledge related to electrical fields and forces and the behavior of charged particles. The questions on the CSE are almost entirely qualitative. About half of the items are also included on the Conceptual Survey in Electricity and Magnetism (CSEM).[19] The creators of the CSEM remark that it contains "a combination of questions probing students' alter-

Table I. Correlation between normalized learning gain and pretest score on CSE.

Sample	N	Correlation coefficient between student learning gain g and CSE pretest score	Statistical significance (two-tailed)
SLU 1997	45	+0.15	$p=0.35$ (not significant)
SLU 1998	37	+0.10	$p=0.55$ (not significant)
ISU 1998	59	0.00	$p=0.98$ (not significant)
ISU 1999	78	+0.10	$p=0.39$ (not significant)

native conceptions and questions that are more realistically described as measuring students' knowledge of aspects of the formalism.'' [19]

On the pretest, students were given enough time to respond to all 33 questions. Neither grades nor answers for this pretest were posted or discussed. On the last day of class, the same CSE was administered as an extra-long in-class quiz. However, students were asked to respond to only 23 of the questions.[30] The CSE was used in this abridged form for various reasons. For example, in some cases, the notational conventions differed from what was used in class (for instance, electric field lines are used on the CSE, but only field vectors were used in class). In other cases, the questions involved material that was covered peripherally or not at all in class. Only the 23 designated items were graded, both on the pretest and the post-test. All CSE scores discussed in this paper (as well as quantities derived from them) refer only to the 23-item abridged CSE.

For the SLU samples, scores on the ACT Mathematics Test were used to assess pre-instruction mathematics skill. This test is a college entrance exam, and so there is typically a 1-3 year gap between the time students take this test and the time they take the CSE. The instrument used at ISU is a 38-item multiple-choice test originally developed by Hudson during the course of his investigations (cited in Sec. II) into the effect of mathematics preparation on students' physics performance. It includes the following topics among others: solving and manipulating one- and two-variable algebraic equations; factoring quadratic equations; unit conversions; elementary trigonometry; straight-line graphs; powers-of-10 notation; simple word problems; and addition of numerical and algebraic fractional expressions. (See Appendix for representative problems.)

All students who register for the first semester course in the algebra-based physics sequence at ISU are required to take this test; it does not count toward the students' grade. Because students take this exam at the beginning of the *first* semester course, there was a gap of at least two months (as in the case of summer-school students) between when they took the mathematics test and when they took the CSE. More often, the gap was 5 to 12 months.

Several modifications were introduced during the ISU 1999 course which, it was hoped, would improve instruction. Both graduate student teaching assistants for the course were members of the Physics Education Research Group and had extensive experience and capabilities in inquiry-based instruction. For many of the recitation-session/tutorials, an additional undergraduate teaching assistant was present. During this course, both the teaching assistants and the course instructor spent many out-of-class hours in individual instruction with students who solicited assistance.

VIII. RESULTS

A. CSE pretest scores are not correlated with individual normalized learning gain

Table I shows the correlation coefficients between individual students' g scores and their CSE pretest score for the four samples. The correlations are very small and none is close to being statistically significant. Figure 1 shows the value of g and the CSE pretest score for all students in the ISU 1998 sample. The correlation coefficient for this relation is $r=0.00$; there is no evidence of any pattern in the data points. This random pattern is typical of all four samples.

Table II presents comparisons of $\langle g \rangle$ for several different subgroups of two different samples.[31] For the 1998 sample in Table II, ''Top half'' refers to the students with the 29 highest scores on the CSE pretest; ''Bottom half'' refers to the group with the 30 lowest CSE pretest scores. (The 59-student sample was divided in this way to form two groups of nearly equal size; the groups had zero overlap in pretest scores. Pretest scores ranking #24–29 were identical [eight correct], and scores in the group #30–43 were equal [seven correct].) This method was used to form the other subgroups represented in Tables II and IV.) The mean CSE pretest scores of

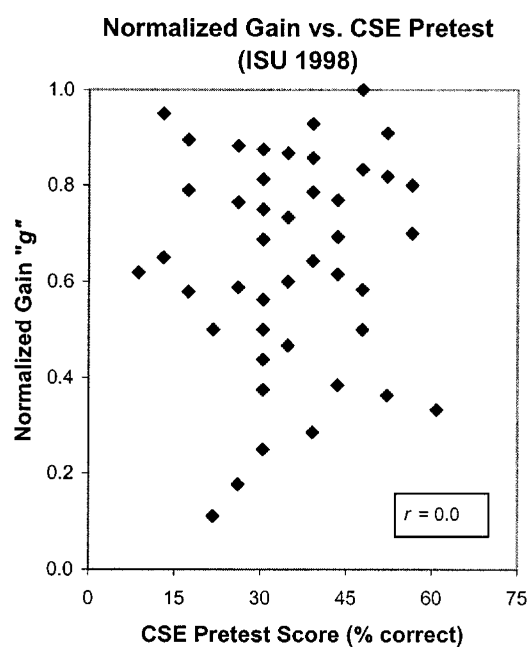


Fig. 1. Scatter plot of ISU 1998 sample; data points correspond to individual students, plotted according to their individual normalized learning gain g score on the Conceptual Survey in Electricity (CSE) and their pretest score on that same exam. Correlation coefficient $r=0.00$.

Table II. ISU samples: Gain comparison, students with high and low CSE pretest scores. $\langle g \rangle$ represents the mean of individual students' normalized gains; s.d.≡standard deviation.

	N	Mean CSE pretest score	$\langle g \rangle$ (s.d.)
1998			
Top half	29	44%	0.68 (0.19)
Bottom half	30	25%	0.63 (0.23)
Top quartile	15	50%	0.65 (0.21)
Bottom quartile	16	20%	0.66 (0.24)
1999			
Top third	30	43%	0.74 (0.18)
Bottom third	27	18%	0.72 (0.17)
Top fifth	14	49%	0.73 (0.20)
Bottom fifth	15	14%	0.67 (0.13)

these two groups were very different, but their normalized gains were not statistically distinguishable according to the one-tailed t-test: $\langle g_{\text{top half}} \rangle = 0.68$, $\langle g_{\text{bottom half}} \rangle = 0.63$, $t = 0.84$, $p = 0.20$. A comparison between even more disparate groups is also shown in Table II. "Top quartile" refers to students with the 15 highest CSE pretest scores in the 1998 sample, while "Bottom quartile" refers to the 16 lowest in that sample. The normalized gains of these two groups were virtually identical. Table II also presents a similar set of comparisons for the ISU 1999 sample. The results for this sample share the main characteristic of the 1998 sample, even for the extreme "Top fifth" and "Bottom fifth" groups: $\langle g_{\text{top fifth}} \rangle = 0.73$, $\langle g_{\text{bottom fifth}} \rangle = 0.67$; these gains are not significantly different according to the one-tailed t-test ($t = 0.98$, $p = 0.17$).

Figure 2 shows the distributions of the normalized gain among the Top half and Bottom half groups from the 1998 sample; there are no striking differences between the pretest groups. A similar result was found for the 1999 sample. This result reinforces the conclusion from the correlation analysis

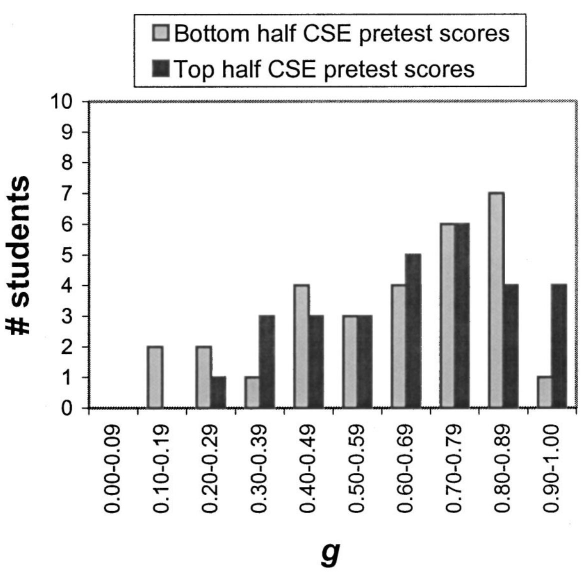


Fig. 2. Distribution of normalized learning gains for ISU 1998 sample: light bars, students with 30 lowest scores on CSE pretest ($\langle g \rangle = 0.63$); dark bars, students with 29 highest scores on CSE pretest ($\langle g \rangle = 0.68$). ($\langle g \rangle$ represents the mean of individual students' normalized gains.)

Table III. Correlation between normalized learning gain and mathematics pretest score.

Sample	N	Correlation coefficient between student learning gain g and mathematics pretest score	Statistical significance (two-tailed)
SLU 1997	45	+0.38	$p < 0.01$
SLU 1998	37	+0.10	$p = 0.55$ (not significant)
ISU 1998	59	+0.46	$p = 0.0002$
ISU 1999	78	+0.30	$p < 0.01$

that the pretest score on the CSE is not a significant factor in determining a student's normalized learning gain.

B. Mathematics pretest scores are correlated with normalized learning gain

Table III presents the correlation coefficient and corresponding statistical significance (that is, p value) for the relation between students' g scores and their scores on the pre-instruction mathematics skills test. The correlation for the SLU 1998 sample was not statistically significant; the correlations for the other three samples were all statistically significant at the $p < 0.01$ level.

Figure 3 shows g as a function of score on the Mathematics Diagnostic Test for the ISU 1998 sample. A positive correlation between the two variables is evident. A similar correlation (though not as large) is also evident in the SLU 1997 and ISU 1999 sample data. Examination of the residuals, that is, the differences between data points and regression fit line, shows that there are no marked nonlinearities evident in the

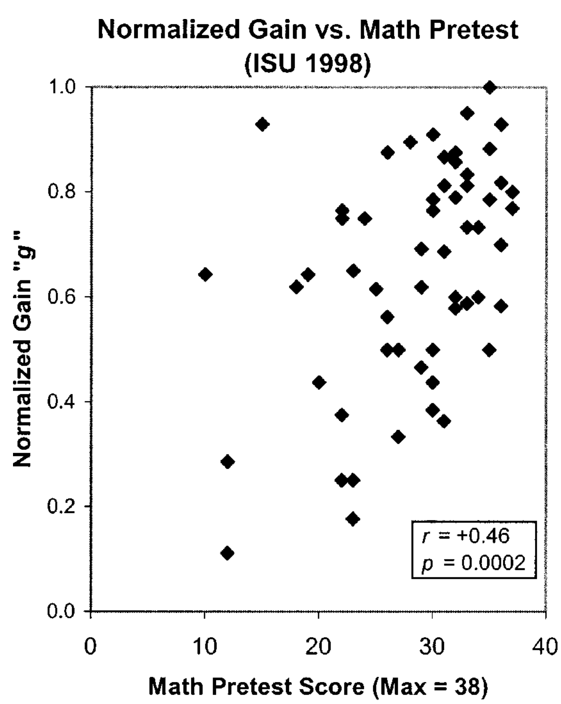


Fig. 3. Scatter plot of ISU 1998 sample. Data points correspond to individual students, plotted according to their individual normalized learning gain g on the CSE and their pre-instruction score on the Mathematics Diagnostic Test. Correlation coefficient $r = +0.46$, $p = 0.0002$; the data are best fit by the linear relation $g = 0.228 + 0.01496\,M$, where M is the number of correct answers on the Mathematics Diagnostic Test (maximum=38).

Table IV. ISU samples: Gain comparison, students with high and low mathematics pretest scores. $\langle g \rangle$ represents the mean of individual students' normalized gains. s.d.≡standard deviation.

	N	Mean mathematics pretest score	$\langle g \rangle$ (s.d.)
1998			
Top half	28	89%	0.75 (0.15)
Bottom half	31	63%	0.56 (0.22)
Top quartile	13	93%	0.77 (0.14)
Bottom quartile	14	49%	0.49 (0.25)
1999			
Top half	37	86%	0.75 (0.20)
Bottom half	36	55%	0.66 (0.22)
Top quartile	21	90%	0.78 (0.17)
Bottom quartile	20	44%	0.60 (0.23)

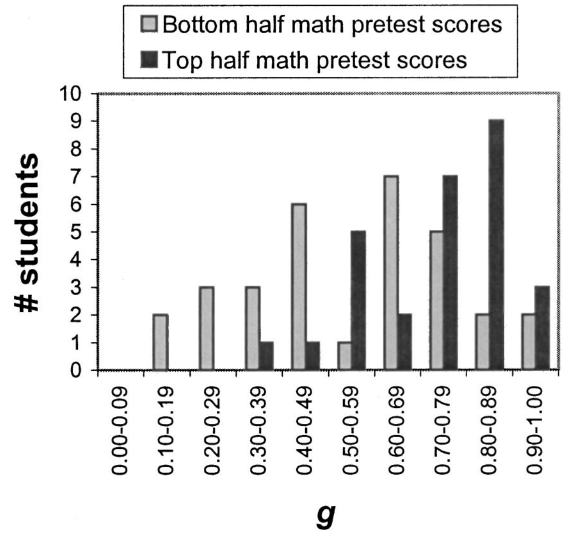


Fig. 4. Distribution of normalized learning gains for ISU 1998 sample: light bars, students with 31 lowest scores on the Mathematics Diagnostic Test ($\langle g \rangle = 0.56$); dark bars, students with 28 highest scores on the Mathematics Diagnostic Test ($\langle g \rangle = 0.75$).

data, and further that the sample variances are fairly uniformly distributed (that is, the data are "homoscedastic").

Table IV presents comparison data for subgroups chosen in a manner analogous to that used in Table II. For instance, the first two lines compare $\langle g \rangle$ for the group of students in the ISU 1998 sample with the highest math pretest scores (Top half, actually the top 47%) to the group with the lowest scores in the same sample (Bottom half, the lowest 53%). In this case—in sharp contrast to the situation in Table II—the learning gains of the two groups are very different, with high statistical significance: $\langle g_{\text{top half}} \rangle = 0.75$, $\langle g_{\text{bottom half}} \rangle = 0.56$; $p = 0.0001$ (one-tailed). When we go to groups even further separated by their mathematics pretest scores—the Top quartile and Bottom quartile groups—we find an even greater difference between their mean normalized gain: $\langle g_{\text{top quartile}} \rangle = 0.77$, $\langle g_{\text{bottom quartile}} \rangle = 0.49$, $p = 0.001$ (one-tailed).

Also shown in Table IV is an analogous set of data for the ISU 1999 sample. The differences in $\langle g \rangle$ between the Top half and Bottom half mathematics pretest groups are substantially smaller than in the 1998 sample, but are still statistically significant: $\langle g_{\text{top half}} \rangle = 0.75$, $\langle g_{\text{bottom half}} \rangle = 0.66$, $p = 0.04$ (one-tailed). Moreover, the difference in learning gain is substantially larger for the groups closer to the extremes of the mathematics pretest score range, that is, the Top quartile and Bottom quartile groups: $\langle g_{\text{top quartile}} \rangle = 0.78$, $\langle g_{\text{bottom quartile}} \rangle = 0.60$, $p = 0.005$ (one-tailed). This difference is consistent with the data from the 1998 sample and significantly strengthens the case that the observed correlation is real and not an artifact produced by the particular selection of the subgroups.

Figure 4 shows the population distributions for the normalized gain for the ISU 1998 sample, portraying the top and bottom mathematics pretest score groups. There is a very noticeable skewing of the distribution toward the high end of the g scale for the high math group. Again, this result is consistent with the correlation analysis and is in striking contrast to the distributions shown in Fig. 2.

It is worth noting another feature of Table IV. Although the normalized gains for the Top half and Top quartile groups in the 1999 sample are nearly identical to those for the corresponding groups in the 1998 sample, that is not the case for the Bottom half and Bottom quartile groups. The g's for those groups are substantially larger in the 1999 sample. It is tempting to ascribe these higher g values to the differences in the instructional methods implemented in 1999, although this is merely speculation.

C. The math score/learning gain correlation is present for both males and females

Table V presents the correlation coefficients and corresponding statistical significance for the male and female subgroups of the two ISU samples (selected because they are larger and contain more reliable data). Although the value of r for males in the ISU 1998 sample is larger than that for females, the difference is not statistically significant ($p = 0.50$, using Fisher transformed values[32]). In the 1999 sample, the correlation coefficients for males and females are nearly identical. All four correlations are statistically significant at the $p < 0.05$ level for a one-tailed test, warranted in this case given the positive correlation observed for both full samples.

IX. DISCUSSION OF RESULTS

The results in this study regarding the *lack* of correlation between normalized learning gain and CSE pretest score are very consistent. However, the results for the mathematics pretest score are in striking contrast to those for the CSE pretest score: in three of the four samples, there is a signifi-

Table V. Correlation between normalized learning gain and mathematics pretest score for males and females (ISU samples).

	N	Correlation coefficient between student learning gain g and mathematics pretest score	Statistical significance (one-tailed test)
ISU 1998: males	22	+0.58	$p < 0.01$
ISU 1998: females	37	+0.44	$p < 0.01$
ISU 1999: males	33	+0.29	$p = 0.04$
ISU 1999: females	45	+0.31	$p = 0.03$

cant positive correlation ($p<0.01$) between normalized learning gain and mathematics pretest score. This relation observed between normalized learning gain and preinstruction mathematics skill is consistent with the preliminary results presented in Refs. 9 and 10; however, the present study represents the first comprehensive examination of this relation.

Another way to look at the data is to compare the mathematics pretest scores for high gainers and low gainers. Hake *et al.*[9] arbitrarily define high and low gainers as those with $g \geqslant 1.3\langle g \rangle$ and $g \leqslant 0.7\langle g \rangle$, respectively, where $\langle g \rangle$ is the mean for the class. They found that high gainers scored 19% higher on the mathematics skills pretest than did the low gainers in their sample. If we apply their definitions and examine mean mathematics pretest scores $\langle m \rangle$ (m is the percentage of correct responses), we find that $\langle m \rangle_{\text{high gainers}} = 81\%$, $\langle m \rangle_{\text{low gainers}} = 60\%$ for ISU 1998 and $\langle m \rangle_{\text{high gainers}} = 80\%$, $\langle m \rangle_{\text{low gainers}} = 65\%$ for ISU 1999. These results are remarkably consistent with those reported in Ref. 9.

The results of Ref. 8 suggested that any observed correlation might *not* be a general characteristic of all students, but of females only. Just as CSE pretest scores were a potentially confounding variable, students' gender has to be considered one as well. With this consideration in mind, the fact that results for *both* ISU samples show statistically indistinguishable correlation coefficients for male and female subpopulations is very significant. Moreover, all four of these correlations were significant at the $p<0.05$ level (one-tailed test) for their individual subpopulation.

The relatively low correlation coefficients found in this study (between $+0.30$ and $+0.46$) yield little predictive power regarding the expected value of the learning gain of an *individual* student, based on his or her pre-instruction score on the mathematics skills test. On the other hand, when assessing the likelihood of a student becoming a high gainer or a low gainer (defined, in this case, as one with gains above or below the class median, respectively), considerably more predictive power is possible. For instance, if we look at the students in the ISU 1998 sample with the lowest mathematics scores (the Bottom quartile in Table IV), we find that only 21% of them (3 of 14) have gains above the class median of $g=0.693$. In comparison, among the group with the highest mathematics scores (Top quartile), 77% (10 of 13) have gains above the class median. Therefore, knowledge of whether a student had unusually high or low mathematics scores could have allowed a fairly high-confidence prediction of whether they would end up with above- or below-average gains.

In striking contrast to this predictability based on mathematics pretest score, the knowledge of a student's CSE pretest score would have allowed no such prediction. The group with the lowest CSE pretest scores (Bottom quartile in Table II) had 50% (8 of 16) with gains above the class median. At the same time, the group with the highest CSE pretest scores (Top quartile in Table II) also had the same number of above-median and below-median gains (7 of each, with one student at exactly the class median).

Higher predictive power is associated with the mean learning gains of the subgroups at the high and low ends of the mathematics scale. The students in the ISU 1998 sample with the lowest mathematics scores have an expected normalized gain (95% confidence interval) ranging from 0.35 to 0.64. In comparison, the expected gain of the group with the highest scores on the mathematics exam range from 0.68 to 0.85. Therefore, we can be highly confident that—for an equivalent sample—the mean gain of the lowest mathematics group would be *below* the class mean of 0.65, while that of the highest mathematics group would be *above* the mean. Obviously, no comparable statement could be made about the groups with the lowest and highest CSE pretest scores. The correlations observed for the other samples are lower, and therefore so is the predictive power, but the same pattern persists.

X. LIMITATIONS OF THIS STUDY

A. Student population

Students enrolled in calculus-based physics courses often have a much more substantial mathematics background than those in the algebra-based course used in this study; this background may be associated with a different relation between mathematics skills and conceptual learning gain in physics. It should also be noted that the population of the two ISU samples was 60% female, a high proportion in comparison to the calculus-based course.

B. Subject matter

Students have considerably less day-to-day experience and accumulated common sense notions regarding electric and magnetic phenomena in comparison with mechanics. Many of the concepts studied (for example, the electromagnetic field) are considerably more abstract than most encountered in the introductory mechanics course. It is conceivable that if a comparable study were done in connection with student learning in a less abstract and more familiar domain, and if assessment relied less on interpretation and analysis of formal representations, the results might be different.

C. Instructional methods

The instructional methods used in this study were certainly not comparable to traditional methods of instruction in widespread national use. They made much use of IE methods, including interactive lecture[29] and group work in the style of the University of Washington tutorials. On the exams, quizzes, and homework, the emphasis was very much on the type of qualitative questions that are used on the Conceptual Survey in Electricity (without teaching to the test). Overall normalized gains were unusually high by national standards. It is possible that the results reported in this study are related in some fashion to the courses' instructional emphasis on qualitative and conceptual problem solving.

D. Hidden variables

It is an inherent limitation of any study that relevant variables might be neglected. For a study such as this one, the particular danger is that some of the neglected variables might actually be so important that their omission is ultimately the source of a spurious apparent correlation that would disappear if these variables had been included. This can happen if the neglected variable is strongly correlated with the targeted dependent variable (learning gain, in this case.)

For example, logical reasoning ability is a variable that some investigators have found to be significant. Suppose that logical reasoning ability is strongly correlated with physics learning gain, and moreover that this reasoning ability is also strongly correlated with pre-instruction mathematics skill. We might find that, for a given level of reasoning ability, there is no separate correlation between mathematics skill and physics learning gain. That would imply that improving reasoning ability might improve learning gain, but that improving mathematics skill would not have such an effect in the absence of any accompanying changes in reasoning ability.

XI. IMPLICATIONS FOR INSTRUCTION

The evidence from this study is that in an IE course, students' normalized learning gains on the CSE are essentially *independent* of their pretest scores. The implication is that, at least with this type of instruction, students' potential to achieve gains in understanding is independent of whether they begin the course with high, low, or even zero initial levels of physics concept knowledge. Knowledge of students' CSE pretest scores might allow some prediction of their probable final level of understanding, but would allow no prediction of their ultimate learning gains. This result is encouraging because it implies that students have an equal chance at learning regardless of their initial knowledge of concepts in electricity.

Although students' initial level of physics concept knowledge may have no impact on their learning gains, the same cannot be said for their initial level of mathematics skill. In three of the four samples in this study, students with higher levels of preinstruction mathematics skill had substantially higher learning gains on the physics concepts—independent of their initial knowledge of those concepts—when compared to students with lower mathematics skill levels (true for both males and females at ISU).

Whether or not this correlation would hold up if other variables, unknown and therefore hidden to us, were included in the analysis is irrelevant to the potential utility of mathematics skill as an indicator of probable high and low gainers. If there are indeed other relevant variables associated with learning gain, it seems likely that they would be correlated with mathematics skill. Until they are known, mathematics skill may be used as a substitute measure for those variables—perhaps not so directly related as those other (hypothetical) variables to the targeted parameter of learning gain, but associated with it nonetheless. (The possibility of using mathematics skill as an indicator of physics learning potential was suggested in Ref. 9 and by many of the investigators cited in Sec. II.) It should be emphasized that the correlation observed between mathematics preparation and normalized learning gain does not imply that mathematics skill is *causally* related to physics concept learning gains. It simply means that whatever factors may ultimately be found to be causally related to learning gain, mathematics skill is probably associated with them in some manner.

