

Chapter 9

Frontiers in Research in Physics Education

Jose Mestre, *University of Massachusetts-Amherst*
Bruce Sherwood, *Carnegie Mellon University*
Ron Thornton, *Tufts University*
Alan Van Heuvelen, *the Ohio State University*
Chair: James H. Stith, *American Institute of Physics*

Jim Stith

In this session, our panelists will talk about what they would like to do in the future. If they had no constraints on money, if they had no constraints on time, what would they really want to do? This topic emerged as we were thinking about the fact that this community is an emerging community. We ought to be thinking about the kind of research that we should be doing, that we want to do in the future, which will move us up to the next tier.

Jose Mestre

I will talk about two things—one will take a little longer than the other—that I think will need attention in physics education research.

The first one is assessment. Let me tell you what I don't mean by assessment so that you don't think in the traditional way. I don't mean traditional exams, multiple choice or problems—the kind given in typical classes. I don't mean evaluation of a curriculum or a program either—that's something else. I mean research in ways to use assessment for many purposes. I'll describe what some of those purposes are. Our group, in Massachusetts, is doing a very small slice of something with assessment, but there is a much wider range of things that need to be done, that no single group could possibly do by itself.

One thing I would recommend is developing assessments and diagnostics for students' conceptual understanding of Physics. As we know, traditional tests are notoriously poor for measuring what a student truly understands—the concepts we are trying to teach. They are perhaps good at having them exhibit procedures for solving problems. Doing well on such a test doesn't mean that a student understands the underlying concepts that go into solving the problems.

We heard yesterday from Arnold Arons that interviews are a great method for getting at student understanding, and many of us who have done clinical interviews with students have certainly been able to diagnose students' problems and misconceptions. You can think of us as experts or doctors who are diagnosing Physics illnesses if you like. However, I think I will not take long to convince you that interviews are a very inefficient way of dealing with diagnosing large groups of students. It just takes a lot of time.

We do have instruments. You might say the Force Concept Inventory is a good diagnostic tool, but I don't think that the FCI was designed to help us help particular students learn Physics better. It was designed for other purposes—perhaps to evaluate instruction rather than students.

I think we need research on developing inventive ways for assessing levels of conceptual understanding, or learning obstacles if you like. Those might consist of paper and pencil instruments. How do we design them so that they're getting at exactly what we want, but they look different from what we are used to?

Questions for use in collaborative lab group work. How do we design instruments for students working in groups—to either diagnose their own misconceptions or learning issues or help the teacher do that? We might think of developing computer-based diagnostics that go after students' specific misunderstandings so that it is tailored to pull out what the students don't know or what they do know.

Finally, you might want to think about questions to diagnose problems in whole classes. How does a teacher develop questions that work well in a large classroom setting?

Some of these kinds of items already exist, and many of them have been coming out of the expert-novice research that you heard a little bit about yesterday. You didn't hear the details, but the same kinds of differences that exist between experts and novices can be used to develop instruments to see if we can tweak out what novices know or don't know, how they organize their knowledge, how they categorize problems—according to principles or according to surface features. All these things can be done to ascertain whether students are understanding the concepts in physics in the ways that we as physicists appreciate.

Another whole area of assessment that I don't think very many people have focused on is research on teachers' use of assessment to inform their practice. Are teachers capable of tailoring instruction based on their knowledge of student understanding of physics? We teach in a certain way—many of us lecture, many of us don't—but think about a high school teacher. If teachers are armed with what the students know or don't know, how will they pitch their instruction to match the students level of understanding or misunderstanding? I don't think this has been looked at very much. And I think this involves a whole different style of teaching—it's not something you can accomplish just by lecturing a different way. It also requires a different style of assessment. How do you diagnose and how do you use that diagnosis to inform your teaching?

How does formative assessment automatically change experienced teachers teaching? Does it change it? Does it have no influence on it? If assessment changes teachers' teaching then we might want to use assessment to drive instruction. If you want to change instruction, maybe the best way to do it is to develop the right kinds of assessments.