In the same sense in which the lack of g versus CSE pretest score correlation was encouraging, the positive correlation between g and a mathematics pretest score is somewhat disconcerting. The implication may be that students with lower levels of preinstruction mathematics skills (whatever the cause) may be unlikely as a group to attain a level of physics learning gain achieved by those with greater mathematics skill, all else being equal. An instructor who transports instructional methods and curricula from one student population to another with much lower mathematics skill levels might find that lower learning gains are achieved. However, the poorer expected outcome of using the *same* instruction with students of lower mathematics skill leaves open the possibility that different instructional methods and curricula might ultimately achieve the same levels of learning gain success with the new population as with the old. The higher learning gains of the low-math group in the ISU 1999 sample (which received modified instruction) might offer some mild support for this speculation.

XII. METHODOLOGICAL IMPLICATIONS

A. The observed correlations might imply that widely diverse populations taught with identical instructional methods might manifest different normalized learning gains

The low-math and high-math subgroups in this study were taught with identical instructional methods (for all practical purposes). And yet it is clear that their mean normalized learning gains were significantly different. If one imagines an entire class populated with low-math students at institution A, and a different class—perhaps at a different institution B—populated with high-math students, it is plausible that instruction carried out with identical methods and materials—perhaps with the identical instructor—might nonetheless result in different values of $\langle g \rangle$ for the two classes.

The extent of the variation in g in a given population that might be ascribed to variations in mathematics preparation would depend on the range of mathematics skills represented in that population; it could be estimated by using the linear regression equation that is a best fit to the g versus M_{pre} data, where M_{pre} is the mathematics pretest score (for example, the data shown in Fig. 3). Using this method, we estimate for the ISU samples that variations in $\langle g \rangle$ ascribable solely to the average variability of students' mathematics preparation (that is, for students having M_{pre} within the range $\langle M_{\mathrm{pre}} \rangle \pm 1.0$ s.d., where s.d. is the standard deviation of the M_{pre} scores) are confined to the range $\langle g \rangle \approx \langle g \rangle_{\mathrm{mean}} \pm 0.15 \langle g \rangle_{\mathrm{mean}}$.

If we speculate that mechanics courses would show correlations between normalized gain and mathematics preparation similar to those in this study, we can estimate that the variation in $\langle g \rangle$ ascribable to mathematics preparation would be ± 0.07 for $\langle g \rangle \approx 0.45$ (a typical value for mechanics courses that employ interactive engagement). This variation is much smaller than the difference commonly found between courses taught with IE and traditional methods, respectively.

B. It may be necessary to consider possible second-order effects due to sample-to-sample differences in preinstruction knowledge state

This particular statement can easily be put in a familiar context. The author measured $\langle g \rangle$ on the CSE to be ≈ 0.48 in his courses at SLU. After attempting to improve his instructional methods and materials, he found $\langle g \rangle \approx 0.67$ in the courses he taught at ISU. (Mean CSE pretest scores were 28% at SLU, 32% at ISU.) Does this difference imply that he succeeded in improving his instruction? Does the large apparent gain in $\langle g\rangle$ perhaps overstate the actual improvement? This type of practical question is one that we often attempt to answer with pre-/post-test data.

If one is actually planning an experiment in which $\langle g\rangle$ is to be a measure of comparative learning gains, it is standard practice to randomize the different samples so that the effects of any potential uncontrolled variables (such as mathematics preparation) may be expected to cancel each other out. One can argue that $\langle g\rangle$ should never be used to compare potentially nonequivalent (that is, nonrandomized) samples. The author's courses at SLU and ISU are a good example of this problem. Should one directly compare the $\langle g\rangle$'s in the two cases, or is some set of hidden variables at work, variables that actually make the two student samples not equivalent?

It is important to emphasize that there is no reason to believe that effects of hidden variables—even combined—are likely to be of the same scale as the two-standard deviation differences in $\langle g\rangle$ on the FCI between traditional instruction and IE instruction documented by Hake. Moreover, with a sample as large as Hake's, it is very unlikely that the IE/non-IE differences in $\langle g\rangle$ could possibly be due to the effects of hidden variables that have not been averaged out. However, when one has much smaller samples in just a few courses taught at widely disparate institutions where the differences in $\langle g\rangle$ may not be so large, there is much more uncertainty in the comparison. To first-order, large differences in $\langle g\rangle$ are probably due to instructional method. However, almost certainly, higher-order effects of unknown scale and origin influence comparative $\langle g\rangle$ statistics in as yet unknown ways.

XIII. SUMMARY

The results of this study provide substantial evidence that factors other than instructional method play a role in determining students' normalized learning gains. Further research to identify and measure these factors should aid in understanding and addressing students' learning difficulties in physics, as well as in analyzing data that result from assessments of student learning.

ACKNOWLEDGMENT

I am very grateful to F. C. Peterson for bringing to my attention the existence of the ISU Mathematics Diagnostic Test data and for many helpful conversations.

APPENDIX

Selected problems from the Mathematics Diagnostic Test used at ISU (author: H. T. Hudson):

(1) $\sqrt{15^2-9^2}=?$

(a) 6, (b) $\sqrt{6}$, (c) 12, (d) $\sqrt{12}$, (e) $\sqrt{135}$.

(2) Find y as a function of x from the following equations: $2x-t=2$, $y-4=3t$,

(a) $y=3x+4$,
(b) $y=10-3x$,
(c) $y=3x+6$,
(d) $y=4-6x$,
(e) $y=6x-2$.

(3) $3/14+7/6=$_____.

(a) 29/21,
(b) 21/20,
(c) 10/21,
(d) 18/49,
(e) 5/21.

(4) If the angle $A=4\pi/6$ radians, what is the value of A in degrees?

(a) 60°, (b) 120°, (c) 90°, (d) 45°, (e) 210°.

(5) $12\times10^8/2\times10^{-2}=$_____.

(a) 6×10^{-4}, (b) 10×10^{10}, (c) 10×10^{-10}, (d) 6×10^{10}, (e) 10×10^6.

[1] Ralph H. Blumenthal, "Multiple instruction and other factors related to achievement in college physics," Sci. Educ. **45**, 336–342 (1961).

[2] H. Daniel Cohen, Donald F. Hillman, and Russell M. Agne, "Cognitive level and college physics achievement," Am. J. Phys. **46**, 1026–1029 (1978).

[3] H. T. Hudson and W. R. McIntire, "Correlation between mathematical skills and success in physics," Am. J. Phys. **45**, 470–471 (1977); H. T. Hudson and Ray M. Rottmann, "Correlation between performance in physics and prior mathematics knowledge," J. Res. Sci. Teach. **18**, 291–294 (1981); Warren Wollman and Frances Lawrenz, "Identifying potential 'dropouts' from college physics classes," *ibid.* **21**, 385–390 (1984); H. T. Hudson, "A comparison of cognitive skills between completes and dropouts in a college physics course," *ibid.* **23**, 41–50 (1986); C. J. Linder and H. T. Hudson, "A comparison of mathematics backgrounds between American and South African physics students," Sci. Educ. **73**, 459–465 (1989).

[4] Audrey B. Champagne, Leopold E. Klopfer, and John H. Anderson, "Factors influencing the learning of classical mechanics," Am. J. Phys. **48**, 1074–1079 (1980); Audrey B. Champagne and Leopold E. Klopfer, "A causal model of students' achievement in a college physics course," J. Res. Sci. Teach. **19**, 299–309 (1982).

[5] H. T. Hudson and Dov Liberman, "The combined effect of mathematics skills and formal operational reasoning on student performance in the general physics course," Am. J. Phys. **50**, 1117–1119 (1982).

[6] W. T. Griffith, "Factors affecting performance in introductory physics courses," Am. J. Phys. **53**, 839–842 (1985).

[7] Ibrahim Abou Halloun and David Hestenes, "The initial knowledge state of college physics students," Am. J. Phys. **53**, 1043–1055 (1985).

[8] Susan McCammon, Jeannie Golden, and Karl K. Wuensch, "Predicting course performance in freshman and sophomore physics courses: women are more predictable than men," J. Res. Sci. Teach. **25**, 501–510 (1988).

[9] R. R. Hake, R. Wakeland, A. Bhattacharrya, and R. Sirochman, "Assessment of individual student performance in an introductory mechanics course," AAPT Announcer **24**(4), 76 (1994).

[10] Michael J. Thoresen and Carl M. Gross, "Factors contributing to success in college physics," AAPT Announcer **30**(4), 128 (2000).

[11] John R. Bolte, "Background factors and success in college physics," J. Res. Sci. Teach. **4**, 74–78 (1966); Gerald E. Hart and Paul D. Cottle, "Academic backgrounds and achievement in college physics," Phys. Teach. **31**, 470–475 (1993); Brian J. Alters, "Counseling physics students: A research basis," *ibid.* **33**, 413–415 (1995).

[12] Philip M. Sadler and Robert H. Tai, "Success in introductory college physics: The role of high school preparation," Sci. Educ. **85**, 111–136 (2001).

[13] George Barnes, "Scores on a Piaget-type questionnaire versus semester grades for lower-division college physics students," Am. J. Phys. **45**, 841–847 (1977); Dov Liberman and H. T. Hudson, "Correlation between logical abilities and success in physics," *ibid.* **47**, 784–786 (1979); Morris A. Enyeart, Dale Baker, and Dave Vanharlingen, "Correlation of inductive and deductive logical reasoning to college physics achievement," J. Res. Sci. Teach. **17**, 263–267 (1980); J. Cilliers, P. Kruger, I. Basson, and P. A. Kirschner, "Psychometric testing as a predictor of student performance in first year physics practicals," J. Sci. Educ. Tech. **5**, 47–57 (1996).

[14] John M. Clement, "Using physics to raise students' thinking skills," AAPT Announcer **30**(4), 81 (2000); "Techniques used to raise thinking skills in high school," *ibid.* **30**(4), 92 (2000); "The correlation between scientific thinking skill and gain in physics conceptual understanding," *ibid.* **31**(2), 82 (2001).

[15] O. J. Ehindero, "Correlates of physics achievement: The role of gender

and non-induced student expectations,'' J. Expt. Educ. **54**, 189–192 (1986).

[16]M. J. Marr, E. W. Thomas, M. R. Benne, A. Thomas, and R. M. Hume, ''Development of instructional systems for teaching an electricity and magnetism course for engineers,'' Am. J. Phys. **67**, 789–802 (1999).

[17]Christoph von Rhöneck, Karl Grob, Gerhard W. Schnaitmann, and Bruno Völker, ''Learning in basic electricity: How do motivation, cognitive and classroom climate factors influence achievement in physics,'' Int. J. Sci. Educ. **20**, 551–565 (1998).

[18]Lillian C. McDermott and Edward F. Redish, ''Resource Letter: PER1: Physics Education Research,'' Am. J. Phys. **67**, 755–767 (1999), and available at ⟨http://www.physics.umd.edu/rgroups/ripe⟩

[19]David P. Maloney, Thomas L. O'Kuma, Curtis J. Hieggelke, and Alan Van Heuvelen, ''Surveying students' conceptual knowledge of electricity and magnetism,'' Phys. Educ. Res., Am. J. Phys. Suppl. **69**, S12–S23 (2001).

[20]Richard R. Hake, ''Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,'' Am. J. Phys. **66**, 64–74 (1998); available at ⟨http://www.physics.indiana.edu/~sdi/⟩

[21]Karen Cummings, Jeffrey Marx, Ronald Thornton, and Dennis Kuhl, ''Evaluating innovations in studio physics,'' Phys. Educ. Res., Am. J. Phys. Suppl. **67**, S38–S44 (1999).

[22]Ronald K. Thornton and David R. Sokoloff, ''Assessing student learning of Newton's laws: the Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula,'' Am. J. Phys. **66**, 338–352 (1998).

[23]Lillian C. McDermott, ''Millikan Lecture 1990: What we teach and what is learned–Closing the gap,'' Am. J. Phys. **59**, 301–315 (1991); Lillian C. McDermott, ''Bridging the gap between teaching and learning: The role of research,'' in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by Edward F. Redish and John S. Rigden [AIP Conf. Proc. **399**, 139–165 (1997)], Pt. 1.

[24]Eric Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997), pp. 3–7, and *passim*.

[25]Walter R. Borg and Meredith D. Gall, *Educational Research, An Introduction* (Longman, New York, 1989), 5th ed., pp. 728–733.

[26]$\langle g \rangle$ is found by using the mean pretest score and mean posttest score of all students in the class. Note that the notation to be adopted in the present paper is unorthodox; see Ref. 31.

[27]Edward F. Redish and Richard N. Steinberg, ''Teaching physics: figuring out what works,'' Phys. Today **52**, 24–30 (1999); Richard R. Hake, ''Lessons from the physics education reform effort,'' Conservation Ecology **5**, 28 (2002), available at ⟨http://www.consecol.org/vol5/iss2/art28⟩

[28]Charles R. Henderson, Kenneth Heller and Patricia Heller, ''Common concerns about the Force Concept Inventory,'' AAPT Announcer **29**(4), 99 (1999), and available at ⟨http://www.physics.umn.edu⟩

[29]David E. Meltzer and Kandiah Manivannan, ''Promoting interactivity in physics lecture classes,'' Phys. Teach. **34**, 72–78 (1996); ''Transforming the lecture-hall environment: The fully interactive physics lecture,'' Am. J. Phys. **70**, 639–654 (2002).

[30]The following item numbers on CSE Form D were used in all cases [more recent versions of the CSE use different numbering; corresponding CSEM item numbers are shown in brackets]: 3, 4, 5, 6, 7, 8, 9[7], 10[8], 11[9], 12, 13[10], 14[11], 15, 16[12], 23[17], 24[18], 25[19], 26, 27[16], 28[20], 31[3], 32[4], 33[5]. In several cases, there are minor differences between the CSE questions and the corresponding CSEM items.

[31]Here the notation $\langle g \rangle$ is used to symbolize the mean value of the *individual* student g's. Its more common use in the literature is to symbolize the g calculated by using mean values of the pretest and post-test scores of all students in a class.

[32]David J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures* (Chapman & Hall/CRC, Boca Raton, FL, 2000), 2nd ed., pp. 782–784.

Issues Related to Data Analysis and Quantitative Methods in PER

David E. Meltzer
Department of Physics & Astronomy, Iowa State University, Ames, IA 50011

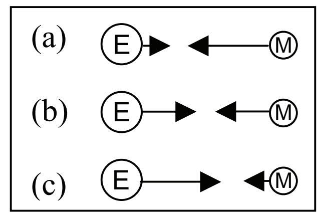
A variety of issues are always relevant (either explicitly or implicitly) in analysis of quantitative data in Physics Education Research. Some specific examples are discussed.

There are a number of issues that always arise, implicitly or explicitly, when conducting quantitative research and carrying out data analysis in Physics Education Research. (Most are relevant for qualitative research as well.)

I. Validity. Broadly speaking, validity refers to the degree to which the conclusions of an investigation truthfully and accurately respond to some specific research questions. Among the particular issues that may arise is: Does your instrument provide data that could actually answer your research question? A common flaw is that the instrument (or test item) is not sufficiently focused, in this sense: To try to answer the question, "Do students understand concept *A*?" the test item (or test instrument) requires knowledge of concepts *A*, *B*, and *C*. Here, *B* and/or *C* might correspond to specific mathematical tools or formal representations. A related question that might arise is: Is your interpretation of the data an accurate representation of students' knowledge?

For example, consider how one might assess students' knowledge of Newton's third law in the context of gravitational forces. At Iowa State I have given a quiz on gravitation on the second day of class for five consecutive years. (The course is the second semester of the algebra-based general physics sequence, focusing on electricity and magnetism. All students in this course have completed their study of mechanics.) Question #1 on the quiz asks whether the magnitude of the gravitational force exerted by the sun on the earth is larger than, the same as, or smaller than the magnitude of force exerted by the earth on the sun. (This question uses words, but no diagrams or equations.) The correct answer ("the same") was given by 10-23% of the students (representing the low and high scores among the five classes). The most popular response by far was "larger," and it was given by 70-83% of all students.

On the very same quiz, Question #8 asks the students to choose a vector diagram that most closely represents the gravitational forces that the earth and moon exert on each other. The three most popular choices are shown in the figure below.



The correct answer "*b*" was given by 6-12% of students. In each of the five independent administrations of the quiz, the proportion of correct responses on Question #8 was about half that on Question #1 (0.43, 0.60, 0.59, 0.50, and 0.50). The implication seems to be that Question #8 was measuring not only students' knowledge of Newton's third law of motion and law of gravitation, but also (in part) students' understanding of vector diagrams. This conclusion is considerably strengthened by the fact that 34-47% of students gave answer "*c*" on Question #8 [answer "*a*": 43-55%]. The "*c*" response corresponds to the force exerted by the more massive object having the *smaller* magnitude, a response that was given by only 3-6% of the same students on Question #1. We see, then, that the validity of two inferences that *might* have been drawn from the results on Question #8 are thrown into question: (1) the proportion of students who misunderstood Newton's third law, and (2) the proportion who believed that in a gravitational interaction involving two masses, the more massive object exerts the *smaller* magnitude force. Although a more definitive analysis of students' reasoning on these questions must await examination of interview data (currently underway), it seems clear that the validity of conclusions that might have been based on only one of these test items would be very uncertain.

This paper appeared in *Proceedings of the 2002 Physics Education Research Conference [Boise, Idaho, 7-8 August 2002]*, edited by Scott Franklin, Karen Cummings, and Jeffrey Marx (PERC Publishing, New York, 2002), pp. 21-24.

The lesson to be drawn from this example is simply the ever-present need to be cautious in collecting and interpreting PER data. Although writers of diagnostic instruments and test items must *always* make some assumptions regarding the previous knowledge of the students being tested, it is important to (1) be aware of what specific assumptions are being made, and (2) have some sound basis (e.g., previous investigation) for believing that the assumptions are accurate.

Another threat to validity of interpretations of test data is associated with analysis of students' answers without regard for explanations of their reasoning. Although there are many good practical reasons for employing diagnostic instruments that yield "answer only" data without students' explanations, it is important for researchers to be aware of possible pitfalls in the data analysis. These dangers are associated most particularly with attempts to draw conclusions from only one or a small number of test items. For example, in a study at the University of Washington [1], students were asked to compare the changes in kinetic energy and momentum of two objects of different mass, acted upon by the same force. For both of these comparisons, the proportion of correct responses observed when ignoring students' explanations was substantially higher than when answers were judged correct only when accompanied by a correct explanation. (KE comparison: 45-65% correct vs. 30-35% correct; momentum comparison: 55-80% correct vs. 45-50% correct.) Many other researchers have reported anecdotal evidence that supports the conclusion suggested by this study, that is, that data regarding students' explanations of their reasoning (whether in written or verbal form) very substantially strengthen the potential validity of conclusions drawn from any given investigation.

II. Reliability. Reliability refers to the consistency of results produced by a specific instrument or investigative protocol. It is related to validity in the sense that an *unreliable* instrument is very unlikely to lead to valid conclusions about a research question. Reliability encompasses several distinct concepts: (1) Is the instrument internally consistent, that is, do different components of the instrument measure (more or less) the same property? This may be investigated with such measures as KR-20 or Cronbach's alpha [2]. Note that an instrument might well be designed that *intentionally* measures two or more distinct conceptual areas and therefore might not be expected to yield similar results on different subsections. (2) If you made the same measurement again (with all conditions apparently identical), would your instruments yield the same result? If a particular test item or a small set of items deal with a concept of which students have little or no knowledge, responses tend to be random. Therefore, even two consecutive administrations of the same instrument might yield substantially different results and analysis should take that into consideration. (3) Would minor variations in your test items (e.g., slight contextual or representational changes, or alterations in question format) lead to large variations in results?

For example: Schecker and Gerdes [3] reported significant differences in student responses to certain FCI questions when the questions were posed in slightly different physical contexts, i.e., a soccer ball instead of a golf ball, or a vertical pistol shot instead of a steel ball thrown upward. Steinberg and Sabella [4] administered final-exam problems in free-response format that were similar to several FCI questions. They found that in some cases, there were significant differences in percent correct responses between the final-exam questions and corresponding FCI items (administered post-instruction) for students who took both tests. In the example discussed in Section I above, two very similar questions on gravitation posed in different representational forms yielded significantly different results, suggesting that the reliability of an instrument that depended on only one or the other type of question might be compromised. With regard to multiple-choice exams, Rebello and Zollman [5] have provided evidence that even well-validated multiple-choice questions might miss categories of responses that students would offer were the questions posed in free-response format. They also show that in some cases, the specific selection of distracters provided to students can significantly affect the proportion of correct responses.

Again, it should be emphasized that researchers are always forced to make some assumptions regarding the reliability of their instruments and methods. Nonetheless, some efforts – however informal – should be made to gauge the reliability of any particular investigative protocol.

More generally, diagnostic items that omit students' explanations may have their reliability threatened for that reason alone. In the University of Washington study discussed above [1], both questions (i.e., KE comparison and momentum comparison) were posed in two separate variants: one in which the different objects experienced forces for the same time period, and one in which the time periods differed. Remarkably, the proportion of correct responses when explanations were required was nearly identical for the two variants (KE: 35% and 30%; momentum: 50% and 45%). However, when explanations were ignored, results on the two variants were significantly different (KE: 65% and 45%; momentum: 80% and 55%). This suggests that reliability, and not merely validity, may be strongly dependent on consideration of student explanation data.

III. Statistical Significance. Before drawing any conclusions from one's data it might be helpful to ask whether there is a substantial probability (10% or more) that your result might have occurred purely by chance. Do you have a measure of variance, or can one be estimated? If standard deviations are available a *t*-test (or similar measures) could be used to assess significance of differences in sample means. If not, an assumption of binomial distribution might be made and a test for difference between binomial proportions could be applied [6].

If many individual variables or inter-sample differences are being tested for significance, then substantial deviations from "null hypothesis" values may be expected to occur, purely by chance, for some tested items. For instance, if 100 different sample means are compared, random fluctuations would dictate that several are likely to show a two-sigma (p = 0.05) effect (i.e., means separated by two or more standard errors).

Another important consideration is that the sample size being utilized may be inadequate to yield a statistically significant result for the specific effect being investigated. In that case, failure to observe a difference between control and experimental groups may *not* imply non-existence of a treatment effect, but merely that the sample size used or the experimental protocol employed is inadequate to demonstrate the existence of the effect at an acceptable level of statistical significance.

IV. Pedagogical Significance. Is the observed effect likely to be of practical significance in the classroom? Are there cost-benefit relationships implied in the magnitude of the effect [7]? Even if an effect is statistically significant (e.g., large "effect size" [8]) the actual learning gains (as measured for instance by Hake's g [8, 9]) might be small and of limited practical pedagogical interest.

V. Representativeness of Sample. Is your student sample representative of the larger group from which it is (implicitly or explicitly) drawn? Are samples from the different student groups that are being compared equivalent in all respects except for the variable being investigated? If sample selection is truly random the expectation is that the answer to both of these questions should be "yes." In random samples that are sufficiently large, the probability that both answers actually *are* "yes" is very high. However, samples are rarely "sufficiently large" nor, for that matter, truly randomly selected. In that case one must consider which relevant population variables may differ among the various student samples, for example: demographic makeup, previous preparation, pre-instruction knowledge, etc. Although some measures of learning gain such as Hake's g explicitly incorporate normalization to reduce the dependence on pretest scores [8, 9], so-called "hidden variables" such as mathematics preparation, gender, spatial visualization ability, reasoning ability, etc. may nonetheless exert an influence for which account should be taken [8, 10]. Even more subtle variables such as whether students are enrolled in an "on-sequence" or "off-sequence" course might have an effect [11].

One should always ask: How have you controlled variables that might be relevant? Have you done random selection? If not, what alternatives were used? In any case, what is the basis for believing that the different population samples being compared are equivalent except for the treatment being tested?

VI. Reproducibility. Just because you saw an effect in one PER experiment does not necessarily mean you will observe it again. In physics, all groups of electrons in identical states are completely equivalent. In PER, different groups of students are never in identical states and are never truly completely equivalent. This reality requires answers to questions such as these: Did you repeat the experiment? Did anybody else repeat the experiment? Are your results substantially different from what others have observed, or are they otherwise very surprising? If so, better check again!