One way of looking at this is that if you diagnose a major conceptual error in a student, what do you do with it? Do you ignore it or do you go after it? How do you use that information as a teacher? I don't think we know very much about that.

How do we incorporate what we know about assessment to teach pre-service teachers, so they will be armed with techniques for dealing with student misunderstandings and armed with ways to diagnose student misunderstandings? I don't think that is done very much in pre-service programs, and I think it needs to be done.

Finally, can teachers devise their own diagnostic assessments for use in a whole class? How do they do that? What forms do they take? How do they work? There is certainly more, but those are some different areas within the assessment arena that I think need attention.

The other separate area I think that we could spend some time looking at is the evaluation of research-based curricula and instructional approaches. I will discuss this in terms a specific example. Many of you are aware of Alan Van Heuvelen's ALPS (Active Learning Problem Sheets) materials. Most students who use those, if not all, learn quite a bit more, as measured by traditional assessments or the FCI or anything else. In the last two or three years there was a study by Gautreau and Novemsky (*American Journal of Physics*, 65, 418-428) at the New Jersey Institute of Technology showing tremendous performance differences between students that he taught using the ALPS approach and control classes taught by other professors using traditional approaches. What made the study interesting to me is that the students were "at risk" students. They were very under prepared minority students. All of us at the UMass PERG stood back and said "Well, what the heck happened here? These are excellent results." But when we read the article very carefully we were left with many unanswered questions.

Here are some examples of what I'm interested in finding out. What were the contributing factors? Was it totally Alan's curriculum that that made the difference? Was it time on task? We know from cognitive research that the more time you spend on something the better you get at it. So was it that students using the curriculum were spending a lot more time learning than students in the control classes? Was it motivation? Maybe Alan's materials motivate students to spend more time or to learn more. So that may have been a contributing factor. Was it self-efficacy? We heard about that yesterday morning from Gregg Schraw. Maybe the materials help students gain a lot more confidence so they're willing to try harder at learning physics. Was it the cooperative learning aspects? Was it teacher's ability to use the curriculum as it was designed to be used? All these issues are things that are not being investigated. All we see are reports that this works a lot better than that as measured by some assessments. That doesn't tell me a heck of a lot.

I recommend studies in the ethnographic tradition, where you follow students around and see what they do, how they spend their time, what it is that makes them do better in these kinds of curriculum. Alan's isn't the only one. There are other curricula that people might want to try this approach on. Why is this important? Well, if something like motivation or self-efficacy is very important for getting

these kinds of gains, we need to know that. You might design a curriculum that is excellent otherwise, but if the students aren't motivated or don't get self-confidence out of it, they're not going to learn very much.

These are two different areas where I think we need to focus some attention.

Bruce Sherwood

Physics education research has been mainly focused on trying to find out how we can teach the traditional curriculum better. There are exceptions, but that has been the dominant paradigm, even from one of the seminal events, which was the FCI measurement showing that people didn't learn the traditional curriculum in the traditional courses as traditionally taught. Along with that emphasis has been the not very much discussed assumption that we should try to move students from novice behavior and problem solving toward expert behavior and problem solving. This seems as American as apple pie; it's what we clearly want.

I suggest that we should start asking "What is it that we think students should be doing? What is worth doing?" In particular, if a thing is not worth doing then it is not worth doing well. And there could be elements in the curriculum which are there for historical reasons, unexamined as to their purpose and their value. We try very hard to get people to be experts at doing these things, even if such expertise is not worth much. There is a danger of improving buggy whips. Do we want to teach people to be really good at manipulating buggy whips and constructing buggy whips at a time when there aren't any buggies?

We should not just ask, "How should we teach for expertness or quasi-expertness?" We should at least ask the question, "What should we be teaching?" This is a different kind of research—it has much more of a synthesis flavor than an analysis flavor. It is difficult to interview a student and ask what should be taught, so it has to come from us. Jonathan Reichert, in an essay in the August 1998 issue of the *American Journal of Physics* (pp. 664-665) which I can interpret for my own purposes here, is suggesting that we somehow should make a better, stronger link with our physics research colleagues about what should be taught. They might have ideas that would inform our attempts and struggles to think about what should be taught. I'll give a particular example of something that is extremely central to introductory physics, a topic that all of us have worked on very hard. It seems absolutely crucial—the key, the centerpiece of everything—and yet I want to construct a scenario where we should even be questioning whether we should be teaching it.