It is important to keep in mind that PER necessarily deals with many variables that are often difficult (and sometimes impossible) either to identify or to control (or both), e.g.: student demographics, instructor style, course logistics, issues of validity and reliability of diagnostic instruments, etc. Moreover, students' mental models of physics concepts are often complex and incorporate overlapping and frequently conflicting themes. Therefore, students' responses to different (though related) questions may be highly variable. Largely due to this assortment of variables, fluctuations from one PER data run to the next tend to be large (and, of course, each data run may require an entire academic quarter or semester). This inherently large scale of fluctuations substantially increases the importance of replication in PER investigations in comparison, for instance, to more traditional physics research. Even investigations that yield large treatment effects with high statistical significance should probably be replicated by the original research group at the same institution, and/or by other researchers working at different institutions with diverse student populations.

SUMMARY

Although the issues that are discussed here often get no explicit attention in Physics Education Research papers and presentations, I believe that PER investigators should formulate responses – at least implicitly and approximately – to all questions of this type. Substantial neglect of one or more of these issues can threaten the validity and usefulness of the results of an investigation, and vitiate the product of hundreds of hours of laborious study.

I am grateful for discussions with Leith Allen and R. Hake. This material is based upon work supported by the National Science Foundation under Grant Number DUE-9981140 and Grant Number REC-0206683.

REFERENCES

1. T. O'Brien Pride, S. Vokos, and L. C. McDermott, "The challenge of matching learning assessments to teaching goals: An example from the work-energy and impulse-momentum theorems." Am. J. Phys. **66**, 147 (1998).

2. W. R. Borg and M. D. Gall, *Educational Research*, *An Introduction* (Longman, New York, 1989), 5th ed., pp. 260-261.

3. H. Schecker and J. Gerdes, "Messung von Konzeptualisierungsfähigkeit in der Mechanik: Zur Aussagekraft des FCI," Zeitschrift für Didaktik der Naturwissenschaften **5** (1), 75-89 (1999).

4. R. Steinberg and M. Sabella, "Performance on multiple-choice diagnostics and complementary exam problems," Phys. Teach. **35**, 150 (1997).

5. N. S. Rebello and D. A. Zollman, "The effect of distracters on student performance on the Force Concept Inventory," Am. J. Phys. (Phys. Educ. Res. section) *in press*.

6. J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 2nd ed. (Brooks/Cole, Monterrey, CA, 1987), Ch. 9.2, 9.4.

7. R. R. Hake, "Effect sizes and economic cost-benefit," AERA-D post at <http://lists.asu.edu/cgi-bin/wa?A2=ind0207&L=aera-d&F=&S=&P=3921>.

8. R. R. Hake, "Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization," <http://www.physics.indiana.edu/~hake/index.html>.

9. R. R. Hake, "Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," Am. J. Phys. **66,** 64-74 (1998); <http://www.physics.indiana.edu/~sdi/>.

10. D. E. Meltzer, "The relationship between mathematics preparation and conceptual learning gains in physics: A possible 'hidden variable' in diagnostic pretest scores," Am. J. Phys. **70**, 1259-1268 (2002) and at: <http://www.physics.iastate.edu/per/index.html>.

11. N-L Nguyen and D. E. Meltzer, "Initial understanding of vector concepts among students in introductory physics courses," Am. J. Phys. (Phys. Educ. Res. section), accepted for publication, <http://www.physics.iastate.edu/per/index.html>.

The Questions We Ask and Why: Methodological Orientation in Physics Education Research

David E. Meltzer
Department of Physics & Astronomy, Iowa State University, Ames, IA 50011

Research methodology is discussed using a simple model of students' knowledge. I argue that the nature of data obtained is closely linked to the type of knowledge being probed.

Objectives of Physics Education Research

The research methodology one employs will necessarily depend on one's particular objectives. Our group's objective is to find ways to help students learn physics more effectively and efficiently, to understand concepts more deeply. To do this we seek to understand the process by which students develop their physics knowledge, and what difficulties they encounter along the way.

A Model for Students' Knowledge Structure

To model students' knowledge, Redish uses the analogy of an archery target [1]. The central black bull's-eye represents what the students know well. It contains a tightly linked, hierarchically structured network of concepts understood in depth. When problems related to knowledge in that region are posed to the students, they answer rapidly, confidently, consistently, and correctly, independent of context or representational mode.

The gray circle surrounding the bull's-eye represents what students understand partially and imperfectly. Some concepts are understood well and some not so well; some firmly held beliefs in this region are inconsistent with physical reality. Some links between concepts are strong, but most are weak, absent, or miswired from the standpoint of an expert's knowledge. Knowledge in this region is dynamic and still in the process of development. When questions from this region are put to students they may answer correctly in some contexts, yet incorrectly or incompletely in others.

The outer white region represents what students don't know at all. It contains disconnected fragments of concepts, poorly understood terms and equations, and few or no links relating one fragment to another. Questions from this region yield responses that are mostly noise: highly context-dependent, inconsistent and unreliable, with deeply flawed or totally incorrect reasoning.

Redish, following Vygotsky – who called the gray region the "zone of proximal development" – says that teaching is most effective when targeted at concepts in the gray. ("The zone of proximal development defines those functions that have not yet matured but are in the process of maturation, functions that will mature tomorrow but are currently in an embryonic state [2].") This region is analogous to a substance near a phase transition: a few key concepts and a handful of crucial links can catalyze substantial leaps in student understanding. Conversely, in the bull's-eye region one is merely refining a well-established body of knowledge, while instruction targeted at the white region yields only infrequent and poorly retained gains, lacking stability and durability.

Probing Students' Knowledge

When we administer diagnostics or carry out interviews in which students' bull's-eye regions are probed, we get consistent, reliable, and rather uninteresting results. When we probe understanding in the white region we get inconsistent, context-dependent responses, also uninteresting from a research or teaching standpoint. In contrast, when we probe the gray area, we tend to get rich, diverse, and potentially interesting and useful data.

Sometimes we find relatively stable, internally consistent conceptual islands which may, or may not, be consistent with physicists' knowledge. These islands are likely to have flawed or broken links to the bull's-eye region. When persistent patterns with well-defined characteristics are found, we identify and analyze them. By necessity, we are probing students' responsiveness to minimal guidance, since even asking a question is a form of guidance. In physics terminology we are trying to determine the student's "response function."

We attempt to map a student's knowledge structure in the gray region, and then amalgamate

CP720, *2003 Physics Education Research Conference*, edited by J. Marx, S. Franklin, and K. Cummings

a set of such individual mappings into an ensemble average. We determine the population *average* of things such as typical reasoning patterns, stability of links, responsiveness to probes, etc. We also gauge the magnitude of the natural "line width" to the distributions, that is, the spread around the mean value of the measured parameters [1].

Applying the Model: Sample Research Design

Our group has recently investigated student learning in thermodynamics. A short written diagnostic was administered to several hundred students in three separate offerings of the calculus-based general physics course, and 32 students from a fourth offering of the course were interviewed.

Analysis of the written responses had indicated several surprising results, including a widely prevalent belief that heat and work behaved as state functions, and a very weak understanding of the first law of thermodynamics [3]. The recently published paper by Loverude, Kautz, and Heron [4] had documented very similar difficulties. These results guided our objectives for the interviews; to focus on "gray region" knowledge:

- pose elementary ***baseline questions*** to determine "lower" bounds on understanding;
- use a pictorial representation of a cyclic process to present ***diverse real-world contexts*** in order to probe students' ideas in depth throughout the gray region;
- gauge resilience and ***stability of students' concepts*** upon minimal probing;
- identify key ***learning difficulties***, and gauge their approximate prevalence.

By contrast, there were several alternative research objectives on which we did *not* focus:

- exactly how students had acquired their knowledge *[would be a* ***very*** *difficult task*];
- students' attitudes towards learning *[separate investigation; not our primary interest]*

Although these are limitations on the completeness of our picture of students' thinking, any investigation must be constrained in *some* manner.

Learning Difficulties, Not Alternative Theories

Even alternative conceptions that are clearly and confidently expressed are unlikely to be defended with the strength of a full-blown "theory." Different contexts or representations, or questions using related concepts, may trigger dormant links and influence students to reconsider their reasoning.

For example, in the thermodynamics interviews a lengthy description of a cyclic process was given, with diagrams portraying varying positions of a piston as a volume of ideal gas was alternately expanded and compressed back to its original state. Students were asked this question:

> Consider ***<u>the entire process</u>*** from time *A* to time *D*. Is the net work done *by* the gas on the environment during that process (a) greater than zero, (b) equal to zero, or (c) less than zero?

A *P-V* diagram of the process referred to in the question (*not* shown to the students) is given in Fig. 1. The magnitude of the net work done by the system is represented by the enclosed area, and since the path is traversed counterclockwise the net work done is negative.

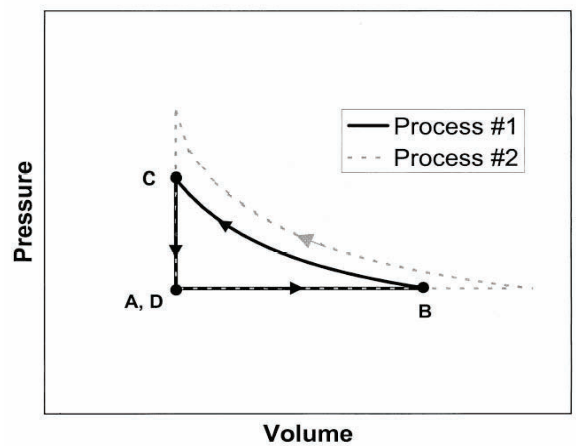


Figure 1. A *P-V* diagram (not shown to students) of the process (Process #1) discussed during interviews.

Most students (over two thirds) quickly and confidently answered that the net work done would be equal to zero. Their explanations expressed just a few common themes, typified by these two:

> "The net work done by the gas…I put equal to zero. I was measuring work as the force over a certain distance, and if your piston is back to your original spot you had a positive work, and you had a negative work. And if you all measured it from the same starting point, you're back to the original point with the same thing. So, you're equal to zero. There *was* work done by the gas on the

environment, but the *total* work over the entire process is equal to zero."

"I think the net work is zero, because no change in volume…Because work is equal to the integral of $P\Delta V$…and $\Delta V = 0$."

Variations of these arguments were readily volunteered and persistently defended by most of the students. However, 17% of those who initially answered "zero" changed their response after they were asked to draw a *P-V* diagram of the process. Some changed to "greater than zero," and some to the correct response. For these students, drawing the diagram triggered a recollection of the relationship between work done and area under the curve. Their original belief – despite being confidently expressed and defended with a plausible physical argument – was not so stable as to resist a counter-argument spontaneously arising from the students themselves with only a minimal external influence. Thus we found that an apparently strong student conception was at least somewhat unstable when confronted with alternative reasoning.

Although this zero-net-work idea reflects a serious misunderstanding of work in a thermodynamic context, there is no basis for ascribing to it attributes of a full-blown alternative theory. There is no reason to think that students had this conception pre-formulated in any consciously articulated form before they were interviewed. They seemed to be offering explanations that had been worked out on the spot, although most of them obtained the same answer and defended it with similar reasoning. However, their explanations lacked the depth that would be expected from a carefully thought-out physical model.

The precise origin of this student idea – how it abruptly crystallized based on previous instruction and experience – is an open question. It is based to some extent on the common-sense notion that properties of a system returned to its original state must have undergone no net change. However, this line of reasoning also includes specific physical arguments based on students' prior knowledge of physics, including overgeneralizations of both net mechanical work done by conservative forces, and of net changes in state functions during a cyclic process. Those arguments would need to be addressed before students could thoroughly resolve their understanding of these concepts. It is quite possible that this conception, however lacking in the attributes of a full-blown alternative theory, may be quite resistant to instruction.

Investigating Stability of a Learning Difficulty

Through research I try to map out conceptions related to learning difficulties, and to understand what systems or situations elicit them with greatest consistency. Some of these conceptions may be pre-existing in students' minds before their first physics class, but more often they are only vaguely and incompletely expressed until encountered in an instructional setting. There, however, one often finds that they arise with monotonous regularity. An example is students' idea that heat is or behaves as a state function.

We asked students to compare the heat absorbed by the same system in two different processes represented on a *P-V* diagram, both processes sharing the same initial and final states. It was clear from the diagram that the work done was different in the two processes, and so the heat absorbed also had to be different [3]. However, 39% of the students asserted that the heat absorbed by the system would be equal for both processes. Many offered explicit arguments regarding the path-independence of heat, for example: "I believe that heat transfer is like energy in the fact that it is a state function and doesn't matter the path since they end at the same point," "they both end up at the same *PV* value so…they both have the same Q or heat transfer." Students offered similar arguments to explain – in response to an interview question – why they believed a system undergoing a cyclic process would receive zero total heat transfer. Thus the belief that heat is or behaves as a state function proved sufficiently persuasive that students' responses in two very different contexts were extremely consistent with each other.

A remarkable aspect of our findings was the popularity of *explicit* statements to the effect that heat was "a state function," "doesn't depend on path," or "depends only on initial and final states." Well over 100 students volunteered statements of this type (either in written responses or during interviews), notwithstanding the virtual certainty that they had never read them in any textbook nor heard them from any instructor [5]. They were synthesized by students on their own, and with startling regularity.

It seems that students have some useful intuitions regarding state functions that they improperly generalize (perhaps unconsciously) to the cases of both heat and work. It would be worthwhile to investigate in more detail just how and why this overgeneralization occurs during the instructional process. However, there is great value simply in knowing that it does tend to occur, in knowing the approximate frequency of its occurrence in a given population, and in knowing the form that students' explanations tend to follow.

Interpretation of Students' Reasoning

When we report the results of research, we do not confine ourselves to a bare statistical summary of the data. We offer qualitative assessments based on an overview of all data sources. In particular, we must determine how *consistent* are the various assessments of student thinking. Are the results qualitatively and quantitatively in agreement with each other? Do students offer the same or similar answers when repeatedly probed with related questions? How confident are they in their responses?

Do students offer numerous lines of unproductive reasoning, or do they gravitate toward just one or two? Are there common themes in students' thinking that are not directly reflected in the tabulated data, or in the selected quotations? Do the data and quotations as presented fairly represent the stability and consistency of students' thinking? I believe that researchers should make clear their answers to these questions based on an overall assessment of their data.

Conclusion

The fundamental challenge of research into student understanding is that we are investigating a moving target. Students are always learning, and their mental states are always undergoing change. It is precisely these *changes* – in response to instructional interventions – that are our primary interest. One might well find that two students, whose *instantaneous* mental states (and ability to answer questions) appear to be identical, are actually following very different learning trajectories, with different learning rates.

All assessments – particularly interviews – probe students' thinking not at a single moment, but over a period of time. Students often alter their initial responses under the most minimal probing. The dynamic nature of any assessment raises profound issues of how to view the student's knowledge at one moment in time from the perspective of the learning trajectory (rate and direction) along which they are moving.

Recognition of the fluid nature of assessment has motivated development of the field of *Dynamic Assessment*, documented in many books and journal articles over the past two decades [6]. Practitioners of Dynamic Assessment – explicitly motivated by Vygotskian thinking – have developed assessment protocols that gauge student responsiveness to short-term instructional interventions. These methodologies hold promise for application within physics education research.

The underlying theme of this methodology is that we are probing student thinking that is truly in a state of flux and development, such that conceptual understanding is constantly undergoing evolution and restructuring. The aim of research is not to portray a misleading picture of firmly rooted student concepts, but to provide a snapshot of the interplay and evolution of student thinking – to gauge which aspects are more clearly defined and persistent, and which are relatively flexible and fluid. The more accurately and thoroughly we accomplish that, the better we will be able to develop improved curricula and instructional methods.

This material is based upon work supported by the National Science Foundation under Grant Numbers DUE-9981140 and REC-0206683.

REFERENCES

1. E. F. Redish, "Implications of cognitive studies for teaching physics," Am. J. Phys. **62**, 796 (1994).
2. L. S. Vygotsky, *Mind in Society* (Harvard University Press, Cambridge, MA, 1978), p. 86.
3. D. E. Meltzer, "Student reasoning regarding work, heat, and the first law of thermodynamics in an introductory physics course," *PERC Proceedings 2001* (S. Franklin, J. Marx, and K. Cummings, eds.), pp. 107-110.
4. M. E. Loverude, C. H. Kautz, & P. R. L. Heron, "Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas," Am. J. Phys. **70**, 137 (2002).
5. Aside from confusion regarding $Q = mc\Delta T$!
6. E.g., Carol S. Lidz, *Practitioner's Guide to Dynamic Assessment* (Guilford, New York, 1991).

How Do You Hit A Moving Target? Addressing The Dynamics Of Students' Thinking

David E. Meltzer

Department of Physics and Astronomy
Iowa State University

Abstract. From the standpoint both of research and instruction, the variable and dynamic nature of students' thought processes poses a significant challenge to PER. It is difficult merely to assess and characterize the diverse phases of students' thinking as they gain and express understanding of a concept. (We might call this the "kinematics" of students' thought processes.) Much harder still is uncovering the various factors (instructional method, student characteristics, etc.) that influence and determine the trajectory of students' thinking. (We could call this the "dynamics" of students' thinking.) The task of deciphering the mutual interaction of these factors adds to the challenge. I will outline some of the initial work that has been done along these lines by various researchers, and I will identify some directions for future research that I think might be fruitful for workers in PER.

INTRODUCTION

Our goal as educators is to better understand the *process* of student learning so as to be able to influence it more effectively. Students' learning of physics is characterized by a knowledge state that is a generally increasing function of time. Often, however, the inherent time-dependence of this process is given inadequate examination, in part due to the difficulty of investigating students' thinking at multiple time points during its evolution.

Characterization of a time-dependent process requires a bare minimum of two probes at different time points, while a varying rate requires three such probes. Alternatively, a probe may be carried out over a continuous (brief) time interval and variations during that interval observed. (This type of probe is characteristic of so-called "dynamic assessment" [1] and the "teaching experiment" [2].) In any case, such repeated probes of student thinking are logistically difficult to implement within actual classroom settings involving ongoing instruction.

In this paper I will outline some of the work that has been done by various researchers in exploring changes in student thinking over time, and I will identify some directions for future research that I think might be fruitful for workers in PER.

ASSESSING STUDENTS' MENTAL STATES AT A PARTICULAR TIME

It is useful to recall the complexity of a thorough probe of students' thinking at even a single point in time. Such a probe would require analysis not only of a students' ideas about a set of physics concepts and the relationships among them, but also of the ways in which the student perceives and implements the learning process itself.

Students' "Knowledge" State

At any given moment a student has a collection of ideas related to specific physics concepts, and a related set of ideas corresponding to the expressed or implied interconnections among those concepts. These ideas are in significant part dependent on context, that is, they often depend on the physical setting of a given problem, the form of representation employed in the problem, and so forth. One can try to assess this

Published in *2004 Physics Education Research Conference [Sacramento, California, 4-5 August 2004]*, edited by Jeffrey Marx, Paula R. L. Heron, and Scott Franklin [American Institute of Physics Conference Proceedings **790**, 7-10 (2005)].

collection of student ideas by posing questions involving diverse contexts and a variety of representations [3-9]. In this fashion one can try to determine the "distribution function" of ideas (sometimes called the "mental model" [4,5]) characteristic of a particular student, or of a particular student population.

Students' "Learning State"

Another key component of students' thinking is the set of their ideas related to the practice of learning physics, along with the methods they actually employ to learn. This includes their study methods, their attitudes toward physics and physics learning, their motivation to learn, etc. One can attempt to assess these factors through a number of methods including observations of learning practices [10], attitudinal surveys [11,12], "dynamic assessment" [1], "teaching experiments" [2], etc.

CHARACTERIZING THE PROCESS OF STUDENT LEARNING

If we are to carry assessment beyond a single time point, we must determine the specific parameters needed for an assessment of the overall learning process. If we can obtain observable data corresponding to those parameters, we then need to determine how exactly to analyze those data.

Qualitative Parameters

The basic elements of a time-dependent analysis of student learning include *sequences* of the various parameters that characterize students' knowledge. These include the following: (1) The *sequence of ideas* and of *sets* of ideas (mental models) developed by a student during the process of learning a set of related concepts; (2) The *sequence of difficulties* encountered by a student during that learning process (difficulties are related to "*ideas*," but are not necessarily the same thing); (3) The *sequence of knowledge resources and study methods* employed by the student during that process; (4) The *sequence of attitudes* developed by a student during that process.

The fundamental assumption in this analysis is that all of the various elements may (and probably do) undergo change over time. There will always be a question of how rapidly this change occurs and, consequently, how frequently an assessment must be made in order not to overlook key stages of the process.

Quantitative Parameters

In addition to qualitative parameters, one can identify a number of potentially relevant measures to which numbers can be attached. These include the following: (1) The progression in *depth of knowledge* as measured by probability of correct response on a set of related questions (e.g., score S, range [0.00,1.00]); (2) The *average rate of learning* R of a set of related concepts (e.g., $R = g/\Delta t$ where g = normalized gain calculated using $S_{pretest}$ and $S_{posttest}$); (3) The *variations in the learning rate* V encountered by a student during that process (e.g., $V = \Delta R/\Delta t$); (4) The *time-dependent distribution function* characterizing the idea set of a student population. (This might be defined through a method analogous to that of Bao [4,5].)

Phase I: "Kinematics" Of Students' Thinking

The first level of investigation is to characterize the pattern of students' thinking as it *evolves* during the learning process. In principle the objective is to determine, at a number of different points in time, the set of students' ideas, difficulties, learning resources, etc. with respect to a well-defined concept or set of related concepts. (For instance, one might acquire data related to students' understanding of Newton's second law of motion.) Then, based on this time-series data, one can try to determine the *normal course of evolution* of those ideas and difficulties under a variety of standard learning situations.

Phase II: "Dynamics" Of Students' Thinking

The second phase of the investigation would be to determine the factors that *influence* the evolutionary pattern of students' thinking during the learning process. One might describe this objective as an attempt to answer the question, "What are the social and pedagogical forces that determine the path of a student's 'learning trajectory'?" More specifically, one could ask: What is the *relative* influence of (a) individual student characteristics (preparation, background, etc.) and (b) instructional method (including pedagogical techniques, classroom environment, etc.), on the observed sequences of ideas, difficulties, and attitudes? A crucial question would be to determine the extent to which the observed sequences might be altered due to efforts of the instructor and/or the students.

PREVIOUS WORK

A number of workers have investigated various aspects of the issues discussed in this paper. However, many related issues have been explored little or not at all. Here I will outline some of this previous work.

Sequence of ideas: A number of investigators have described shifts in mental models by analyzing the differences in typical student response patterns between pretests and posttests [4,5,7-9]. Savinainen et al. have also explored such patterns at mid-instruction points (between pre- and post-instruction) [8,9], while other workers have attempted to describe and characterize the sequence of ideas acquired by students during the learning process in a more detailed, step-by-step fashion [13-15]. Some workers (e.g., Thornton [3] and Dysktra [6]) have postulated the existence of specific "transitional states," which are well-defined sets of ideas occurring during the transition from novice to expert thinking.

Sequence of difficulties: The generalizability of *patterns* of learning difficulties is well established [16], but that of difficulty *sequences* has not been thoroughly investigated. In general, there has not been much detailed exploration into how the specific learning difficulties students encounter may change and evolve over the course of a semester or year.

Sequence of attitudes: There is evidence of regularities in attitude change during instruction [11], but also evidence that these regularities are dependent upon instructional context [12].

DYNAMIC ASSESSMENT

As an alternative to assessment of student thinking at a single instant (through a quiz, exam, etc.), a preplanned sequence of questions, hints, and answers may be provided and the students' responses observed throughout a time interval. This method has been formalized under the rubric "Dynamic Assessment" [1]. One first attempts to determine what types of problems the students can solve on their own, without additional assistance. One then continues by providing carefully measured and sequenced assistance through hints and answers, in order to assess the students' ability to respond to instructional cues with efficient learning. Among the assessment criteria are the amount of assistance required, the rapidity and depth of response, etc. A similar method is the "teaching experiment" [2], in which a mock instructional setting is used as a means to probe students' responses to various instructional interventions.

QUESTIONS FOR FUTURE WORK

Here I will list a number of questions that might serve as a basis for future investigations on these topics. For convenience, I will divide them according to whether they refer primarily to *characterizations* of the evolutionary process of students' thinking ("kinematics"), or to the factors that *influence* that process ("dynamics").

I. Kinematics

(1) Can one confirm the existence of *well-defined* "transitional mental states" related to learning of specific concepts, that is, sets of ideas concerning those concepts that are intermediate between those of a novice and those of an expert? If such transitional states do exist, do they vary among individuals according to differences in their background and preparation? Are different transitional states observed in traditional and reformed instruction?

(2) More broadly, one can ask: Does the individual "mental model" distribution function evolve according to some characteristic pattern? (This "distribution function" refers to the collection of student concepts related to a specific topic, as reflected for instance in the set of responses to a group of related diagnostic questions [4,5].) Is the evolution pattern correlated with individual characteristics (demographics, preparation, etc.) and/or with the nature of the instructional method?

(3) How does the population "mental model" distribution function evolve in general? (Here we refer to the average set of responses given by an entire class of students, or a number of similar classes.) Is the evolution pattern correlated with population demographics?

(4) Are there common patterns of variation in learning rates? For example, do learning rates typically increase or decrease monotonically throughout the course of a semester?

(5) Is the magnitude of the learning rate at an early phase of the process correlated with the long-term learning rate [17]?

(6) Is the picture of a student's learning trajectory provided by "dynamic assessment" (or teaching experiments) over a brief time interval more complete and accurate than that provided by a single standard quiz or exam?

II. Dynamics

(1) Can one trace back, in a causal fashion, the set of student ideas at a particular time, to the specific set of ideas and difficulties that had been acquired at an earlier time? More specifically: To what extent does the student's present set of ideas and difficulties determine the pattern of his or her thinking in the future?

(2) Are transitional states (if they exist) actually influenced by differences in students' preparation, and/or by the nature of the instructional method?

(3) Are the sequences of individual and population "idea distribution functions" (mental models) influenced by individual background and/or instructional mode?

(4) Are learning-rate variations influenced by individual background and/or instructional mode? More broadly, what are the factors that influence the trajectory of student learning, and what is the nature of the interaction among the various determining factors?

SUMMARY

The dynamic, time-dependent aspects of the student learning process are essential features of that process, and yet they are logistically difficult to observe and analyze. Future investigations in this area have the potential to yield valuable information that could help instructors increase the effectiveness of instruction in physics.

ACKNOWLEDGMENTS

The work of the Iowa State Physics Education Research Group has been supported by grants from the Iowa State University Center for Teaching Excellence, and from the National Science Foundation through the Division of Undergraduate Education, Division of Research, Evaluation and Communication, and Division of Physics.

REFERENCES

1. Lidz, C. S., *Practitioner's Guide to Dynamic Assessment*, New York: Guilford, 1991.
2. Engelhardt, P. V., Rebello, N. S., Corpuz, E., and Ozimek, D., "The Teaching Experiment: What It Is and What It Isn't," in *Proceedings of the 2003 Physics Education Research Conference*, edited by J. Marx et al., AIP Conference Proceedings, New York: American Institute of Physics, *in press*.
3. Thornton, R. K., "Conceptual Dynamics: Following Changing Student Views of Force and Motion," in *The Changing Role of Physics Departments in Modern Universities: Proceedings of International Conference on Undergraduate Physics Education*, edited by E. F. Redish and J. S. Rigden, AIP Conference Proceedings 399, New York: American Institute of Physics, 1997, pp. 241-266.
4. Bao, L. and Redish, E. F., *Am. J. Phys*. **69**, S45-S53 (2001).
5. Bao, L., Hogg, K., and Zollman, D., *Am. J. Phys*. **70**, 766-778 (2002).
6. Dykstra, D. I., "Why Teach Kinematics?" Parts I and II, preprint (2002).
7. Meltzer, D. E., *AAPT Announcer* **33 (4)**, 98 (2003).
8. Savinainen, A., *High School Students' Conceptual Coherence of Qualitative Knowledge in the Case of the Force Concept*, Ph.D. Dissertation, Department of Physics, Joensuu (Finland): University of Joensuu, 2004.
9. Savinainen, A., Scott, P., and Viiri, J., *Sci. Educ*. (*in press*).
10. Thornton, R. K., "Uncommon Knowledge: Student Behavior Correlated to Conceptual Learning," preprint (2004).
11. Redish, E. F., Saul, J. M., and Steinberg, R. N., *Am. J. Phys*. **66**, 212-224 (1998).
12. Elby, A., *Am. J. Phys*. **69**, S54-S64 (2001).
13. Hrepic, Z., Zollman, D. A., and Rebello, N. S., *AAPT Announcer* **33 (4)**, 134 (2003).
14. Itza-Ortiz, S. F., Rebello, N. S., and Zollman, D., *Euro. J. Phys*. **25**, 81-89 (2004).
15. There are also a number of "case studies" in the literature in which very detailed, diary-like expositions follow small numbers of students on almost a week-to-week basis. It is hard to know whether and to what extent such descriptions might be generalizable to broader student populations.
16. For example, see the many studies referenced in McDermott, L. C., and Redish, E. F., *Am. J. Phys*. **67**, 755-767 (1999).
17. Meltzer, D. E., *AAPT Announcer* **27 (4)**, 89 (1997).

VII.

PER in Overview

Mini-Course on Physics Education Research and Research-Based Innovations in Physics Instruction

David E. Meltzer
Department of Physics and Astronomy
Iowa State University
Ames, Iowa, U.S.A.

Supported by the U.S. National Science Foundation, Division of Undergraduate Education, and Division of Research, Evaluation, and Communication

I. Physics Education Research: Laying the Basis for Improved Physics Instruction

Over the past 20 years, systematic investigations have helped to clarify the dynamics of students' thinking during the process of learning physics. This research has revealed students' learning difficulties, as well as aiding in the development of more effective instructional strategies. I will describe the principal goals and methods of Physics Education Research, and discuss some of the methodological issues related to this work. With examples drawn from investigations we have carried out at Iowa State University, I will illustrate this research process and show how it can lead to improved curricula and instructional methods.

Within the past 20 years, physicists have begun to treat the teaching and learning of physics as a research problem. This includes (1) systematic observation and data collection, and carrying out of reproducible experiments, (2) identification and control of variables, and (3) in-depth probing and analysis of students' thinking. This field of study has come to be known as "Physics Education Research" (PER). Broadly speaking, the goals of PER are to improve the effectiveness and efficiency of physics instruction. This is carried out primarily by developing and assessing instructional methods and materials that address obstacles which impede students' learning of physics. The methods of PER include the development and testing of diagnostic instruments that assess student understanding, and the utilization of these instruments to investigate student learning. Students' thinking is probed through analysis of written and verbal explanations of their reasoning, supplemented by multiple-choice diagnostics. Learning is assessed through measures derived from pre- and post-instruction testing.

It is important to realize that there are certain things PER can *not* do: PER can not determine an instructor's "philosophical" approach toward education, such as whether one should focus on improving the achievement of the majority of enrolled students, or instead focus on a subgroup, such as high-ability or low-ability students. PER can not specify the goals of instruction in particular learning environments, such as the appropriate balance between learning of "concepts," and development of mathematical problem-solving skills. PER may help instructors make informed choices about these goals, but it can not determine what they should be.

There are now more than 60 PER groups in U.S. physics departments, including more than 30 in Ph.D.-granting departments. The primary activities of PER groups include (1) research into student learning, (2) research-based curriculum development, (3) assessment of instructional methods, and (4) preparation of K-12 physics and science teachers. Curriculum development is directed both at introductory and advanced courses, lab- and non-lab courses, and courses for teacher preparation. There are many different research themes, including investigations of students' conceptual understanding, development and assessment of diagnostic instruments, students' attitudes and beliefs about learning physics, and many others.

Among the specific issues addressed by PER are these: many (if not most) students (1) develop weak qualitative understanding of physics concepts after standard introductory courses, and (2) lack a "functional" understanding of concepts that would allow them to solve problems in unfamiliar contexts. There are many reasons for this. For one, students hold (or develop during instruction) many firm ideas about the physical world that may conflict with physicists' views. (Examples: an object in motion must be experiencing a force; a given battery always produces the same current in any electric circuit.) Beyond that, most introductory students need a great deal of guidance in developing scientific reasoning skills and using abstract concepts. Most of these students lack "active-learning" skills that would permit more efficient mastery of physics concepts.

One of the ways that PER researchers address these problems is through research-based curriculum development. This involves investigation of student learning with standard instruction, with a focus on probing learning difficulties encountered by students during this instruction. Based on this research, new curricular materials are developed, tested, and modified. Student understanding is assessed to determine whether the new materials actually result in improved learning. I will discuss a simple example of how this process is carried out by outlining some of the work done at Iowa State University to investigate student learning of concepts in gravitation. I will also briefly sketch out another project related to student learning of thermodynamics, and in my next presentation I will describe that project in detail.

In addressing the issues involved in curriculum development, it is useful to remember that at least *some* students learn efficiently. Highly successful physics students are "active learners": they continuously probe their own understanding by posing their own questions, scrutinizing implicit assumptions, examining varied contexts, etc. By contrast, most introductory students are unable to do efficient active learning on their own. They don't know "what questions they need to ask," and they require considerable assistance by instructors using appropriate curricular materials.

To help students become active learners, several principles can be used as a guide: (1) students are led to engage in deeply thought-provoking activities during class time ["interactive engagement"]; (2) students' preexisting "alternative conceptions" and other common learning difficulties are recognized and deliberately elicited; (3) the process of science (exploration and discovery) is used as a means for learning science; students are not necessarily "told" things are true; instead, they are prodded to figure them out for themselves as much as possible ("inquiry-based" learning). The term "Interactive Engagement" [originated by R. Hake] usually implies very high levels of interaction between students and instructor, collaborative group work among students during class time, and intensive active participation by students in learning activities during class time.

Some strategies used to elicit students' preconceptions and learning difficulties include: (1) having students make predictions of the outcome of experiments; (2) requiring students to give written explanations of their reasoning; and (3) posing specific problems that are known to consistently trigger certain learning difficulties. Incorporating inquiry-based learning can be done by giving students an opportunity to investigate or think about concepts before the instructor actually discusses the concept in detail. This may be done either by leading students to draw conclusions based on evidence they acquire in the instructional laboratory, or – in lecture courses – by guiding students through chains of reasoning using printed worksheets. Research-based instruction emphasizes qualitative, non-numerical questions to reduce students' unthinking reliance on algebraic "plug-and-chug." Extensive use is made of multiple representations (graphs, diagrams, computer simulations, verbal descriptions, etc.) and diverse physical contexts in order to deepen students' understanding. Requiring students to explain their reasoning (verbally or in writing) helps them to more clearly expose their thought processes.

I will describe some of the research that has been done on improving students' problem-solving abilities, and I will outline some instructional strategies that have been developed based on that research (e.g., use of multiple representations by Alan Van Heuvelen, and "Context-Rich Problems" by Pat and Ken Heller). I will also outline some instructional strategies using active-learning laboratories ("Workshop Physics" by Laws *et al.*; "Socratic-Dialogue-Inducing Labs" by R. Hake), and active-learning textbooks (*Matter and Interactions* by Chabay and Sherwood; *Understanding Basic Mechanics* by Reif; *Physics: A Strategic Approach* by Knight). Perhaps the oldest and most thoroughly tested instructional approach is that developed at the University of Washington by Lillian C. McDermott and her co-workers. Their method (sometimes known as "Elicit, Confront, Resolve") has led to the development of the widely used research-based curricular materials *Physics by Inquiry* and *Tutorials in Introductory Physics*. Implementing active-learning instructional strategies in large lecture classes is a particular challenge; I will discuss that subject in detail during my third presentation.

Finally, I will discuss some methodological issues involved in PER. A key question for teachers is how to assess the effectiveness of instruction. A single exam measures only a students' instantaneous knowledge state, but instructors are interested in *learning*, i.e., the transition between states. For that, one needs a measure of learning gain that has maximum dependence on instruction, and minimum dependence on students' pre-instruction state. A widely used measure that addresses these needs is Hakes' "normalized gain" or *g*, defined as the learning gain (pre-instruction to post-instruction), divided by the maximum possible gain. I will discuss some of the properties of normalized gain, and some of the issues that are involved in making use of it.

II. Developing Improved Curricula and Instructional Methods based on Physics Education Research

In many research-based curricula, physics students are guided to work their way through carefully designed and tested sequences of questions, exercises, and/or laboratory activities. Utilizing these materials, and interacting frequently during class with instructors and with each other, students have often achieved significant gains in understanding when compared with instruction based on lecture alone. In this presentation I will describe in some detail the process of developing these research-based curricula, as carried out by our group at Iowa State over the past several years. I will show how our research into students' reasoning in thermodynamics is helping guide the development of improved curricular materials. Similarly, investigations of the pedagogical role played by diverse representational modes (mathematical, verbal, diagrammatic, etc.) are also helping us lay the basis for developing more effective instructional methods.

In this presentation I will describe in considerable detail some of the investigations we have carried out regarding student learning of specific topics in physics, and how we have begun to use the results of that research to develop improved instructional materials.

In collaboration with Prof. Tom Greenbowe of the Iowa State Chemistry Education Research Group, we initiated a project to develop improved curricular materials for teaching thermodynamics. To lay the basis for that work, we carried out extensive investigations of student learning in courses using standard instruction. Here I'll discuss an investigation of reasoning regarding heat, work, and the first law of thermodynamics among students in an introductory calculus-based general physics course. We found that responses to written questions by 653 students in three separate courses were very consistent with results of detailed individual interviews carried out with 32 students in a fourth course. Although most students seemed to acquire a reasonable grasp of the state-function concept, it was found that there was a widespread and persistent tendency to improperly over-generalize this concept to apply to both work and heat. A large majority of interviewed students thought that net work done and/or net heat absorbed by a system undergoing a cyclic process must be zero, while only 20% or fewer were able to make effective use of the first law of thermodynamics even after instruction was completed. Students' difficulties seemed to stem in part from the fact that heat, work, and internal energy all share the same units. Results were consistent with those of previously published studies of students in U.S. and European universities, but portray a pervasiveness of confusion regarding process-dependent quantities that was previously unreported. The implication is that significant enhancements of current standard instruction may be required for students to master basic thermodynamic concepts.

Loverude, Kautz, and Heron (University of Washington) have pointed out that a crucial first step to improving student learning of thermodynamics concepts lies in solidifying the student's understanding of the concept of work in the more familiar context of mechanics, with particular attention to the distinction between positive and negative work [*Am. J. Phys*. **70**, 137 (2002)]. Beyond that first step, it seems clear that little progress can be made without first guiding the student to a clear understanding (1) that work in the thermodynamic sense can alter the internal energy of a system, and (2) that "heat" or "heat transfer" in the context of thermodynamics refers to a *change* in some system's internal energy, or equivalently that it represents a quantity of energy that is being *transported* from one system to another.

I will describe some of our initial efforts to develop improved curricular materials and instructional methods for these topics. We are planning to extend this work to more advanced topics, including student learning of statistical physics.

In a related investigation we have explored students' approaches to solving calorimetry problems involving two substances with differing specific heats. We found that students often employ various context-dependent rules-of-thumb such as "*equal energy transfer implies equal temperature change*," and "*temperature changes are directly proportional to specific heat*." Through interviews we found that students frequently get confused by, or tend to overlook, the detailed proportional reasoning or algebraic procedures that could lead to correct solutions. Instead, they often proceed with semi-intuitive reasoning that at times may be productive, but more often leads to inconsistencies and non-uniform conceptual understanding. We have developed new curricular materials that are designed to address these and related learning difficulties. I will illustrate and discuss some of these materials, and describe some of the preliminary testing we have carried out.

Another project done in collaboration with Tom Greenbowe is an investigation of the role played by diverse representational modes in the learning of physics and chemistry. There are two major phases of this work: (1) Probe students' reasoning with widely used representations, such as free-body diagrams, *P-V* diagrams, vector diagrams of various types, etc., and (2) compare student reasoning with different forms of representation of the same concept (verbal, diagrammatic, mathematical, graphical, etc.). In an initial phase of this work with graduate student Ngoc-Loan Nguyen, we investigated the understanding of vector concepts in graphical form among students enrolled in general physics courses at Iowa State University. We found a number of significant learning difficulties related to addition of vectors and ability to manipulate vectors without a coordinate system or grid. Many students had an imprecise understanding of vector direction and a vague notion of vector addition.

In further investigations, we compared students' ability to solve similar (or identical) problems when presented using different forms of representation. We used a "multi-representation quiz" in which a single problem is presented in several different versions, utilizing either words only ("verbal" version), mathematical symbols, graphs, or diagrams. We found significant differences in student performance on some questions, in particular verbal and diagrammatic questions involving Newton's third law. The proportion of students making errors when responding to the diagrammatic version of the questions was consistently higher than in the case of the verbal version. Moreover, many students had difficulty in translating certain phrases such as "exerted on" or "exerted by" into vector-diagram form, and this led to other discrepancies between responses in the two cases. We also found some preliminary evidence that there might be differences between the performance of males and females on electrical circuit-diagram questions: the error rate for females was about 50% greater than that of males, even after identical instruction.

III. Research-Based Active-Learning Instructional Methods in Large-Enrollment Physics Classes

A long-standing challenge has been to incorporate active-learning instructional methods in large-enrollment physics classes traditionally taught in a lecture format. I will describe the methods we have introduced to develop a "fully interactive physics lecture," and discuss the curricular materials that we have created to support this form of instruction. This involves both carefully designed sequences of multiple-choice conceptual questions, and free-response worksheets designed to be used by students working in collaborative groups.

SEE SLIDES BEGINNING NEXT PAGE

Research-Based Active-Learning Instructional Methods in Large-Enrollment Physics Classes

David E. Meltzer

Department of Physics and Astronomy
Iowa State University
Ames, Iowa
U.S.A.

Supported by NSF DUE #0243258 and DUE #0311450

Research in physics education and other scientific and technical fields suggests that:

- "Teaching by telling" has only limited effectiveness
  - can inform students of isolated bits of factual knowledge
- For understanding of
  - inter-relationships of diverse phenomena
  - deep theoretical explanation of concepts

→ **students have to "figure it out for them-selves" by struggling intensely with ideas**

Research in physics education and other scientific and technical fields suggests that:

- "Teaching by telling" has only limited effectiveness
  - listening and note-taking have relatively little impact
- Problem-solving activities with rapid feedback yield improved learning gains
  - student group work
  - frequent question-and-answer exchanges with instructor

Goal: Guide students to "figure things out for themselves" as much as possible

What Role for Instructors?

- Introductory students often don't know what questions they need to ask
  - or what lines of thinking may be most productive
- Instructor's role becomes that of guiding students to ask and answer useful questions

What needs to go on in class?

- Clear and organized presentation by instructor is not *at all* sufficient
- Must find ways to guide students to synthesize concepts in their own minds
- Instructor's role becomes that of guiding students to ask and answer useful questions
  - aid students to work their way through complex chains of thought

Keystones of Innovative Pedagogy

- problem-solving activities during class time
- deliberately elicit and address common learning difficulties
- guide students to "figure things out for themselves" as much as possible

The Biggest Challenge: Large Lecture Classes

- Very difficult to sustain active learning in large classroom environments
- Two-way communication between students and instructor becomes paramount obstacle
- Curriculum development must be matched to innovative instructional methods

Active Learning in Large Physics Classes

- **De-emphasis of lecturing**; Instead, ask students to respond to many questions.
- Use of classroom communication systems to obtain **instantaneous feedback** from entire class.
- Cooperative **group work** using carefully structured free-response worksheets

Goal: *Transform large-class learning environment into "office" learning environment (i.e., instructor + one or two students)*

"Fully Interactive" Physics Lecture

DEM and K. Manivannan, Am. J. Phys. ***70,*** *639 (2002)*

- Very high levels of student-student and student-instructor interaction
- Simulate one-on-one dialogue of instructor's office
- Use numerous structured question sequences, focused on specific concept: small conceptual "step size"
- Use student response system to obtain instantaneous responses from all students simultaneously (e.g., "flash cards")

Sequence of Activities

- Very brief introductory lectures (≈10 minutes)
- Students work through sequence of multiple-choice questions, signal responses using flash cards
- Some "lecture" time used for group work on worksheets
- Recitations run as "tutorials": students use worksheets with instructor guidance
- Homework assigned out of workbook

Features of the Interactive Lecture

- High frequency of questioning
- Must often create unscripted questions
- Easy questions used to maintain flow
- Many question variants are possible
- Instructor must be prepared to use diverse questioning strategies

Video (18 minutes)

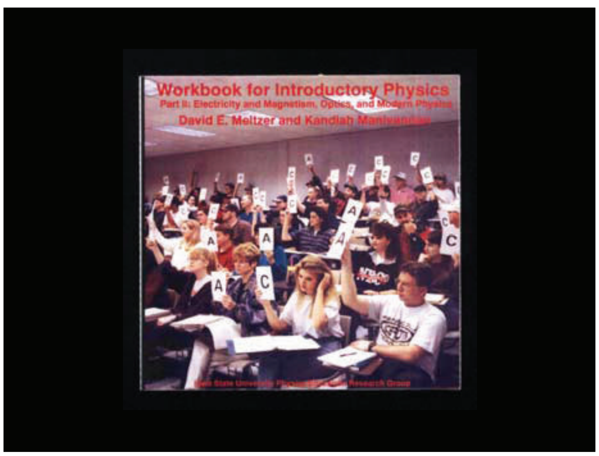
- Excerpt from class taught at Southeastern Louisiana University in 1997
- Algebra-based general physics course
- *First Part:* Students respond to questions written on blackboard.
- *Second Part:* Students respond to questions printed in their workbook.

Curriculum Requirements for Fully Interactive Lecture

- Many question sequences employing multiple representations, covering full range of topics
- Free-response worksheets adaptable for use in lecture hall
- Text reference ("Lecture Notes") with strong focus on conceptual and qualitative questions

→ *Workbook for Introductory Physics (DEM and K. Manivannan, CD-ROM, 2002)*

Supported by NSF under "Assessment of Student Achievement" program



Curriculum Development on the Fast Track

- Need curricular materials for complete course ⇒ must create, test, and revise "on the fly"
- Daily feedback through in-class use aids assessment
- Pre- and post-testing with standardized diagnostics helps monitor progress

Curricular Material for Large Classes
"Workbook for Introductory Physics"

- Multiple-choice "Flash-Card" Questions
  - Conceptual questions for whole-class interaction
- Worksheets for Student Group Work
  - Sequenced sets of questions requiring written explanations
- Lecture Notes
  - Expository text for reference
- Quizzes and Exams
  - some with worked-out solutions

High frequency of questioning

- Time per question can be as little as 15 seconds, as much as several minutes.
  - similar to rhythm of one-on-one tutoring
- Maintain small conceptual "step size" between questions for high-precision feedback on student understanding.

Must often create unscripted questions

- Not possible to pre-determine all possible discussion paths
- Knowledge of probable conceptual sticking points is important
- Make use of standard question variants
- Write question and answer options on board *(but can delay writing answers, give time for thought)*

Easy questions used to maintain flow

- Easy questions (> 90% correct responses) build confidence and encourage student participation.
- If discussion bogs down due to confusion, can jump start with easier questions.
- Goal is to maintain continuous and productive discussion with and among students.

Many question variants are possible

- Minor alterations to question can generate provocative change in context.
  - add/subtract/change system elements (force, resistance, etc.)
- Use standard questioning paradigms:
  - greater than, less than, equal to
  - increase, decrease, remain the same
  - left, right, up, down, in, out

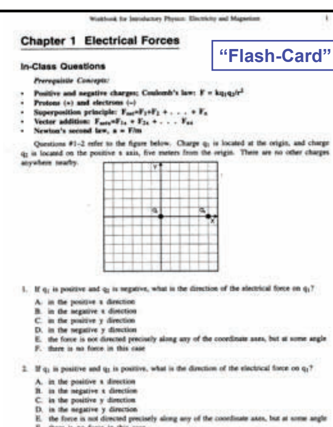
Instructor must be prepared to use diverse questioning strategies

- If discussion dead-ends due to student confusion, might need to backtrack to material already covered.
- If one questioning sequence is not successful, an alternate sequence may be helpful.
- Instructor can solicit suggested answers from students and build discussion on those.

Interactive Question Sequence

- Set of closely related questions addressing diverse aspects of single concept
- Progression from easy to hard questions
- Use multiple representations (diagrams, words, equations, graphs, etc.)
- Emphasis on qualitative, not quantitative questions, to reduce “equation-matching” behavior and promote deeper thinking

“Flash-Card” Questions


Workbook for Introductory Physics: Electricity and Magnetism 1

Chapter 1 Electrical Forces

In-Class Questions

Prerequisite Concepts:

- Positive and negative charges; Coulomb's law: $F = kq_1q_2/r^2$
- Protons (+) and electrons (–)
- Superposition principle: $F_{net} = F_1 + F_2 + \ldots + F_n$
- Vector addition: $F_{net} = F_{1x} + F_{2x} + \ldots F_{nx}$
- Newton's second law, $a = F/m$

Questions #1–2 refer to the figure below. Charge q_1 is located at the origin, and charge q_2 is located on the positive x axis, five meters from the origin. There are no other charges anywhere nearby.

1. If q_1 is positive and q_2 is negative, what is the direction of the electrical force on q_1?
   A. in the positive x direction
   B. in the negative x direction
   C. in the positive y direction
   D. in the negative y direction
   E. the force is not directed precisely along any of the coordinate axes, but at some angle
   F. there is no force in this case

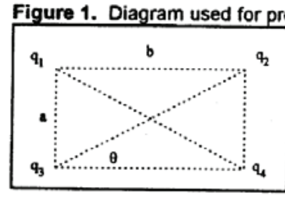
2. If q_1 is positive and q_2 is positive, what is the direction of the electrical force on q_1?
   A. in the positive x direction
   B. in the negative x direction
   C. in the positive y direction
   D. in the negative y direction
   E. the force is not directed precisely along any of the coordinate axes, but at some angle
   F. there is no force in this case

Problem “Dissection” Technique

- Decompose complicated problem into conceptual elements
- Work through problem step by step, with continual feedback from and interaction with the students
- May be applied to both qualitative and quantitative problems

Example: Electrostatic Forces

Figure 1. Diagram used for problem dissection

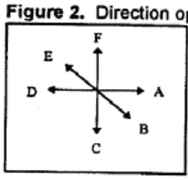


Four charges are arranged on a rectangle as shown in Fig. 1. ($q_1 = q_3 = +10.0\ \mu C$ and $q_2 = q_4 = -15.0\ \mu C$; a = 30 cm and b = 40 cm.) Find the magnitude and direction of the resultant electrostatic force on q_1.

Question #1: How many forces (due to electrical interactions) are acting on charge q_1?

(A) 0 (B) 1 (C) 2 (D) 3 (E) 4 (F) Not sure/don’t know

Figure 2. Direction options



For questions #2-4 refer to Fig. 2 and pick a direction from the choices A, B, C, D, E, and F.

Question #2: Direction of force on q_1 due to q_2
Question #3: Direction of force on q_1 due to q_3
Question #4: Direction of force on q_1 due to q_4

Assessment Data

Scores on Conceptual Survey of Electricity and Magnetism, 14-item electricity subset

Sample	*N*	Mean pre-test score	Mean post-test score	<g>
National sample (algebra-based)	402	27%	43%	0.22
National sample (calculus-based)	1496	37%	51%	0.22
ISU 1998	70	30%	75%	0.64
ISU 1999	87	26%	79%	0.71
ISU 2000	66	29%	79%	0.70

Quantitative Problem Solving: Are skills being sacrificed?

ISU Physics 112 compared to ISU Physics 221 (calculus-based), numerical final exam questions on electricity

	N	Mean Score
Physics 221: F97 & F98 ***Six final exam questions***	*320*	56%
Physics 112: F98 **Six final exam questions**	*76*	77%
Physics 221: F97 & F98 **Subset of three questions**	*372*	59%
Physics 112: F98, F99, F00 **Subset of three questions**	*241*	78%

Summary

- Focus on ***what the students are doing in class,*** not on what the instructor is doing
- Guide students to answer questions and solve problems during class
- Maximize interaction between students and instructor ***(use communication system)*** and among students themselves ***(use group work)***

GUEST EDITORIAL

The future of physics education research: Intellectual challenges and practical concerns

During the World Year of Physics, much effort is being made to celebrate the unprecedented advances in our understanding of the physical world made during the past century. However, we have not yet seen comparable advances in our understanding of student learning of our discipline. One possible explanation is that learning is inherently more complex than most physical processes. Although this explanation is plausible, we have not made similar systematic efforts to understand student learning. The enormous effort expended by many physics instructors over the past century was not harnessed in a way that made cumulative progress likely. As Lillian McDermott has observed, ‘‘Unless we are willing to apply the same rigorous standards of scholarship to issues related to learning and teaching that we regularly apply in more traditional research, the present situation in physics education is unlikely to change.’’[1]

In the past few decades, an increasing number of physicists have taken up this challenge by applying methods of research based on those that have been employed successfully in investigations of the physical world. This endeavor is broadly known as ‘‘physics education research’’ (PER). Systematic studies of student learning have revealed a wide gap between the objectives of most physics instructors engaged in traditional forms of instruction and the actual level of conceptual understanding attained by most of their students.[2] But PER has gone beyond documenting shortcomings in student learning and traditional instruction. Researchers have developed instructional materials and methods that have been subjected to repeated testing, evaluation, and redesign. Numerous reports have documented significant and reproducible learning gains from the use of these materials and methods in courses ranging from large-enrollment classes at major public universities to small classes in two-year colleges and high schools.[1–3] Still, there remain inadequacies in even the most recent instructional approaches and many unanswered questions. In this Guest Editorial we will identify some of the current and emerging research directions that we consider promising. We also argue for the importance of doing research on the learning and teaching of physics in physics departments. We do not mean to suggest that PER should not be conducted in schools of education, but, as we argue later, we do not believe that the field is viable without a critical mass of faculty in physics departments. Finally, we identify some practical and political challenges and propose some steps that could be taken to help ensure the stability, growth, and productivity of PER.

Current and future research directions. We first briefly mention some of the research directions that have potential for deepening our understanding of how students learn physics. This understanding should lead to more effective instructional tools, techniques, and materials. We highlight those directions that address intellectual issues that are specific, but not necessarily unique, to the subject matter and reasoning patterns of physics. Therefore we omit important work on investigating gender-equity issues, for example. Moreover, we focus on the college and university level, although some issues we mention have implications for K-12 instruction. We do not wish to neglect the large and vigorous PER community outside the U.S. However, although many fundamental issues of student learning are largely invariant across cultures, the diversity of approaches to education and, consequently, of research goals is too broad to be addressed satisfactorily here.

Most early PER work focused on student ability to apply the concepts covered in typical introductory university physics courses. The results of these studies have proven invaluable in guiding improvements in instruction. The breadth of topics covered, their importance as a foundation for future study, and the many students involved ensure that the introductory course will continue to be a major emphasis for the foreseeable future. Current research efforts range from extensions of earlier studies of student ability to interpret and apply kinematical concepts[4] to investigations of student understanding of basic electromagnetism and modern physics.

In recent years, there has been an increasing focus on student learning in upper-level courses such as quantum mechanics,[5] thermal physics,[6] relativity,[7] and advanced mechanics.[8] This research should lead to learning gains for physics majors similar to those found for research-based instruction at the introductory level.

We also expect to see a greater emphasis on tracing students’ intellectual development as they progress through the undergraduate curriculum, both in physics and in related disciplines such as engineering. Although a few relevant studies have been conducted[9] (the results of which are consistent), most are unpublished. It is important that these studies be conducted and the results be widely disseminated. These investigations should lead to the development of strategies that help students apply the knowledge and skills developed in their physics courses to their subsequent studies or nonacademic pursuits.

Helping students to approach novel problems in a systematic fashion is a major goal of physics instruction. It also is one of the most difficult goals to achieve, although significant success has been reported.[10] However, much remains unknown. Efforts to understand the interrelationships among conceptual knowledge, mathematical skills, and logical reasoning ability should significantly enhance our progress toward helping students become better problem solvers.[11]

The rapid proliferation of computer-based technologies represents both an opportunity and a challenge. Technically sophisticated simulations, animations, and multimedia representations of physics concepts are being developed and implemented by many instructors and curriculum designers, but research into the effectiveness of these technologies lags far behind development.[12] It will be a major challenge to assess the effects of these technologies on student understanding of abstract physics concepts, the nature of scientific models, and the relation of both to the natural world. Such research is crucial for informing the implementation and further development of computer-based instructional tools.

In recent years, students’ beliefs about the nature of knowledge in physics and how it is acquired have become a major focus of interest.[13] There is reason to suspect that such epistemological beliefs can influence students’ learning of physics and their development of more generalized reasoning skills. Future directions will include efforts to understand these relationships and to incorporate the results in practical instructional strategies and materials.[14]

Although it has long been recognized that student knowledge is complex, there is now an increasing amount of research that focuses on the organization of this knowledge, the elements that it comprises, and the mechanisms by which it evolves.[15] In particular, the dynamics of learning are being investigated in studies that range from the construction of statistical and/or qualitative models of the knowledge states of students[16,17] to qualitative analyses of student thinking over the course of a single interview. The systematic analysis of student behavior during instruction will be an increasing focus for many workers.[18] The identification of common learning ''trajectories'' and strategies for promoting those that are productive would provide valuable assistance in the design of instructional methods and materials.

The findings of empirical investigations of student learning are usually accompanied by some speculation as to the underlying causes of common student errors or the nature of the learning process. In many cases this speculation is situation specific and is not tightly linked to an over-arching structure or theory. In this frequently successful approach, one attempts to affect *what* students do without being able to explain fully *why*. However, even this minimal-interpretation approach is carried out within a framework of specific ideas regarding the nature of the processes involved in learning physics.[19]

The refinement of such frameworks, with the ultimate goal of elucidating a few fundamental principles from which broad explanatory if not predictive power can be derived, is the focus of some PER workers.[20] Although this effort is potentially fruitful, it is important that theoretical descriptions remain firmly linked to empirically observable phenomena. The relationship between experiment and theory in PER will continue to be very different from that in traditional areas of physics from the standpoint of providing precise operational definitions and predictive power. In fact, in the context of PER we prefer to use the phrases ''models'' or ''theoretical frameworks'' to clearly differentiate generalizations about learning from the physical theories with which physicists are familiar. We expect that additional data from detailed studies of the dynamics of student learning will enhance efforts to establish useful theoretical frameworks. At the same time, we believe that empirical studies that are not necessarily closely identified with a specific theoretical framework will continue to lead to significant advances in instruction.

Whereas PER tends to focus on problems associated with the teaching of physics, cognitive science considers the nature of knowledge and learning in general. There is rough agreement on general principles between the two fields, but there has been relatively little cross fertilization, in part because differing goals have led to studies that have little detailed overlap. However, some PER researchers are working to build stronger connections between these two disciplines.[21] As more is learned about memory and learning, it will be a challenge to incorporate those findings into new lines of investigation within PER. An even greater challenge will be to incorporate these findings in practical classroom applications. Collaboration between members of the PER and cognitive science communities in designing and conducting experiments relevant to physics education could be useful and productive.

Physics is at the forefront, but discipline-based education research is growing in the other sciences and engineering. We believe that the PER community should actively cultivate connections with these related fields. Moreover, as we will discuss, lobbying for increased funding is more likely to be successful when broadly based.

Necessity for PER physicists within physics departments. Research on education in general, and on science teaching in particular, has been carried out for nearly a century. However, the impact of this research on undergraduate physics instruction is small compared to that from PER. The explanation is simple: education research conducted by physicists in physics departments is more credible, more accessible, and, in general, more relevant to physics faculty than that conducted in colleges of education or departments of psychology (although the conclusions are typically consistent). Thus for PER to be influential, it is essential that its researchers maintain close ties with the traditional physics community.

For PER to be both valid and useful, it is important that researchers have close, sustained, and day-to-day contact with physics students. Graduate students who work in this field need advanced training in physics and physics research methods, in addition to specialized training in PER. It is difficult to imagine that this training could occur without a firm base in a college or university physics department, for which undergraduate (and graduate) education is a central mission. In contrast, the mission of colleges of education is focused almost exclusively on K-12 instruction, with much less attention to discipline-specific instruction at the undergraduate level.

The close links to the rest of the physics community have enabled PER to make a contribution to education research that is unique.[22] Physicists have deep knowledge about physics concepts as well as familiarity with the methods and culture of the physics research community and the goals of physics instructors. These conditions have helped workers in PER to gain insights about physics learning and to develop instructional materials and methods that, although informed by work in related fields, have gone beyond those fields in terms of their direct impact on instructional practice. It is worth noting that ''the research-based development of tools and processes for use by practitioners''[23]—long the primary goal of most PER workers—is a relative rarity in traditional educational research. One of the strengths of PER is that it is not simply traditional education research conducted by individuals with a strong subject matter background, but rather it is a unique enterprise in which the techniques are strongly colored by the discipline in which it is embedded.

Practical and political issues facing the PER community. In the past seven years, more than 50 people who were trained in PER through Ph.D. or postdoctoral studies have obtained new tenure-track faculty positions in institutions ranging from four-year liberal arts colleges to research-oriented universities. At the same time, a number of physicists who had already achieved tenure through research in traditional areas have ''converted'' to PER. The pace of such conversions has increased in recent years, and such individuals form a significant fraction of PER workers. This dual-track expansion has allowed the field to grow rapidly. Although the numbers suggest that the field is thriving, there are several serious hurdles that must be overcome for PER to become a viable subfield of physics.

The fact that a significant fraction of PER faculty are tenure-track assistant professors is a concern. Although all tenure-track faculty have uncertain futures, there is an additional potential danger in PER. That is, there is a tendency in some departments for PER faculty to be viewed as resource people whose major responsibility is to provide local support for instruction rather than to conduct scholarly research. The responsibilities of PER faculty should be consistent with those of the other faculty in their departments, and they should have the same opportunities for promotion and tenure as faculty in other areas of physics. Although standards for teaching and service are primarily locally determined, criteria regarding publication can be set relative to national norms for PER, just as in other subfields of physics. These conditions are necessary for ensuring that the quality of PER is high and for ensuring that talented people continue to enter the field.

The current level of activity in PER requires a stable source of support to be sustained. Work in PER is primarily funded by the National Science Foundation (NSF) but the research aspect of funded projects is typically secondary to curriculum development, teacher education courses and workshops, and other applications of interest to the various funding programs. There is no source of funding for physics education research *per se*. When the research phase of a project is subservient to teacher education workshops or the production of curricular materials, the overall research and development endeavor is weakened. There are NSF programs that support science education research, but many PER projects are not competitive because they are perceived by the reviewers to be too narrowly focused. (Reviewers in these programs are drawn primarily from the traditional science education and cognitive science communities, instead of the physics community.) The traditional models of physics research funding, such as the renewable three-year grants provided to individual researchers by the NSF Divisions of Physics and of Materials Research, are virtually unknown in PER. However, the NSF Directorate for Mathematical and Physical Sciences (MPS) has recently taken tentative steps to support a small number of PER projects. If this initiative leads to increased and sustained support, it could have a significant impact.

We would like to see the Directorate for Mathematical and Physical Sciences support fundamental research on the learning and teaching of physics through competitive proposals submitted through standard procedures and peer-reviewed by experts in PER. A new program is not necessary—an explicit expansion of the types of projects considered suitable for submission would suffice. We recognize that the suggestion that MPS spread its limited funds over a larger number of areas is unlikely to find favor with much of the physics community. However, the lack of a funding base within NSF for discipline-based education research, despite the documented successes of this research, is a problem not just for physics but also for the other sciences and engineering. We would like to see physicists at NSF take the lead in establishing mechanisms for funding discipline-based education research within NSF. These programs could be jointly administered by the Division of Undergraduate Education and the appropriate divisions within the traditional research directorates.

A research field must have mechanisms to support the documentation, peer review, and dissemination of findings. For more than 25 years, the American Journal of Physics has served this function for PER, and also has served as the principal link between the PER community and the broader community of physics educators. (There are other journals in which research on physics teaching and learning is reported, but most have a limited readership in the U.S. among physics instructors at the postsecondary level.) There are now frequent special sections in AJP, overseen by an editor with expertise in PER, that provide a venue for PER articles that are more technically oriented than those in the main body of the journal. This development is an important acknowledgment of the role that AJP plays in the PER community. The proceedings of the annual Physics Education Research Conference provides a useful forum for the publication of short, preliminary accounts of investigations. The publication of the proceedings by the American Institute of Physics (starting with the 2003 conference) will make them much more widely accessible. An additional on-line archival journal with the tentative title *Physical Review Special Topics—Physics Education Research* is planned in partnership with the American Physical Society. Although a secure, long-term funding mechanism has not yet been established, we are hopeful that this new journal will greatly enhance the ability of members of the PER community to publish new and important results with a minimum of delay. Because it is critical that this new journal establish credibility in the physics community, we believe that the review criteria should resemble as closely as possible those in place for Physical Review as a whole.

While growing in size, the PER community also has diversified in terms of research themes, with both positive and negative future implications. The complex problem of improving physics learning requires that many and varied approaches be investigated and tested; not all will be fruitful, but that is the nature of research. However, the community is still relatively small and resources are limited. Too broad a dispersion of effort may result in research areas that fall below the critical mass needed to sustain a viable, self-critical, and productive research field. Collaborations could increase the impact of individual efforts and ensure that important issues receive adequate attention.

The growing number of faculty positions indicates that PER is increasingly viewed as a legitimate field for scholarly research by physicists in physics departments. However, many physicists still question whether effective teaching, long considered a skill or even an art, is amenable to scientific study. The large number of variables involved in student learning in the classroom is usually assumed to render the scientific study of physics education more difficult than most investigations of the physical world. We do not dispute this assumption, but we note that research in traditional areas of physics also is characterized by difficulties in identifying and controlling variables and by the necessity of making and assessing assumptions, approximations, and models. Physicists deal with these issues on a regular basis. Resolution comes only through the continual testing of models and assumptions by many research groups over the long term. In practice, the situation may well be significantly more challenging in PER, but it does not differ in principle.

As in traditional areas of physics, there are many careful experiments in PER and some that are not. Critical review of evidence by expert peers, the open debate of alternative interpretations, and experimental challenges to reported findings are the only way to ensure legitimacy. Therefore, it is especially crucial for members of the PER community to document their findings in sufficient detail to permit replication, to consider alternate interpretations explicitly, to cite the work of others, and to draw conclusions that are only as general as the scope of the given study warrants. A relatively new field such as PER has a special responsibility in these matters. At the same time, it is reasonable to expect that respectful consideration by the broader community of physicists will be given to well-executed PER investigations, just as would be given to such investigations in other areas of physics.

There are numerous examples of PER results that are highly robust and reproducible across diverse student populations, institutions, instructors, and nations. It is tempting to believe that the growing weight of such evidence will eventually overcome lingering doubts about the validity of PER within the larger physics community. These doubts reflect intellectual concerns and perhaps a generally conservative attitude about what and how we teach. However, efforts to convince skeptics by ''drowning them in data'' can engender further resistance. A backlash effect is created when the message heard by physics instructors is that *they* are ineffective and that *we*, the PER community, are the only ones who know how to teach. Results from a pilot study of attitudes toward PER held by mainstream physics faculty suggest that this type of miscommunication may be a significant issue.[24] There is a clear lesson here for physics education researchers. When communicating with the physics community, we must pay attention to the message received as much as the message that we intend to transmit. We must increase our efforts to assure our colleagues that PER results do not imply either that they are wasting their efforts in the classroom, or that their ideas are without merit. We also must try to correct the common inference that research-based instruction has no room for the creativity, intuition, or experience of individual instructors. And we must be careful not to over-generalize or over-simplify our results. Instead we should try to convey the simple premise on which PER rests: systematic research is an appropriate way to learn as much as possible about what students are learning and to guide improvements in instruction where indicated.

Conclusions. We have argued that it is important for PER to preserve and cultivate close connections with the traditional physics community, both to further the unique contributions made by physicists to the understanding of the learning of physics and to strengthen and widen the impact of PER on physics instruction in colleges and universities.

The regular inclusion of PER in AAPT and APS meetings and the growth in attendance at the annual Physics Education Research Conference are among the many signs of vigorous activity in this field. Physics education researchers are frequently invited to give colloquia in physics departments and PER is highlighted at AAPT-sponsored conferences including the New Faculty Workshop and the Conference of Physics Department Chairs. Prominent physics education researchers have been awarded the Oersted and Millikan awards, the highest honors of the AAPT. The Executive Committee of the APS Forum on Education is working to create stronger links between the AAPT, which is the traditional home of PER, and the APS. By maintaining high standards for PER and reaching out to the general physics community, we are optimistic that PER can become a firmly established and productive subfield of physics. The APS Council explicitly endorsed this outlook in its 1999 statement supporting PER in physics departments.[25] However, the differences in outlook between PER faculty and faculty in traditional areas of physics cannot be bridged solely by efforts from the PER community. Physicists in traditional areas need to acknowledge that the specialist knowledge of the PER community on instructional issues merits special consideration when physics pedagogy is the subject of discussion.

ACKNOWLEDGMENTS

We are grateful to Professor L. C. McDermott and Professor E. L. Jossem, and a number of other colleagues who read and commented on an earlier version of this Guest Editorial.

[1]L. C. McDermott, ''Research in physics education,'' APS News **7**(1), (1998); Education Outreach, p. 8.

[2]L. C. McDermott, ''Millikan Lecture 1990: What we teach and what is learned—Closing the gap,'' Am. J. Phys. **59**, 301–315 (1991); ''Guest Comment: How we teach and how students learn—A mismatch?,'' *ibid.* **61**, 295–298 (1993); R. R. Hake, ''Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,'' *ibid.* **66**, 64–74 (1998).

[3]L. C. McDermott and E. F. Redish, ''Resource Letter: PER-1: Physics Education Research,'' Am. J. Phys. **67**, 755–767 (1999); L. C. McDermott, ''Oersted Medal lecture 2001: 'Physics Education Research—The key to student learning','' *ibid.* **69**, 1127–1137 (2001).

[4]These were first published in AJP in the 1980s; see D. E. Trowbridge and L. C. McDermott, ''Investigation of student understanding of the concept of velocity in one dimension,'' Am. J. Phys. **48**, 1020–1028 (1980); ''Investigation of student understanding of the concept of acceleration in one dimension,'' *ibid.* **49**, 242–253 (1981). Also see F. Reif and S. Allen, ''Cognition for interpreting scientific concepts: a study of acceleration,'' Cognit. Instruct. **9**, 1–44 (1992).

[5]C. Singh, ''Student understanding of quantum mechanics,'' Am. J. Phys. **69**, 885–895 (2001); M. C. Wittmann, R. N. Steinberg, and E. F. Redish, ''Investigating student understanding of quantum physics: Spontaneous models of conductivity,'' *ibid.* **70**, 218–226 (2002); E. Cataloglu and R. W. Robinett, ''Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career,'' *ibid.* **70**, 238–251 (2002).

[6]M. E. Loverude, C. H. Kautz, and P. R. L. Heron, ''Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas,'' Am. J. Phys. **70**, 137–148 (2002).

[7]R. E. Scherr, P. S. Shaffer, and S. Vokos, ''Student understanding of time in special relativity: Simultaneity and reference frames,'' Am. J. Phys. **69**, S24–S35 (2001).

[8]B. S. Ambrose, ''Investigating student understanding in intermediate mechanics: Identifying the need for a tutorial approach to instruction,'' Am. J. Phys. **72**, 453–459 (2004).

[9]See, for example, G. E. Francis, J. P. Adams, and E. J. Noonan, ''Do they stay fixed?,'' Phys. Teach. **36**, 488–490 (1998).

[10]A. Van Heuvelen, ''Learning to think like a physicist: A review of research-based instructional strategies,'' Am. J. Phys. **59**, 891–897 (1991); ''Overview, Case Study Physics,'' **59**, 898–907 (1991); P. Heller, R. Keith, and S. Anderson, ''Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving,'' *ibid.* **60**, 627–636 (1992); P. Heller and M. Hollabaugh, ''Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups,'' *ibid.* **60**, 637–644 (1992); F. Reif and L. A. Scott, ''Teaching scientific thinking skills: Students and computers coaching each other,'' *ibid.* **67**, 819–831 (1999); A. Van Heuvelen and Xueli Zou, ''Multiple representations of work-energy processes,'' *ibid.* **69**, 184–194 (2001).

[11]W. J. Leonard, R. J. Dufresne, and J. P. Mestre, ''Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems,'' Am. J. Phys. **64**, 1495–1503 (1996); F. Reif and L. A. Scott, Ref. 10; C. Henderson, E. Yerushalmi, V. H. Kuo, P. Heller, and K. Heller, ''Grading student problem solutions: The challenge of sending a consistent message,'' Am. J. Phys. **72**, 164–169 (2004).

[12]D. J. Grayson and L. C. McDermott, ''Use of the computer for research on student thinking in physics,'' Am. J. Phys. **64**, 557–565 (1996); R. J. Beichner, ''The impact of video motion analysis on kinematics graph interpretation skills,'' *ibid.* **64**, 1272–1277 (1996); R. N. Steinberg, G. E. Oberem, and L. C. McDermott, ''Development of a computer-based tutorial on the photoelectric effect,'' *ibid.* **64**, 1370–1379 (1996); E. F. Redish, J. M. Saul, and R. N. Steinberg, ''On the effectiveness of active-

engagement microcomputer-based laboratories,'' *ibid.* **65**, 45–54 (1997); R. K. Thornton and D. R. Sokoloff, ''Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula,'' *ibid.* **66**, 338–352 (1998); F. Reif and L. A. Scott, Ref. 10; K. Cummings, J. Marx, R. Thornton, and D. Kuhl, ''Evaluating innovation in studio physics,'' Am. J. Phys. **67**, S38–S44 (1999); R. N. Steinberg, ''Computers in teaching science: To simulate or not to simulate?,'' *ibid.* **68**, S37–S41 (2000); Z. Zacharia and O. R. Anderson, ''The effects of an interactive computer-based simulation prior to performing a laboratory inquiry-based experiment on students' conceptual understanding of physics,'' *ibid.* **71**, 618–629 (2003); S. Yeo, R. Loss, M. Zadnik, A. Harrison, and D. Treagust, ''What do students really learn from interactive multimedia? A physics case study,'' *ibid.* **72**, 1351–1358 (2004).

[13]E. F. Redish, J. M. Saul, and R. N. Steinberg, ''Student expectations in introductory physics,'' Am. J. Phys. **66**, 212–224 (1998); D. Hammer, ''Student resources for learning introductory physics,'' *ibid.* **68**, S52–S59 (2000); D. B. May and E. Etkina, ''College physics students' epistemological self-reflection and its relationship to conceptual learning,'' *ibid.* **70**, 1249–1258 (2002).

[14]A. Elby, ''Another reason that physics students learn by rote,'' Am. J. Phys. **67**, S52–S57 (1999); ''Helping physics students learn how to learn,'' *ibid.* **69**, S54–S64 (2001).

[15]F. Reif, ''Millikan Lecture 1994: Understanding and teaching important scientific thought processes,'' Am. J. Phys. **63**, 17–32 (1995).

[16]L. Bao and E. F. Redish, ''Concentration analysis: A quantitative assessment of student states,'' Am. J. Phys. **69**, S45–S53 (2001); L. Bao, K. Hogg, and D. Zollman, ''Model analysis of fine structures of student models: An example with Newton's third law,'' *ibid.* **70**, 766–778 (2002).

[17]E. Bagno and B.-S. Eylon, ''From problem solving to a knowledge structure: An example from the domain of electromagnetism,'' Am. J. Phys. **65**, 726–736 (1997); N. S. Rebello and D. A. Zollman, ''The effect of distracters on student performance on the force concept inventory,'' *ibid.* **72**, 116–125 (2004).

[18]R. K. Thornton, ''Uncommon knowledge: Student behavior correlated to conceptual learning,'' in *Research on Physics Education: Proceedings of the International School of Physics ''Enrico Fermi'' Course CLVI*, edited by E. F. Redish and M. Vicentini (IOS, Amsterdam, 2004), pp. 591–601.

[19]P. R. L. Heron, ''Empirical investigations of learning and teaching, Part I: Examining and interpreting student thinking,'' in Redish and Vicentini, Ref. 18, pp. 341–350; ''Part II: Developing research-based instructional materials,'' *ibid.* pp. 351–365; D. E. Meltzer, ''The questions we ask and why: methodological orientation in physics education research,'' in *2003 Physics Education Research Conference (Madison, Wisconsin, 6-7 August 2003)*, edited by J. Marx, S. Franklin, and K. Cummings [AIP Conf. Proc. **720**, 11–14 (2004)]; P. R. L. Heron, ''Empirical investigations of student understanding,'' *ibid.* pp. 15–18.

[20]D. Hestenes, ''Toward a modeling theory of physics instruction,'' Am. J. Phys. **55**, 440–454 (1987); E. F. Redish, ''Millikan lecture 1998: Building a science of teaching physics,'' *ibid.* **67**, 562–573 (1999).

[21]E. F. Redish, ''Implications of cognitive studies for teaching physics,'' Am. J. Phys. **62**, 796–803 (1994); E. F. Redish, Ref. 20; E. F. Redish, ''A theoretical framework for physics education research: Modeling student thinking,'' in Redish and Vicentini, Ref. 18, pp. 1–63.

[22]L. C. McDermott, ''A view from physics,'' in *Toward a Scientific Practice of Science Education*, edited by M. Gardner, J. G. Greeno, F. Reif, A. H. Schoenfeld, A. diSessa, and E. Stage (Lawrence Erlbaum, Hillsdale, New Jersey, 1990), pp. 3–30.

[23]H. Burkhardt and A. H. Schoenfeld, ''Improving educational research: toward a more useful, more influential, and better-funded enterprise,'' Educational Researcher **32**, 3–14 (2003).

[24]C. Henderson and M. Dancy, ''Teaching, learning, and PER: Views from mainstream faculty,'' Poster CP-IP08, Physics Education Research Conference 2004, Sacramento, California, 4–5 August, 2004.

[25]http://www.aps.org/statements/99_2.cfm.

Paula R. L. Heron
Department of Physics, University of Washington, Seattle, Washington 98195-1560

David E. Meltzer
Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011-3160

MAXWELL'S GENIUS

In 1861, James Clerk Maxwell had a scientific idea that was as profound as any work of philosophy, as beautiful as any painting, and more powerful than any act of politics of war. Nothing would be the same again.

In the middle of the nineteenth century the world's best physicists had been searching for a key to the great mystery of electricity and magnetism. The two phenomena seemed to be inextricably linked but the ultimate nature of the linkage was subtle and obscure, defying all attempts to winkle it out. Then Maxwell found the answer with as pure a shaft of genius as has ever been seen.

Basil Mahon, *The Man Who Changed Everything: The Life of James Clerk Maxwell* (Wiley, 2003), p. 1.

Chemical Education and Physics Education: Facing Joint Challenges and Practical Concerns*

Paula R. L. Heron [pheron@phys.washington.edu]
David E. Meltzer [dem@iastate.edu]

Recently there have been discussions between chemists and physicists directed at organizing a series of joint meetings on learning and teaching, with chemical educators presenting their work at conferences of the American Association of Physics Teachers, and physicists reciprocating at ACS meetings. During these discussions we have learned of the large common ground we share in our work, particularly in the area of discipline-based education research at colleges and universities. Our common experiences and interests suggest that much may be gained by further joint discussions and activities.

For over a hundred years, colleges and universities in the United States have carried out intensive instruction in chemistry, physics, and other technical subjects. Systematic research in these fields has expanded tremendously during this period as well. However, it is a relatively recent phenomenon that the process of learning and teaching these subjects at the college level has been regarded as a potentially fruitful area of research in its own right. As readers of this *Newsletter* are probably aware, chemical education research (CER) has become a research specialty pursued at about two dozen research-oriented chemistry departments in the U.S. More than half of these departments currently award the Ph.D. degree for dissertation research in chemical education. Similarly, there has been explosive growth in the field of physics education research (PER) during the past fifteen years. There are now over 80 physics departments at U.S. colleges and universities employing at least one faculty member pursuing research on the teaching and learning of physics. About 35 of these departments are research-intensive Ph.D.-granting departments, and approximately 15 such departments have already awarded or plan to award a Ph.D. degree in physics for research in physics education.

The development of PER reflects the growing realization by many physicists that in contrast to the efforts that have led to remarkable advances in science and technology, the enormous effort expended by many physics instructors over the past century was not harnessed in a way that made cumulative progress in education likely. Instead, innumerable individual instructors discovered and re-discovered inadequacies in popular teaching methods and instructional materials, and developed their own ways to address these problems based on personal experience through trial-and-error methods. However, as one leading physics education researcher has observed, "Unless we are willing to apply the same rigorous standards of scholarship to issues related to learning and teaching that we regularly apply in more traditional research, the present situation in physics education is unlikely to change."[1] As is occurring in chemistry, an increasing number of physicists have taken up this challenge by applying methods of research based on those that have been employed successfully in investigations of the physical world. Systematic studies of student learning have been carried out that incorporate careful collection and analysis of data based on deep probes of students' reasoning. Many such studies have been replicated with widely diverse institutions, instructors, and student populations. These investigations have revealed a wide gap between the objectives of most physics instructors engaged in traditional forms of instruction and the actual level of conceptual understanding attained by most of their students.[2]

But PER has gone beyond documenting shortcomings in student learning and traditional instruction. Physics education researchers have developed instructional materials and methods that have been subjected to repeated testing, evaluation, and re-design. Numerous reports have documented significant and reproducible learning gains from the use of these materials and methods in courses ranging from large-enrollment classes at major public universities to small classes in two-year colleges and high schools.[2] In what follows, we will identify some of the current and emerging research directions in PER that we consider promising. We also identify some practical and political challenges to the growth of PER, and we argue that these are virtually identical to challenges facing the CER community. With that in mind, we will propose some common steps that could be taken to help ensure the stability, growth, and productivity of both CER and PER.

Current and emerging research directions in PER
Most early PER work focused on student ability to apply the concepts covered in typical introductory university physics

* *This article is adapted from a paper to be published in the American Journal of Physics.*

courses. The results of these studies have proven invaluable in guiding improvements in instruction. The breadth of topics covered, their importance as a foundation for future study, and the many students involved ensures that research in the introductory course will continue to be a major emphasis for the foreseeable future. However, in recent years, there has been an increasing focus on student learning in upper-level courses such as quantum mechanics, thermal physics, relativity, and advanced mechanics. This research should lead to learning gains for physics majors similar to those found for research-based instruction at the introductory level.

Helping students to approach novel problems in a systematic fashion is a major goal of physics instruction. It also is one of the most difficult goals to achieve, although a few approaches have had significant success.[3] Efforts to understand the interrelationships among conceptual knowledge, mathematical skills, and logical reasoning ability should significantly enhance our progress toward helping students become better problem solvers.

In both physics and chemistry, the rapid proliferation of computer-based technologies represents both an opportunity and a challenge. Technically sophisticated simulations, animations, and multimedia representations of physics and chemistry concepts are being developed and implemented by many instructors and curriculum designers, but research into the effectiveness of these technologies lags far behind development. It will be a major challenge to assess the effects of these technologies on student understanding of abstract concepts, the nature of scientific models, and the relation of both to the natural world. Such research is crucial for informing the implementation and further development of computer-based instructional tools.

In recent years, students' beliefs about the nature of knowledge in physics and how it is acquired have become a major focus of interest.[4] There is reason to suspect that such epistemological beliefs can influence students' learning of physics and their development of more generalized reasoning skills. Future directions will include efforts to understand these relationships and to incorporate the results in practical instructional strategies and materials.

The empirical investigations of student learning in PER are usually carried out within a framework of ideas regarding the underlying causes of common student errors and the nature of the learning process. The refinement of such frameworks, with the ultimate goal of elucidating a few fundamental principles from which broad explanatory if not predictive power can be derived, is the focus of some physics education researchers.[5]

Necessity for Discipline-Based Education Research

Research on education in general, and on science teaching in particular, has been carried out for nearly a century. However, the impact of this research on undergraduate science instruction is small compared to the influence of education research originating from within the disciplines. The explanation is simple: education research conducted by scientists in science departments is more credible, more accessible, and more relevant to college and university science faculty than that conducted in colleges of education or departments of psychology, although the conclusions are typically consistent. Moreover, we note that "the research-based development of tools and processes for use by practitioners"[6]—long a primary goal of most CER and PER workers—is a relative rarity in traditional educational research. Thus while we view increased collaboration with cognitive psychologists, education researchers and neuroscientists to be of potential benefit, we believe that for CER and PER to continue to be influential, it is important for researchers to maintain close ties with the traditional science community.

Practical and Political Issues Facing the PER Community

Although the rapid growth of PER suggests that the field is thriving, there are several serious hurdles that must be overcome for it to become a viable subfield of physics. The fact that a significant fraction of PER faculty are tenure-track assistant professors is a concern. There is a tendency in some departments for PER faculty to be viewed as resource people, whose major responsibility is to provide local support for instruction rather than to conduct scholarly research. We believe that faculty in PER and CER need to have the same opportunities for advancement based on their scholarly work as those in other areas of research. This will ensure that the quality of physics and chemistry education research remains high and that talented people continue to enter these fields.

Another area of concern is the availability of funding. The current level of activity in CER and PER requires a stable source of support to be sustained. Such work is now funded primarily by the National Science Foundation, but in general the research aspect of funded projects is secondary to curriculum development, teacher education courses and workshops, and other programs of interest to the various funding agencies. There is no source of funding for chemistry or physics education research *per se*. The traditional models of chemistry and physics research funding, such as the renewable three-year grants available to individual researchers by the NSF Divisions of Physics and Chemistry, are virtually unknown in PER and CER. The lack of a funding base within NSF for discipline-based education research, despite its documented successes, is a problem not just for chemistry and physics but also for the other sciences and engineering. We believe that chemists and physicists at NSF could be effective in promoting the establishment of mechanisms for funding discipline-based education research within the Foundation. The NSF Directorate for Mathematical and Physical Sciences (MPS) has recently taken tentative steps to support a small number of PER projects. If this initiative leads to increased and sustained support, it could have a significant impact.

The growing number of faculty positions indicates that CER and PER are increasingly viewed as legitimate fields for scholarly research by chemists and physicists in college and university science departments. However, many still question whether effective teaching, long considered a skill or even an art, is amenable to scientific study. The large number of variables involved in student learning in the classroom is usually assumed to render the scientific study of science education more difficult than most investigations of the physical world. We do not dispute this assumption, but we note that research in traditional areas of science also is characterized by difficulties in identifying and controlling variables and by the necessity of making and assessing assumptions, approximations, and models. Chemists and physicists deal with these issues on a regular basis. Resolution comes only through the continual testing of models and assumptions by many research groups over the long term. Therefore, it is reasonable to expect that respectful consideration by the broader community of scientists will be given to well executed CER and PER investigations, just as would be given to such investigations in other areas of science. After all, CER and PER both rest on a simple, fundamental premise that should have wide credibility among research scientists, that is: systematic research is an appropriate way to learn as much as possible about what students are learning and to guide improvements in instruction where indicated.

Conclusion

By promoting joint discussions and collaborative projects, workers in CER and PER have the opportunity to enhance significantly the impact of their work on the traditional chemistry and physics communities. Some such collaborations have already resulted in joint NSF-supported projects and joint publications in science education journals.[7] We believe that extension of such collaborations on a national scale would be an important development in furthering discipline-based science education research at the undergraduate level.

References

[1] L. C. McDermott, "Research in Physics Education," APS News **7** (1), (1998); Education Outreach, p. 8.

[2] L. C. McDermott and Edward F. Redish, "Resource Letter: PER-1: Physics Education Research," Am. J. Phys. **67**, 755–767 (1999).

[3] A. Van Heuvelen, "Learning to think like a physicist: A review of research-based instructional strategies," Am. J. Phys. **59**, 891-897 (1991); P. Heller, R. Keith, and S. Anderson, P. Heller and M. Hollabaugh, "Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups," Am. J. Phys. **60**, 637-644 (1992); F. Reif and L. A. Scott, "Teaching scientific thinking skills: Students and computers coaching each other," Am. J. Phys. **67**, 819–831 (1999).

[4] E. F. Redish, J. M. Saul, and R. N. Steinberg, "Student expectations in introductory physics," Am. J. Phys. **66**, 212–224 (1998).

[5] E. F. Redish, "Millikan lecture 1998: Building a science of teaching physics," Am. J. Phys. **67**, 562–573 (1999).

[6] H. Burkhardt and A. H. Schoenfeld, "Improving educational research: toward a more useful, more influential, and better-funded enterprise," Educational Researcher **32**, 3–14 (2003).

[7] See, for example, T.J. Greenbowe and D.E. Meltzer, "Student learning of thermochemical concepts in the context of solution calorimetry," Int. J. Sci. Educ. **25**, 779-800 (2003). ■

A Call to the AAPT Executive Board and Publications Committee to Expand Publication of Physics Education Research Articles within the American Journal of Physics

David E. Meltzer
Iowa State University

Lillian C. McDermott
University of Washington

Paula R. L. Heron
University of Washington

Edward F. Redish
University of Maryland

Robert J. Beichner
North Carolina State University

This document has been endorsed by the
AAPT Committee on Research in Physics Education, Stamatis Vokos, Chair.

REVISED December 29, 2003

A Call to the AAPT Executive Board and Publications Committee to Expand Publication of Physics Education Research Articles within the American Journal of Physics

Summary: *The recent dramatic expansion of activity in physics education research among AAPT members has not been matched by commensurate increases in publication venues. Although the impact of this research field within the broader physics community has sharply increased, the viability of its continued existence is dependent upon substantially expanded publication opportunities in the near future. The American Journal of Physics has served for three decades as the primary publication venue for results in physics education research. An increased number of pages devoted to physics education research is consistent both with AJP's historical role and with the greater prominence in recent years of the PER community within AAPT. We recommend (1) considering PER submissions to the main section of AJP on a par with submissions in other subject areas, (2) increasing the number of pages allocated to the PER Section, and (3) allowing the option of increasing the publication frequency of the PER Section from its present rate.*

Introduction: Evaluating and improving the teaching and learning of physics is a prime concern for a large proportion of all physicists, and is the central focus of the AAPT. Significant numbers of physicists have begun to apply to the problems involved in teaching and learning physics the same systematic methods of research and analysis they have employed so successfully in investigating the physical world. They have carried out detailed, systematic, and reproducible studies involving the collection and analysis of data reflecting student thinking and performance. This endeavor, broadly known by the term "physics education research" (PER), has in recent years undergone rapid expansion both in the numbers of physicists involved, and in the recognition and impact of its results within the broader physics community.

The role of physics education research in advancing the teaching of physics: The role of PER within AAPT is perhaps best understood by examining the goals of AAPT itself. The AAPT Mission Statement, posted on the AAPT home page [http://www.aapt.org/aboutaapt/mission.cfm], stresses that it is "*committed to providing the most current resources and up-to-date research needed to enhance a physics educator's professional development*." The Mission Statement continues: "*The Association has identified four critical issues that will guide our future activities,*" among which it includes the following: "*#3: Improve the pedagogical skills and physics knowledge of teachers at all levels; #4: Increase our understanding of physics learning and of ways to improve teaching effectiveness.*" Physics education research is devoted to achieving precisely these objectives.

The goal of physicists working in PER is, broadly speaking, to increase the effectiveness and efficiency of physics education at all levels, from the pre-secondary level up to the graduate level, and for the public in non-academic settings. In recent years, PER has had a dramatic impact on the way in which physics is taught, on the ways in which many physics educators view the issues involved in their profession, and in the preparation of physics teachers at both the high-school and university level. The published findings of physics education research, based on rigorous and reproducible testing and measurement, have disclosed heretofore unknown or under-appreciated aspects of the traditional process of physics education. Research has revealed the broad gap that often exists between the objectives physics instructors have for their courses, and the actual level of conceptual understanding attained by most students engaged in traditional forms of instruction. Ongoing research has clarified the dynamics of student thinking during the process of learning physics, revealing both particular learning difficulties, as well as effective strategies for guiding student insight and understanding. Based directly or indirectly on this research, many new forms of curricular materials and instructional methods have been developed and disseminated throughout the nation and the world. Countless reports have documented improved learning gains resulting from the use of research-based curricula and instructional methods.

The results of research and of research-based instructional methods have thrust the concept of "active engagement" or inquiry-based learning into the forefront of the entire physics education community. Led by workers in PER, innumerable studies have demonstrated the effectiveness of forms of instruction that supplement (in some cases, replace) traditional lecture-based methods with inquiry-based learning based on cooperative groups. Students are guided to work their way through carefully designed and tested sequences of questions, exercises, and/or laboratory activities. Utilizing these research-based curricula, and interacting frequently during class with instructors and with each other, students have often achieved significant gains in understanding when compared with instruction based on lecture alone. By basing the design of curricula and instructional methods on the results of physics education research, and by subjecting them to repeated testing, evaluation, and re-design, dramatic learning gains have been made in physics courses of all types, from large-enrollment classes at huge public universities to small-group laboratory courses in junior colleges and high-school classrooms. Many workshops involving hundreds of new college and university faculty members have been held by AAPT in which the new forms of research-based instruction have been placed at the forefront, and PER researchers have led the majority of plenary sessions.

Due in significant part to the efforts of the physics education research community over the past 20 years, the field of physics education is enjoying a heretofore unknown degree of growth and prominence at all levels, from the elementary and middle schools, through high schools, junior colleges, four-year colleges, and universities. The rapid influx of new participants into the PER community, now occurring to an extent never seen before, offers the promise of additional dramatic advances in physics education in the future based on and guided by new research findings. The degree to which this dynamic expansion in impact and outreach can be sustained will depend, in large part, on the well-being and growth of the physics education research community itself. As is true for any research field, a central issue for the PER community is the effectiveness and flexibility of its means for documentation and dissemination of research results – that is, its form of publication. For the field of PER over the past few decades, the American Journal of Physics has been the central link between researchers in physics education, and the broader community of physics educators worldwide.

In order to implement AAPT's mission of providing the most up-to-date research needed by physics educators, and of increasing our understanding of physics learning and of ways to improve teaching effectiveness, some form of archival record is needed. Only such a record can ensure wide and continuing dissemination of the results obtained by workers in physics education, and can serve as a basis on which to build future advances. The unique tool available to the AAPT for providing this archival record has been and continues to be the American Journal of Physics.

The place of PER within the American Journal of Physics: In a recent editorial introducing the PER section in the American Journal of Physics, some specific criteria were given to characterize research papers in PER:

> Articles . . . are expected to focus more on questions of not only what we think we know about student learning, but how we know and why we believe what we think we know. Articles in PERS can be expected to address a wide range of topics from theoretical frameworks for analyzing student thinking to developments of research instruments for the assessment of the effectiveness of instruction and to the development and comparison of

> different teaching methods. Articles should include careful discussions of research methodology and how the work was done.[1]

A somewhat broader characterization of PER was given by the editor of the PER Supplement to AJP (introduced in 1999 and merged into AJP itself as a special section in 2002):

> It focuses on using the methods and culture of science to help us understand how students learn physics and how to make our instruction more effective. By the methods of science, I mean careful observation and analysis of the phenomenon under study. By the culture of science, I mean documenting and publishing research to evaluate and critique the work for the purpose of building a community consensus of what we know.[2]

It is important to recognize that research falling under the broad definition of PER has been carried out and published not only recently, but rather for several decades. For over 30 years, the primary means of documentation of physics education research and of communicating its results to the worldwide physics community has been the American Journal of Physics. More than 120 papers describing the methods and results of research into physics learning were published in AJP from 1972 to 1998.[3] An approximate breakdown of these papers is as follows: 1972-1979: 35 papers (4.4 per year; range: 1-8 per year); 1980-1989: 38 (3.8 per year; range: 0-9 per year); 1990-1998: 54 papers (6.0 papers per year; range: 3-9 per year). Some of these early papers are listed in Appendix A.

Although the official policy of AJP has always been that it is not a “research journal,” actual editorial practice has long acknowledged, in effect, that the exclusion of research papers adopted by the journal’s founders was aimed at research in the traditional subfields of physics (nuclear, high-energy, condensed matter, etc.). As is demonstrated by the figures cited above, papers devoted to research investigations in the teaching and learning of physics have been *continuously* published in AJP for over three decades. Many of these papers (including dozens published before 1999) incorporate extensive data tables, complex methodologies for data collection and analysis, and lengthy discussions of methods and results. These features are characteristic of papers published in archival physics research journals, and demonstrate that it has long been considered appropriate for AJP papers devoted to physics education research to adopt the format and style of research papers in traditional physics areas. (Such papers occasionally may be viewed as less readily accessible to an ordinary physics teacher “practitioner.” However, this is surely no different from the similarly limited accessibility of many highly specialized papers currently published in AJP, often readable only by physicists with advanced-level training in very specific areas.)

In fact, the American Journal of Physics has long served as the dominant English-language forum for publication of investigations carried out by physicists that focus on research into teaching and learning of physics at the college and university level. Certainly there are other journals in which research regarding physics teaching and learning is and has been reported. However, most of these journals are primarily devoted to research carried out by non-physicists in broad areas of science instruction at the pre-college level, and they have extremely limited readership among physics instructors at the post-secondary level. By any measure, the circulation, readership, and recognition of the American Journal of Physics among the university physics community is *overwhelmingly* greater than any comparable publication.

The role of PER within the physics community: The reality of AJP’s dominant role has in recent years taken on increased significance as the size and impact of the physics education research community has grown. A very important indication of this increased impact was the May 21, 1999 statement by the Council of the American Physical Society:

99.2 RESEARCH IN PHYSICS EDUCATION

(Adopted by the Council, 21 May 1999)

In recent years, physics education research has emerged as a topic of research within physics departments. This type of research is pursued in physics departments at several leading graduate and research institutions, it has attracted funding from major governmental agencies, it is both objective and experimental, it is developing and has developed publication and dissemination mechanisms, and Ph.D. students trained in the area are recruited to establish new programs. Physics education research can and should be subject to the same criteria for evaluation (papers published, grants, etc.) as research in other fields of physics. The outcome of this research will improve the methodology of teaching and teaching evaluation.

The APS applauds and supports the acceptance in physics departments of research in physics education. Much of the work done in this field is very specific to the teaching of physics and deals with the unique needs and demands of particular physics courses and the appropriate use of technology in those courses. The successful adaptation of physics education research to improve the state of teaching in any physics department requires close contact between the physics education researchers and the more traditional researchers who are also teachers. The APS recognizes that the success and usefulness of physics education research is greatly enhanced by its presence in the physics department.[4]

In fact, as this statement suggests, the growth of physics education research as a research subfield within U.S. physics departments has been extraordinarily rapid over the past six years. There has been approximately a fourfold expansion in the number of physics departments that now include among their faculty one or more members whose scholarly efforts are devoted primarily or entirely to work in physics education. More than fifty tenure-track faculty positions in the U.S. have been filled during this period by physics education researchers,[5] with almost all of these at the junior-faculty level. At least 30 Ph.D.-granting physics departments now include tenured or tenure-track PER faculty, most of whom are guiding (or preparing to guide) graduate students toward Masters or Ph.D. degrees in physics education research.[6]

The explosion of interest and participation in physics education research has also been dramatically apparent at the national meetings of the AAPT. For most of the past decade, sessions devoted to PER papers have routinely been filled to overflowing, and increasingly large proportions of both invited and contributed presentations at AAPT meetings have been devoted to physics education research. Ever more workshops are being sponsored by the Research in Physics Education committee. Attendance at the annual Physics Education Research Conference – extending an extra day beyond the end of the summer AAPT meeting – has now nearly reached 200 physicists.

The current status of publication outlets for PER: In startling contrast to the rapid growth of activity in physics education research, the availability of publication venues has not kept pace. A PER Supplement to AJP began publication in 1999 and has recently transformed into a separate, twice-yearly section within AJP itself. Although we have made some progress since the early years (see next paragraph), the growth has been modest and it is clear that there is increasing demand. Further, the hiring patterns described above suggest that we need to be prepared to respond quickly and effectively to increasing demand.

The number of PER papers published in AJP since 1999 is as follows: ***1999***: main section, 9; PER Supplement, 8; ***2000***: main section, 0; PER Supplement, 7; ***2001***: main section, 6; PER Supplement, 6; ***2002***: main section, 7; PER Section, 7; ***2003***: main section, 3; PER Section, 4; ***2004***: main section, 0; PER Section, 2. The average number of PER papers in the main section of AJP is now actually *lower* than typical rates from earlier years.

This is not due to a lack of publishable work; rather, the artificial limitation on the number of pages allowed for PER papers in AJP has itself served to constrain the efforts of researchers within the field. The increasingly long backlog-induced delays for the PER Section – now at approximately two years – and the impression that PER papers appear only infrequently within the main section of AJP, have in some cases led researchers to delay writing and submitting mature research results that had already been widely disseminated through other means such as invited and contributed presentations, workshops, web sites, etc. Often, the only practical and rapid publication option for researchers has been to submit short summary reports of their work to the annual *Proceedings* of the Physics Education Research Conference.

The rapid increase in number of submissions to the *Proceedings* (47 papers were submitted to this year's edition) is evidence of the pent-up demand within the PER community for publication venues. However, the extremely limited circulation of the *Proceedings* (now and for the foreseeable future) implies both a much-lessened impact for this work within the broader physics community, as well as uncertain acceptance by departmental tenure and promotion committees upon whose decisions the continued employment of PER researchers depends. In many research-oriented departments, *Proceedings* papers are not counted as being on a par with publication in established journals, and in some departments they may not count at all.

Very recently the possibility has arisen of an electronic publication venue coming into existence based on limited-term funding from the National Science Foundation. This electronic journal forms one component of the PER-CENTRAL project (Community Enhancing Network for Teaching, Research, and Learning). [Funding has been approved for one year, with the possibility of an additional two years of funding.] The PER community has hopes that this outlet may grow, in the long term, into a significant alternative publication venue for research papers in the field. However, the overall project has a wide scope and will require substantial time to ramp up from its start-up phases into full functioning. The journal component will require assembling additional editorial and production resources, a process that necessarily requires time and some initial testing. A significant challenge will be to develop, over time, a long-term funding mechanism that could sustain the new publication into the indefinite future, beyond the initial three-year period of NSF funding.

For now, the question of the reputation and ultimate acceptance of the new publication venue within the broader physics community is unresolved. This acceptance is critical to the journal's potential viability as an effective publication outlet for PER researchers. The current physics culture includes very few "electronic-only" journals; the vast majority of physics research papers are expected and required to have parallel paper publication. A new community still working to become generally accepted may incur a significant risk by concentrating a large fraction of its output in a venue seen as novel and somewhat "pioneering." The risk of depending primarily on a publication venue whose acceptance in the physics community is unproven may be particularly acute when considering the possible response of physics departments at major research universities. Their willingness to hire and promote PER faculty is critical to maintaining the credibility and influence of the field.

It is true that some purely electronic journals have become the primary publication routes in their field. Thus far, in the U.S. physics community, such publications have well-established and stable funding sources. Over time, the new electronic PER journal may evolve to a point where it can play a similar role. However, it is unrealistic to expect that any solely electronic journal can, in the short term, fulfill the role of primary publication outlet for all PER research articles. This is as much for "reputability" reasons as for logistical ones. A more practical approach might be for the new journal to take on a gradually increasing portion of the publication burden, as its production mechanisms and community acceptance grow and strengthen. Thus this would represent more of an "evolutionary," rather than a "revolutionary" approach. At best, it will be several years before the critical questions regarding the new venture can be answered positively and definitively. It will take some time before the new publication can be established as a legitimate counterpart to AJP within the PER and the broader physics communities

Meanwhile, the number of graduate students, post-doctoral researchers, and junior and senior faculty in PER continues its steady increase. The quantity of research being carried out is rapidly expanding, and adequate publication venues are an urgent, critical necessity to the continued viability of the field. The PER community has previously expressed its strong sentiment that the number of pages within AJP allocated to PER needs to expand at a rate commensurate with the rate of high-quality articles submitted. An increase in the amount of AAPT publication resources devoted to PER is more than adequately justified by the soaring levels of interest and participation in PER work by AAPT members that have been repeatedly demonstrated at the national and regional meetings of the AAPT. Moreover, any further delay in increasing PER publication within AJP will likely have devastating effects on the ability of workers in the field to maintain their effectiveness – if not their very existence – within their respective institutions.

As we have pointed out above, the recent trend has been that the number of PER papers within the main section of AJP has declined (in some years, to zero), while at the same time the number allowed in the PER Section has been rigidly constrained. The net result has been to greatly underserve the needs of the PER community within AAPT with regard to publication opportunities. A research community that has grown by more than an order of magnitude is being forced to operate with fewer publishing opportunities than existed 20 years ago.

Conclusion: *For these reasons we call on the AAPT Executive Board and Publications Committee to take measures sufficient to allow rapid publication in AJP of PER papers accepted through the normal review process, including the following: (1) recommending very strongly to the editors of AJP that PER submissions be considered for inclusion in the main section of AJP on a par with submissions in other subject areas, and that they not be automatically directed to the PER section unless explicitly requested by the authors, (2) immediately increasing the total number of pages within AJP allocated to the PER Section, (3) allowing the option of including several shorter PER Sections within AJP more often than the current twice-per-year rate.*

References

1. Jan Tobochnik, Editor and Edward F. Redish PERS Editor, "Physics Education Research Section (PERS)," Am. J. Phys. **70**, 666 (2002).
2. Edward F. Redish, "EDITORIAL," Am. J. Phys. **67**, S3 (1999).
3. A listing of PER papers published in AJP since 1972 is available as a separate document.
4. http://www.aps.org/statements/99_2.cfm
5. See Appendix B.
6. See Appendix C.

Appendix A

A selection of PER papers published in AJP between 1976 and 1994:

F. Reif, Jill H. Larkin, and George C. Brackett, "Teaching general learning and problem-solving skills," Am. J. Phys. **44**, 212 (1976).

John W. Renner and William C. Paske, "Comparing two forms of instruction in college physics," Am. J. Phys. **45**, 851 (1977).

Richard Vawter, "Entropy state of a multiple choice examination and the evaluation of understanding," Am. J. Phys. **47**, 320 (1979).

David E. Trowbridge and Lillian C. McDermott, "Investigation of student understanding of the concept of velocity in one dimension," Am. J. Phys. **48**, 1020 (1980).

Audrey B. Champagne, Leopold E. Klopfer, and John H. Anderson, "Factors influencing the learning of classical mechanics," Am. J. Phys. **48**, 1074 (1980).

John Clement, "Students' preconceptions in introductory mechanics," Am. J. Phys. **50**, 66 (1982).

P. C. Peters, "Even honors students have conceptual difficulties with physics," Am. J. Phys. **50**, 501 (1982).

Robert J. Whitaker, "Aristotle is not dead: Student understanding of trajectory motion," Am. J. Phys. **51**, 352 (1983).

R. Cohen, B. Eylon, and U. Ganiel, "Potential difference and current in simple electric circuits: A study of students' concepts," Am. J. Phys. **51**, 407 (1983).

L. Viennot, "Analyzing students' reasoning: Tendencies in interpretation," Am. J. Phys. **53**, 432 (1985).

Peter W. Hewson, "Diagnosis and remediation of an alternative conception of velocity using a microcomputer program," Am. J. Phys. **53**, 684 (1985).

W. T. Griffith, "Factors affecting performance in introductory physics courses," Am. J. Phys. **53**, 839 (1985).

Ibrahim Abou Halloun and David Hestenes, "The initial knowledge state of college physics students," Am. J. Phys. **53**, 1043 (1985).

Ibrahim Abou Halloun and David Hestenes, "Common sense concepts about motion," Am. J. Phys. **53**, 1056 (1985).

Fred M. Goldberg and Lillian C. McDermott, "An investigation of student understanding of the real image formed by a converging lens or concave mirror," Am. J. Phys. **55**, 108 (1987).

Monica G. M. Ferguson-Hessler and Ton de Jong, "On the quality of knowledge in the field of electricity and magnetism," Am. J. Phys. **55**, 492 (1987).

Lillian C. McDermott, Mark L. Rosenquist, and Emily H. van Zee, "Student difficulties in connecting graphs and physics: Examples from kinematics," Am. J. Phys. **55**, 503 (1987).

Richard F. Gunstone, "Student understanding in mechanics: A large population survey," Am. J. Phys. **55**, 691 (1987).

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

Alan Van Heuvelen , "Overview, Case Study Physics," Am. J. Phys. **59**, 898 (1991).

Patricia Heller, Ronald Keith, and Scott Anderson, "Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving," Am. J. Phys. **60**, 627 (1992).

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

S. Törnkvist, K.-A. Pettersson, and G. Tranströmer, "Confusion by representation: On student's comprehension of the electric field concept," Am. J. Phys. **61**, 335 (1993).

Beth Thacker, Eunsook Kim, Kelvin Trefz, and Suzanne M. Lea, "Comparing problem solving performance of physics students in inquiry-based and traditional introductory physics courses," Am. J. Phys. **62**, 627 (1994).

Robert J. Beichner, "Testing student interpretation of kinematics graphs," Am. J. Phys. **62**, 750 (1994).

Edward F. Redish, "Implications of cognitive studies for teaching physics," Am. J. Phys. **62**, 796 (1994).

S. Rainson, G. Tranströmer, and L. Viennot, "Students' understanding of superposition of electric fields," Am. J. Phys. **62**, 1026 (1994).

Appendix B

The following departments have filled tenure-track PER positions within the past six and a half years; numbers in parentheses indicate number of positions filled. (In some cases the positions are joint between the physics department and another department.) Only five of these positions were filled with faculty who were hired with tenure:

1. American University (DC)
2. University of Arizona (2) [Physics Department; Astronomy Department]
3. Arizona State University
4. Buffalo State College (SUNY)
5. California State University, Chico
6. California State University, Fullerton
7. California State University, San Marcos (2)
8. University of Central Florida
9. Chicago State University
10. City College of New York (2)
11. University of Colorado (2) [Physics Department; School of Education]
12. Concordia College (MN)
13. Davidson College (NC)
14. Dickinson College (PA) (2)
15. Drury University (MO)
16. Grand Valley State University (MI)
17. Hawai'i Pacific University
18. High Point University (NC)
19. Iowa State University
20. Kansas State University
21. University of Maine (2)
22. University of Maryland
23. McDaniel College (MD)
24. University of Minnesota [General College]
25. New Mexico State University
26. North Carolina State University (2)
27. University of Northern Iowa (2)
28. The Ohio State University
29. Rochester Institute of Technology (NY)
30. Rutgers University (NJ) (2) [Physics and Astronomy Department; Graduate School of Education]
31. Seattle Pacific University
32. Southeastern Louisiana University
33. Southern Connecticut State University
34. Southern Illinois University, Edwardsville (3)
35. Southwest Missouri State University
36. University of Texas at Dallas [Department of Science/Mathematics Education]
37. University of Texas at El Paso
38. Texas Tech University
39. Towson University (MD) (2)

40. U.S. Air Force Academy (CO) (2)
41. University of Washington
42. Western Carolina University (NC)
43. Western Kentucky University
44. Western Michigan University
45. University of Wisconsin-Oshkosh
46. University of Wisconsin-Stout
47. Worcester Polytechnic Institute (MA)

Appendix C

The following Ph.D.-granting physics (or physics and astronomy) departments have tenured, tenure-track, or research or teaching faculty members with a substantial or a primary interest in physics education research (* indicates non-tenure-track):

1. University of Arizona
2. University of Arkansas
3. Arizona State University
4. University of California, Davis*
5. University of Central Florida
6. City College of New York
7. University of Colorado
8. Harvard University
9. University of Illinois
10. Iowa State University
11. Kansas State University
12. University of Maine
13. University of Maryland
14. University of Massachusetts, Amherst
15. University of Minnesota
16. Mississippi State University
17. Montana State University
18. University of Nebraska
19. New Mexico State University
20. North Carolina State University
21. The Ohio State University
22. University of Oregon
23. Oregon State University
24. University of Pittsburgh*
25. Rutgers University
26. San Diego State University (joint program with University of California at San Diego)
27. Texas Tech University
28. Tufts University
29. University of Virginia
30. University of Washington
31. Western Michigan University
32. Worcester Polytechnic Institute

PER papers published in the *American Journal of Physics* since 1972

The following papers have a significant or primary focus on collection and analysis of data regarding thinking or performance of students or instructors, or a focus on the methodologies or models appropriate to collecting and analyzing data of this type. Underlined titles are electronically linked to the AJP-online reference. This list is not intended to be comprehensive.

1972

Roger L. Ptak and Ronald E. Stoner, "A physics survey course with a large entropy," Am. J. Phys. **40**, 611 (1972).

J. H. Munsee, "An evaluation of a nonlecture technique in the teaching of physics," Am. J. Phys. **40**, 1119 (1972).

Owen T. Anderson and Robert A. Artman, "A self-paced, independent study, introductory physics sequence – description and evaluation," Am. J. Phys. **40**, 1737 (1972).

1973

Allan D. Franklin, "Physics for educationally disadvantaged students," Am. J. Phys. **41**, 3 (1973).

Sam M. Austin, "Student performance in a Keller-Plan course in introductory electricity and magnetism," Am. J. Phys. **41**, 12 (1973).

1974

C. S. Kalman, D. Kaufman, and R. Smith, "Introductory CAI dialogue in differential calculus for freshman physics," Am. J. Phys. **42**, 392 (1974).

Allan D. Franklin and Franz Mohling, "Physics for educationally disadvantaged students: II," Am. J. Phys. **42**, 553 (1974).

Frank A. Smith, Jr. and Joseph S. Schmuckler, "Transfer of learning and the conservation laws," Am. J. Phys. **42**, 830 (1974).

David E. Golden, Robert G. Fuller, and Donald D. Jensen, "Repeatable testing – A tool for learning physics," Am. J. Phys. **42**, 941 (1974).

Jacqueline Spears and Dean Zollman, "Orientation for the new teaching assistant – a laboratory based program," Am. J. Phys. **42**, 1062 (1974).

Michael Zeilik II, "A PSI astronomy course," Am. J. Phys. **42**, 1095 (1974).

J. L. Aubel, "Competency-based general physics instruction with a one man staff," Am. J. Phys. **42**, 1101 (1974).

1975

Jacqueline D. Spears and C. E. Hathaway, "Student attitudes toward science and society," Am. J. Phys. **43**, 343 (1975).

1976

Mary Diederich Ott, "Evaluation of methods of instruction and procedures for assigning students to methods," Am. J. Phys. **44**, 12 (1976).

F. Reif, Jill H. Larkin, and George C. Brackett, "Teaching general learning and problem-solving skills," Am. J. Phys. **44**, 212 (1976).

John W. Renner, "Significant physics content and intellectual development---cognitive development as a result of interacting with physics content," Am. J. Phys. 44, 218 (1976).

Elmer G. Redford, "Attitudes toward physics in the high school curriculum," Am. J. Phys. **44**, 337 (1976).

Lillian C. McDermott, "Teacher education and the implementation of elementary science curricula,"" Am. J. Phys. **44**, 434 (1976).

Donald F. Kirwan and Jack Willis, "Co-teaching experiment at the University of Rhode Island," Am. J. Phys. **44**, 651 (1976).

Ronald James Bieniek and Michael Zeilik II, "Follow-up study of a PSI astronomy course," Am. J. Phys. **44**, 695 (1976).

Thomas C. Taveggia, "Personalized instruction: A summary of comparative research, 1967--1974," Am. J. Phys. **44**, 1028 (1976).

1977

Albert B. Smith, Gray Ward, and Joseph S. Rosenshein, "Improving instruction by measuring teacher discussion skills," Am. J. Phys. **45**, 83 (1977).

H. T. Hudson and W. R. McIntire, "Correlation between mathematical skills and success in physics," Am. J. Phys. **45**, 470 (1977).

Graeme D. Putt, "Testing the mastery concept of self-paced learning in physics," Am. J. Phys. **45**, 472 (1977).

S. S. Jaswal, "Results of a two-trial examination system in a general physics course," Am. J. Phys. **45**, 575 (1977).

Robert Gerson and Russell A. Primrose, "Results of a remedial laboratory program based on a Piaget model for engineering and science freshmen," Am. J. Phys. **45**, 649 (1977).

George Barnes, "Scores on a Piaget-type questionnaire versus semester grades for lower-division college physics students," Am. J. Phys. **45**, 841 (1977).

John W. Renner and William C. Paske, "Comparing two forms of instruction in college physics," Am. J. Phys. **45**, 851 (1977).

G. Will Pfeiffenberger and Christopher C. Modu, "A validity study of the multiple-choice component of the advanced placement physics C examination," Am. J. Phys. **45**, 1066 (1977).

1978

Larry Hedges, "Personalized introductory courses: A longitudinal study," Am. J. Phys. **46**, 207 (1978).

Judith D. Aubrecht, "Teacher effectiveness: Self-determined change," Am. J. Phys. **46**, 324 (1978).

Richard R. Boedeker, "Report on the use of follow-up orals in undergraduate and graduate physics teaching," Am. J. Phys. **46**, 643 (1978).

George Barnes and George Bruce Barnes, "Students' scores on Piaget-type questionnaires before and after taking one semester of college physics," Am. J. Phys. **46**, 807 (1978).

H. Daniel Cohen, Donald F. Hillman, and Russell M. Agne, "Cognitive level and college physics achievement," Am. J. Phys. **46**, 1026 (1978).

E. L. McFarland, F. R. Hallett, and J. L. Hunt, "The development of a personalized-system-of-instruction introductory physics course for life-science students," Am. J. Phys. **46**, 1095 (1978).

1979

Richard Vawter, "Entropy state of a multiple choice examination and the evaluation of understanding," Am. J. Phys. **47**, 320 (1979).

Dov Liberman and H. T. Hudson, "Correlation between logical abilities and success in physics," Am. J. Phys. **47**, 784 (1979).

1980

Ted E. Senator, "Goals, aspirations, and attitudes of MIT undergraduate physics students," Am. J. Phys. **48**, 152 (1980).

George Barnes and George Bruce Barnes, "Students' scores on Piaget-type questionnaires before and after taking two semesters of college physics," Am. J. Phys. **48**, 774 (1980).

David E. Trowbridge and Lillian C. McDermott, "Investigation of student understanding of the concept of velocity in one dimension," Am. J. Phys. **48**, 1020 (1980).

Audrey B. Champagne, Leopold E. Klopfer, and John H. Anderson, "Factors influencing the learning of classical mechanics," Am. J. Phys. **48**, 1074 (1980).

1981

David P. Maloney, "Piagetian testing," Am. J. Phys. **49**, 190 (1981).

David E. Trowbridge and Lillian C. McDermott, "Investigation of student understanding of the concept of acceleration in one dimension," Am. J. Phys. **49**, 242 (1981).

David P. Maloney, "Comparative reasoning abilities of college students," Am. J. Phys. **49**, 784 (1981).

Michael Zeilik II, "Flexible, mastery-oriented astrophysics sequence," Am. J. Phys. **49**, 827 (1981).

1982

John Clement, "Students' preconceptions in introductory mechanics," Am. J. Phys. **50**, 66 (1982).

P. C. Peters, "Even honors students have conceptual difficulties with physics," Am. J. Phys. **50**, 501 (1982).

Kenneth W. Wood, "An investigation of CAI teaching methods in an electronics course," Am. J. Phys. **50**, 683 (1982).

H. T. Hudson and Dov Liberman, "The combined effect of mathematics skills and formal operational reasoning on student performance in the general physics course," Am. J. Phys. **50**, 1117 (1982).

1983

Robert J. Whitaker, "Aristotle is not dead: Student understanding of trajectory motion," Am. J. Phys. **51**, 352 (1983).

R. Cohen, B. Eylon, and U. Ganiel, "Potential difference and current in simple electric circuits: A study of students' concepts," Am. J. Phys. **51**, 407 (1983).

Nathan O. Folland, Robert R. Marchini, Charles R. Rhyner, and Michael Zeilik, "Report on using TIPS (Teaching Information Processing System) in teaching physics and astronomy," Am. J. Phys. **51**, 446 (1983).

L. M. Jones, Dennis Kane, Bruce Arne Sherwood, and R. A. Avner, "A final-exam comparison involving computer-based instruction," Am. J. Phys. **51**, 533 (1983).

Gordon J. Aubrecht II and Judith D. Aubrecht, "Constructing objective tests," Am. J. Phys. **51**, 613 (1983).

1984

[--]

1985

Dale D. Long and Jo C. Bedard, "Evaluation of a discussion technique used for both classroom instruction and grade assignment," Am. J. Phys. **53**, 401 (1985).

Jean Lythcott, " "Aristotelian" was given as the answer, but what was the question?," Am. J. Phys. **53**, 428 (1985).

L. Viennot, "Analyzing students' reasoning: Tendencies in interpretation,"Am. J. Phys. **53**, 432 (1985).

Peter W. Hewson, "Diagnosis and remediation of an alternative conception of velocity using a microcomputer program," Am. J. Phys. **53**, 684 (1985).

W. T. Griffith, "Factors affecting performance in introductory physics courses," Am. J. Phys. **53**, 839 (1985).

Ibrahim Abou Halloun and David Hestenes, "The initial knowledge state of college physics students," Am. J. Phys. **53**, 1043 (1985).

Ibrahim Abou Halloun and David Hestenes, "Common sense concepts about motion," Am. J. Phys. **53**, 1056 (1985).

1986

Dale D. Long, Gerald W. McLaughlin, and Allan M. Bloom, "The influence of physics laboratories on student performance in a lecture course," Am. J. Phys. **54**, 122 (1986).

H. A. Buckmaster, "Evaluating the effectiveness of an elementary undergraduate laboratory," Am. J. Phys. **54**, 702 (1986).

1987

Fred M. Goldberg and Lillian C. McDermott, "An investigation of student understanding of the real image formed by a converging lens or concave mirror," Am. J. Phys. **55**, 108 (1987).

Mark L. Rosenquist and Lillian C. McDermott, "A conceptual approach to teaching kinematics," Am. J. Phys. **55**, 407 (1987).

David Hestenes, "Toward a modeling theory of physics instruction," Am. J. Phys. **55**, 440 (1987).

Ibrahim Abou Halloun and David Hestenes, "Modeling instruction in mechanics," Am. J. Phys. **55**, 455 (1987).

Monica G. M. Ferguson-Hessler and Ton de Jong, "On the quality of knowledge in the field of electricity and magnetism," Am. J. Phys. **55**, 492 (1987).

Lillian C. McDermott, Mark L. Rosenquist, and Emily H. van Zee, "Student difficulties in connecting graphs and physics: Examples from kinematics," Am. J. Phys. **55**, 503 (1987).

Richard F. Gunstone, "Student understanding in mechanics: A large population survey," Am. J. Phys. **55**, 691 (1987).

Ronald A. Lawson and Lillian C. McDermott, "Student understanding of the work-energy and impulse-momentum theorems," Am. J. Phys. **55**, 811 (1987).

R. R. Hake, "Promoting student crossover to the Newtonian world," Am. J. Phys. **55**, 878 (1987).

1988

Renate C. Lippert, Patricia A. Heller, Roger S. Jones, and Russell K. Hobbie, "An evaluation of classroom teaching practices one year after a workshop for high-school physics teachers," Am. J. Phys. **56**, 505 (1988).

Sheila Tobias and R. R. Hake, "Professors as physics students: What can they teach us?," Am. J. Phys. **56**, 786 (1988).

1989

R. B. Hicks and H. Laue, "A computer-assisted approach to learning physics concepts," Am. J. Phys. **57**, 807 (1989).

1990

Lillian C. McDermott, "Research and computer-based instruction: Opportunity for interaction," Am. J. Phys. **58**, 452 (1990).

Lillian C. McDermott, "A perspective on teacher preparation in physics and other sciences: The need for special science courses for teachers," Am. J. Phys. **58**, 734 (1990).

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

1991

Mary Beth Ruskai, "Guest Comment: Are there innate cognitive gender differences? Some comments on the evidence in response to a letter from M. Levin," Am. J. Phys. **59**, 11 (1991).

Jack Lochhead, "Guest Comment: Why is good teaching so clearly confusing? A tale of two theories," Am. J. Phys. **59**, 969 (1991).

Lillian Christie McDermott, "Millikan Lecture 1990: What we teach and what is learned---Closing the gap," Am. J. Phys. **59**, 301 (1991).

J. W. Halley, Azin Adjoudani, Patricia Heller, and James S. Terwilliger, "The Graduate Record Examination as an indicator of learning of the curriculum taught to physics majors in US institutions," Am. J. Phys. **59**, 403 (1991).

Alan Van Heuvelen, "Learning to think like a physicist: A review of research-based instructional strategies," Am. J. Phys. **59**, 891 (1991).

Alan Van Heuvelen , "Overview, Case Study Physics," Am. J. Phys. **59**, 898 (1991).

1992

J. Bowden, G. Dall'Alba, E. Martin, D. Laurillard, F. Marton, G. Masters, P. Ramsden, A. Stephanou, and E. Walsh, "Displacement, velocity, and frames of reference: Phenomenographic studies of students' understanding and some implications for teaching and assessment," Am. J. Phys. **60**, 262 (1992).

Ernst Breitenberger, "The mathematical knowledge of physics graduates: Primary data and conclusions," Am. J. Phys. **60**, 318 (1992).

Patricia Heller, Ronald Keith, and Scott Anderson, "Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving," Am. J. Phys. **60**, 627 (1992).

Patricia Heller and Mark Hollabaugh, "Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups," Am. J. Phys. **60**, 637 (1992).

Lillian C. McDermott and Peter S. Shaffer, "Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding," Am. J. Phys. **60**, 994 (1992); Erratum: Am. J. Phys. **61**, 81 (1993).

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

1993

Lillian C. McDermott, "Guest Comment: How we teach and how students learn---A mismatch?," Am. J. Phys. **61**, 295 (1993).

Robert G. Fuller, "Millikan Lecture 1992: Hypermedia and the knowing of physics: Standing upon the shoulders of giants," Am. J. Phys. **61**, 300 (1993).

S. Törnkvist, K.-A. Pettersson, and G. Tranströmer, "Confusion by representation: On student's comprehension of the electric field concept," Am. J. Phys. **61**, 335 (1993).

1994

Lillian C. McDermott, Peter S. Shaffer, and Mark D. Somers, "Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwood's machine," Am. J. Phys. **62**, 46 (1994).

Beth Thacker, Eunsook Kim, Kelvin Trefz, and Suzanne M. Lea, "Comparing problem solving performance of physics students in inquiry-based and traditional introductory physics courses," Am. J. Phys. **62**, 627 (1994).

G. Rebmann and L. Viennot, "Teaching algebraic coding: Stakes, difficulties, and suggestions," Am. J. Phys. **62**, 723 (1994).

Robert J. Beichner, "Testing student interpretation of kinematics graphs," Am. J. Phys. **62**, 750 (1994).

Edward F. Redish, "Implications of cognitive studies for teaching physics," Am. J. Phys. **62**, 796 (1994).

Lorella M. Jones and Dennis J. Kane, "Student evaluation of computer-based instruction in a large university mechanics course," Am. J. Phys. **62**, 832 (1994).

Roger Nanes and John W. Jewett, Jr., "Southern California Area Modern Physics Institute (SCAMPI): A model enhancement program in modern physics for high school teachers," Am. J. Phys. **62**, 1020 (1994).

S. Rainson, G. Tranströmer, and L. Viennot, "Students' understanding of superposition of electric fields," Am. J. Phys. **62**, 1026 (1994).

1995

Frederick Reif, "Millikan Lecture 1994: Understanding and teaching important scientific thought processes," Am. J. Phys. **63**, 17 (1995).

Elaine Seymour, "Guest Comment: Why undergraduates leave the sciences," Am. J. Phys. **63**, 199 (1995).

R. A. Lewis, "Video introductions to laboratory: Students positive, grades unchanged," Am. J. Phys. **63**, 468 (1995).

Malcolm Wells, David Hestenes, and Gregg Swackhamer, "A modeling method for high school physics instruction," Am. J. Phys. **63**, 606 (1995).

Mohammad Samiullah, "Effect of in-class student--student interaction on the learning of physics in a college physics course," Am. J. Phys. **63**, 944 (1995).

Fred Goldberg and Sharon Bendall, "Making the invisible visible: A teaching/learning environment that builds on a new view of the physics learner," Am. J. Phys. **63**, 978 (1995).

Jeffry V. Mallow, "Students' confidence and teachers' styles: A binational comparison," Am. J. Phys. **63**, 1007 (1995).

1996

R. Di Stefano, "Preliminary IUPP results: Student reactions to in-class demonstrations and to the presentation of coherent themes," Am. J. Phys. **64**, 58 (1996).

R. Di Stefano, "The IUPP evaluation: What we were trying to learn and how we were trying to learn it," Am. J. Phys. **64**, 49 (1996).

Dean Zollman, "Millikan Lecture 1995: Do they just sit there? Reflections on helping students learn physics," Am. J. Phys. **64**, 114 (1996).

Diane J. Grayson and Lillian C. McDermott, "Use of the computer for research on student thinking in physics," Am. J. Phys. **64**, 557 (1996).

Frederick Reif, "Guest Comment: Standards and measurements in physics---Why not in physics education?," Am. J. Phys. **64**, 687 (1996).

Robert J. Beichner, "The impact of video motion analysis on kinematics graph interpretation skills," Am. J. Phys. **64**, 1272 (1996).

David Hammer, "More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research," Am. J. Phys. **64**, 1316 (1996).

Richard N. Steinberg, Graham E. Oberem, and Lillian C. McDermott, "Development of a computer-based tutorial on the photoelectric effect," Am. J. Phys. **64**, 1370 (1996).

William J. Leonard, Robert J. Dufresne, and Jose P. Mestre, "Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems," Am. J. Phys. **64**, 1495 (1996).

1997

P. W. Laws, "Millikan Lecture 1996: Promoting active learning based on physics education research in introductory physics courses," Am. J. Phys. **65**, 14 (1997).

Edward F. Redish, Jeffery M. Saul, and Richard N. Steinberg, "On the effectiveness of active-engagement microcomputer-based laboratories," Am. J. Phys. **65**, 45 (1997).

Ronald Gautreau and Lisa Novemsky, "Concepts first---A small group approach to physics learning," Am. J. Phys. **65**, 418 (1997).

Esther Bagno and Bat-Sheva Eylon, "From problem solving to a knowledge structure: An example from the domain of electromagnetism," Am. J. Phys. **65**, 726 (1997).

Jrène Rahm and Paul Charbonneau, "Probing stereotypes through students' drawings of scientists," Am. J. Phys. **65**, 774 (1997).

Michael Zeilik, Candace Schau, Nancy Mattern, Shannon Hall, Kathleen W. Teague, and Walter Bisard, "Conceptual astronomy: A novel model for teaching postsecondary science courses," Am. J. Phys. **65**, 987 (1997).

1998

Richard R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," Am. J. Phys. **66**, 64 (1998).

Tara O'Brien Pride, Stamatis Vokos, and Lillian C. McDermott, "The challenge of matching learning assessments to teaching goals: An example from the work-energy and impulse-momentum theorems," Am. J. Phys. **66**, 147 (1998).

Edward F. Redish, Jeffery M. Saul, and Richard N. Steinberg, "Student expectations in introductory physics," Am. J. Phys. **66**, 212 (1998).

Ronald K. Thornton and David R. Sokoloff, "Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula," Am. J. Phys. **66**, 338 (1998).

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

Karen Wosilait, Paula R. L. Heron, Peter S. Shaffer, and Lillian C. McDermott, "Development and assessment of a research-based tutorial on light and shadow," Am. J. Phys. **66**, 906 (1998).

1999

Bradley S. Ambrose, Peter S. Shaffer, Richard N. Steinberg, and Lillian C. McDermott, "An investigation of student understanding of single-slit diffraction and double-slit interference," Am. J. Phys. **67**, 146 (1999).

Edward F. Redish, "Millikan Lecture 1998: Building a Science of Teaching Physics," Am. J. Phys. **67**, 562 (1999).

L. A. Collins, N. H. Magee, H. C. Bryant, and M. Zeilik, "A conference experience for undergraduates," Am. J. Phys. **67**, 685 (1999).

Lillian C. McDermott and Edward F. Redish, "Resource Letter: PER-1: Physics Education Research," Am. J. Phys. **67**, 755 (1999).

M. J. Marr, E. W. Thomas, M. R. Benne, A. Thomas, and R. M. Hume, "Development of instructional systems for teaching an electricity and magnetism course for engineers," Am. J. Phys. **67**, 789 (1999).

E. Etkina, K. Gibbons, B. L. Holton, and G. K. Horton, "Lessons learned: A case study of an integrated way of teaching introductory physics to at-risk students at Rutgers University," Am. J. Phys. **67**, 810 (1999).

Frederick Reif and Lisa A. Scott, "Teaching scientific thinking skills: Students and computers coaching each other," Am. J. Phys. **67**, 819 (1999).

Bradley S. Ambrose, Paula R. L. Heron, Stamatis Vokos, and Lillian C. McDermott, "Student understanding of light as an electromagnetic wave: Relating the formalism to physical phenomena," Am. J. Phys. **67**, 891 (1999).

Michael Zeilik, Candace Schau, and Nancy Mattern, "Conceptual astronomy. II. Replicating conceptual gains, probing attitude changes across three semesters," Am. J. Phys. **67**, 923 (1999).

Karen Wosilait, Paula R. L. Heron, Peter S. Shaffer, and Lillian C. McDermott, "Addressing student difficulties in applying a wave model to the interference and diffraction of light," Am. J. Phys. **67**, S5 (1999).

R. Beichner, L. Bernold, E. Burniston, P. Dail, R. Felder, J. Gastineau, M. Gjertsen, and J. Risley, "Case study of the physics component of an integrated curriculum," Am. J. Phys. **67**, S16 (1999).

Beth Ann Thacker, Uri Ganiel, and Donald Boys, "Macroscopic phenomena and microscopic processes: Student understanding of transients in direct current electric circuits," Am. J. Phys. **67**, S25 (1999).

Priscilla W. Laws, Pamela J. Rosborough, and Frances J. Poodry, "Women's responses to an activity-based introductory physics program," Am. J. Phys. **67**, S32 (1999).

Karen Cummings, Jeffrey Marx, Ronald Thornton, and Dennis Kuhl, "Evaluating innovation in studio physics," Am. J. Phys. **67**, S38 (1999).

Calvin S. Kalman, Stanley Morris, Christopher Cottin, and Robert Gordon, "Promoting conceptual change using collaborative groups in quantitative gateway courses," Am. J. Phys. **67**, S45 (1999).

Andrew Elby, "Another reason that physics students learn by rote," Am. J. Phys. **67**, S52 (1999).

Randal Harrington, "Discovering the reasoning behind the words: An example from electrostatics," Am. J. Phys. **67**, S58 (1999).

2000

Igal Galili and Amnon Hazan, "The influence of an historically oriented course on students' content knowledge in optics evaluated by means of facets-schemes analysis," Am. J. Phys. **68**, S3 (2000).

Esther Bagno, Bat-Sheva Eylon, and Uri Ganiel, "From fragmented knowledge to a knowledge structure: Linking the domains of mechanics and electromagnetism," Am. J. Phys. **68**, S16 (2000).

Jill A. Marshall and James T. Dorward, "Inquiry experiences as a lecture supplement for preservice elementary teachers and general education students," Am. J. Phys. **68**, S27 (2000).

Richard N. Steinberg, "Computers in teaching science: To simulate or not to simulate?," Am. J. Phys. **68**, S37 (2000).

Stamatis Vokos, Peter S. Shaffer, Bradley S. Ambrose, and Lillian C. McDermott, "Student understanding of the wave nature of matter: Diffraction and interference of particles," Am. J. Phys. **68**, S42 (2000).

David Hammer, "Student resources for learning introductory physics," Am. J. Phys. **68**, S52 (2000).

David S. Abbott, Jeffery M. Saul, George W. Parker, and Robert J. Beichner, "Can one lab make a difference?," Am. J. Phys. **68**, S60 (2000).

2001

Alan Van Heuvelen and Xueli Zou, "Multiple representations of work--energy processes," Am. J. Phys. **69**, 184 (2001).

Donald F. Holcomb, "Should I pay attention to the output from physics education R&D?," Am. J. Phys. **69**, 407 (2001).

Chandralekha Singh, "Student understanding of quantum mechanics," Am. J. Phys. **69**, 885 (2001).

Catherine H. Crouch and Eric Mazur, "Peer Instruction: Ten years of experience and results," Am. J. Phys. **69**, 970 (2001).

Lillian Christie McDermott, "Oersted Medal Lecture 2001: ``Physics Education Research---The Key to Student Learning''," Am. J. Phys. **69**, 1127 (2001).

Alan Van Heuvelen, "Millikan Lecture 1999: The Workplace, Student Minds, and Physics Learning Systems," Am. J. Phys. **69**, 1139 (2001).

Molly Johnson, "Facilitating high quality student practice in introductory physics," Am. J. Phys. **69**, S2 (2001).

David P. Maloney, Thomas L. O'Kuma, Curtis J. Hieggelke, and Alan Van Heuvelen, "Surveying students' conceptual knowledge of electricity and magnetism," Am. J. Phys. **69**, S12 (2001).

Rachel E. Scherr, Peter S. Shaffer, and Stamatis Vokos, "Student understanding of time in special relativity: Simultaneity and reference frames," Am. J. Phys. **69**, S24 (2001).

P. Colin and L. Viennot, "Using two models in optics: Students' difficulties and suggestions for teaching," Am. J. Phys. **69**, S36 (2001).

Lei Bao and Edward F. Redish, "Concentration analysis: A quantitative assessment of student states," Am. J. Phys. **69**, S45 (2001).

Andrew Elby, "Helping physics students learn how to learn," Am. J. Phys. **69**, S54 (2001).

2002

Michael E. Loverude, Christian H. Kautz, and Paula R. L. Heron, "Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas," Am. J. Phys. **70**, 137 (2002).

Rainer Müller and Hartmut Wiesner, "Teaching quantum mechanics on an introductory level," Am. J. Phys. **70**, 200 (2002).

Lei Bao and Edward F. Redish, "Understanding probabilistic interpretations of physical systems: A prerequisite to learning quantum physics," Am. J. Phys. **70**, 210 (2002).

Michael C. Wittmann, Richard N. Steinberg, and Edward F. Redish, "Investigating student understanding of quantum physics: Spontaneous models of conductivity," Am. J. Phys. **70**, 218 (2002).

E. Cataloglu and R. W. Robinett, "Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career," Am. J. Phys. **70**, 238 (2002).

David E. Meltzer and Kandiah Manivannan, "Transforming the lecture-hall environment: The fully interactive physics lecture," Am. J. Phys. **70**, 639 (2002).

Ian D. Beatty and William J. Gerace, "Probing physics students' conceptual knowledge structures through term association," Am. J. Phys. **70**, 750 (2002).

Eunsook Kim and Sung-Jae Pak, "Students do not overcome conceptual difficulties after solving 1000 traditional problems," Am. J. Phys. **70**, 759 (2002).

Lei Bao, Kirsten Hogg, and Dean Zollman, "Model analysis of fine structures of student models: An example with Newton's third law," Am. J. Phys. **70**, 766 (2002).

David J. Van Domelen and Alan Van Heuvelen, "The effects of a concept-construction lab course on FCI performance," Am. J. Phys. **70**, 779 (2002).

Richard Hake**, "**Comment on ``How do we know if we are doing a good job in physics teaching?," by Robert Ehrlich [Am. J. Phys. 70 (1), 24--29 (2002)]**,"** Am. J. Phys. **70**, 1058-1059 (2002).

Chandralekha Singh, "When physical intuition fails," Am. J. Phys. **70**, 1103 (2002).

Rachel E. Scherr, Peter S. Shaffer, and Stamatis Vokos, "The challenge of changing deeply held student beliefs about the relativity of simultaneityAm. J. Phys. **70**, 1238 (2002).

David B. May and Eugenia Etkina, "College physics students' epistemological self-reflection and its relationship to conceptual learning," Am. J. Phys. **70**, 1249 (2002).

David E. Meltzer, "The relationship between mathematics preparation and conceptual learning gains in physics: A possible ``hidden variable" in diagnostic pretest scores," Am. J. Phys. **70**, 1259 (2002).

2003

Beth Ann Thacker, "A study of the nature of students' models of microscopic processes in the context of modern physics experiments," Am. J. Phys. **71**, 599 (2003).

Chandralekha Singh and David Rosengrant, "Multiple-choice test of energy and momentum concepts," Am. J. Phys. **71**, 607 (2003).

Zacharias Zacharia and O. Roger Anderson, "The effects of an interactive computer-based simulation prior to performing a laboratory inquiry-based experiment on students' conceptual understanding of physics," Am. J. Phys. **71**, 618 (2003).

Ngoc-Loan Nguyen and David E. Meltzer, "Initial understanding of vector concepts among students in introductory physics courses,"Am. J. Phys. **71**, 630 (2003).

I. Galili and Y. Lehavi, "The importance of weightlessness and tides in teaching gravitation," Am. J. Phys. **71**, 1127 (2003).

Michael E. Loverude, Christian H. Kautz, and Paula R. L. Heron, "Helping students develop an understanding of Archimedes' principle. I. Research on student understanding," Am. J. Phys. **71**, 1178 (2003).

P. R. L. Heron, M. E. Loverude, P. S. Shaffer, and L. C. McDermott, "Helping students develop an understanding of Archimedes' principle. II. Development of research-based instructional materials," Am. J. Phys. **71**, 1188 (2003).

2004

Paula Vetter Engelhardt and Robert J. Beichner, "Students' understanding of direct current resistive electrical circuits," Am. J. Phys. **72**, 98 (2004).

N. Sanjay Rebello and Dean A. Zollman, "The effect of distracters on student performance on the force concept inventory," Am. J. Phys. **72**, 116 (2004).

Charles Henderson, Edit Yerushalmi, Vince H. Kuo, Patricia Heller, and Kenneth Heller, "Grading student problem solutions: The challenge of sending a consistent message," Am. J. Phys. **72**, 164 (2004).

Bradley S. Ambrose, "Investigating student understanding in intermediate mechanics: Identifying the need for a tutorial approach to instruction," Am. J. Phys. **72**, 453 (2004).

Sergio Flores, Stephen E. Kanim, and Christian H. Kautz, "Student use of vectors in introductory mechanics," Am. J. Phys. **72**, 460 (2004).

Calvin S. Kalman, Shelley Rohar, and David Wells, "Enhancing conceptual change using argumentative essays," Am. J. Phys. **72**, 715 (2004).

Catherine H. Crouch, Adam P. Fagen, J. Paul Callan, and Eric Mazur, "Classroom demonstrations: Learning tools or entertainment?," Am. J. Phys. **72**, 835 (2004).

Diane J. Grayson, "Concept substitution: A teaching strategy for helping students disentangle related physics concepts," Am. J. Phys. **72**, 1126 (2004).

Leonardo Hsu, Eric Brewe, Thomas M. Foster, and Kathleen A. Harper, "Resource Letter RPS-1: Research in problem solving," Am. J. Phys. **72**, 1147 (2004).

Shelley Yeo, Robert Loss, Marjan Zadnik, Allan Harrison, and David Treagust, "What do students really learn from interactive multimedia? A physics case study," Am. J. Phys. **72**, 1351 (2004).

David E. Meltzer, "Investigation of students' reasoning regarding heat, work, and the first law of thermodynamics in an introductory calculus-based general physics course," Am. J. Phys. **72**, 1432 (2004).

K. Kelvin Chang, Beth Ann Thacker, Richard L. Cardenas, and Catherine Crouch, "Using an online homework system enhances students' learning of physics concepts in an introductory physics course," Am. J. Phys. **72**, 1447 (2004).

2005

Laura Lising and Andrew Elby, "The impact of epistemology on learning: A case study from introductory physics," Am. J. Phys. **73**, 372 (2005).

Paula R. L. Heron and David E. Meltzer, "Guest Editorial--The future of physics education research: Intellectual challenges and practical concerns," Am. J. Phys. **73**, 390 (2005).

Chandralekha Singh, "Impact of peer interaction on conceptual test performance," Am. J. Phys. **73**, 446 (2005).

Chance Hoellwarth, Matthew J. Moelter, and Randall D. Knight, "A direct comparison of conceptual learning and problem solving ability in traditional and studio style classrooms," Am. J. Phys. **73**, 459 (2005).

David E. Meltzer, "Relation between students' problem-solving performance and representational format," Am. J. Phys. **73**, 463 (2005).

Luanna G. Ortiz, Paula R. L. Heron, and Peter S. Shaffer**, "**Student understanding of static equilibrium: Predicting and accounting for balancing," Am. J. Phys. **73**, 545 (2005).

Neville W. Reay, Lei Bao, Pengfei Li, Rasil Warnakulasooriya, and Gordon Baugh, "Toward the effective use of voting machines in physics lectures," Am. J. Phys. **73**, 554 (2005).

Janelle M. Bailey and Timothy F. Slater, "Resource Letter AER-1: Astronomy education research," Am. J. Phys. **73**, 677 (2005).

Rebecca Lippmann Kung, "Teaching the concepts of measurement: An example of a concept-based laboratory course," Am. J. Phys. **73**, 771 (2005).

Charles Henderson, "The challenges of instructional change under the best of circumstances: A case study of one college physics instructor," Am. J. Phys. **73**, 778 (2005).