It is the analysis of dynamics problems using free body analysis. This is quite central to the Newtonian synthesis, or can be considered to be so. But let me just question even that, as an exemplar of the kind of discussion we might be carrying on with ourselves and each other about what should be in the curriculum and why. We all know some of the reasons why it should be in the curriculum. Are there any reasons why it should not be in the curriculum? I think so. For example, it is very far away from most other physics and chemistry and sciences. It is true that there are some key aspects of the nature of motion of a system and flows across the system

versus changes within a system, but in fact if you look at the broad sweep of contemporary physics, free body analysis is not where it's at. It is just not a major piece that connects with other things. That means that for it to have value in a more general setting there will have to be transfer, and it will be rather far transfer. We know that even near transfer is very difficult to achieve. So one could ask whether this is the right thing to be spending a lot of time on, particularly when it is very difficult to teach to expertness.

In fact, there is an opportunity cost. If we spend the time and we do our studies carefully, we will see how we could teach this topic to the point where students could reliably do it. There's an opportunity cost because that time could have been spent on something else. What is that something else? That is something for us and our physics research APS colleagues to think about.

What are other kinds of things we can do? There are many options. But I cite this example since it is one of our most sacred cows. I hope it gives you the flavor of the type of question I am suggesting that we consider. It is at least worth asking whether or not that is where we want to put our money. It is a big investment to make and there is an opportunity cost associated with it.

As an example of new kinds of thinking about what should be in the curriculum, look at Tom Moore's new textbook *Six Ideas that Shaped Physics* (McGraw-Hill 1998). It is a really good re-thinking of what the canon could or should be now that it is almost the year 2000. And I think that that Tom has done us all a good service by being one of the people really out in front asking the question "What should we be teaching and why?"

One other thing. The head of our department recently commented that in an evaluation form there was yet another one of these student comments, "Physics is really dull and boring." I would guess that one of the things that convinced the student of that—a tool for inducing that belief—was in fact free body analysis in a situation where the student was hoping to hear something about contemporary physics.

So there are some problems there, some opportunity costs in our curriculum, and we ought to be having some discussion about it, but we haven't been having much discussion about it. Even IUPP, which said, "Let's get some modern flavor into it," wasn't really about what should we teach. It was mainly about how we could teach the regular material better.

A second (not unrelated) comment is that it would be appropriate for some of us to spend more of our time trying to measure not just performance on content but also something about student attitudes, student beliefs and the student construction of their epistemology. There are some good starts on this. We know about the interesting work of the Maryland Group. The Halloun/Hestenes measurement that students didn't learn physics from introductory physics was a very important galvanizing event for all of us. And perhaps equally galvanizing can be the recent Maryland measurement that students also think worse and worse about physics as a result of being hit around the head by it. Halloun is working in this area as well.

It's not that all of us should be studying that, but there might well turn out to be a connection with my first point, that what we teach has a big impact on how students construct their beliefs about learning physics and the nature of physics, that there is a strong interaction between the attitude issues and the what we teach issues. We need to work on both of those if we are to get into a situation where, as is the case with Workshop Physics, we will actually see some attitudes improving as a result of studying physics.

This could be absolutely crucial and I'm really glad to see that that some of this work has started. We need more of it. Tom Foster, our post doc, has apprised us of some interesting work in the science education area, which has been looking at these things longer than we've been looking at them. We have things to learn from them.

In conclusion, I say we ought to be thinking what we should be teaching, and we should be looking at attitudinal issues and belief issues and not just content issues. Our own group (Ruth Chabay, Tom Foster and I) is in fact working in both those areas of trying to develop introductory curriculum and thinking about what should be in that curriculum. And then, a very challenging thing, which Tom Foster will figure out, is how to measure whether we're making some impact by changing the curriculum.

Ron Thornton

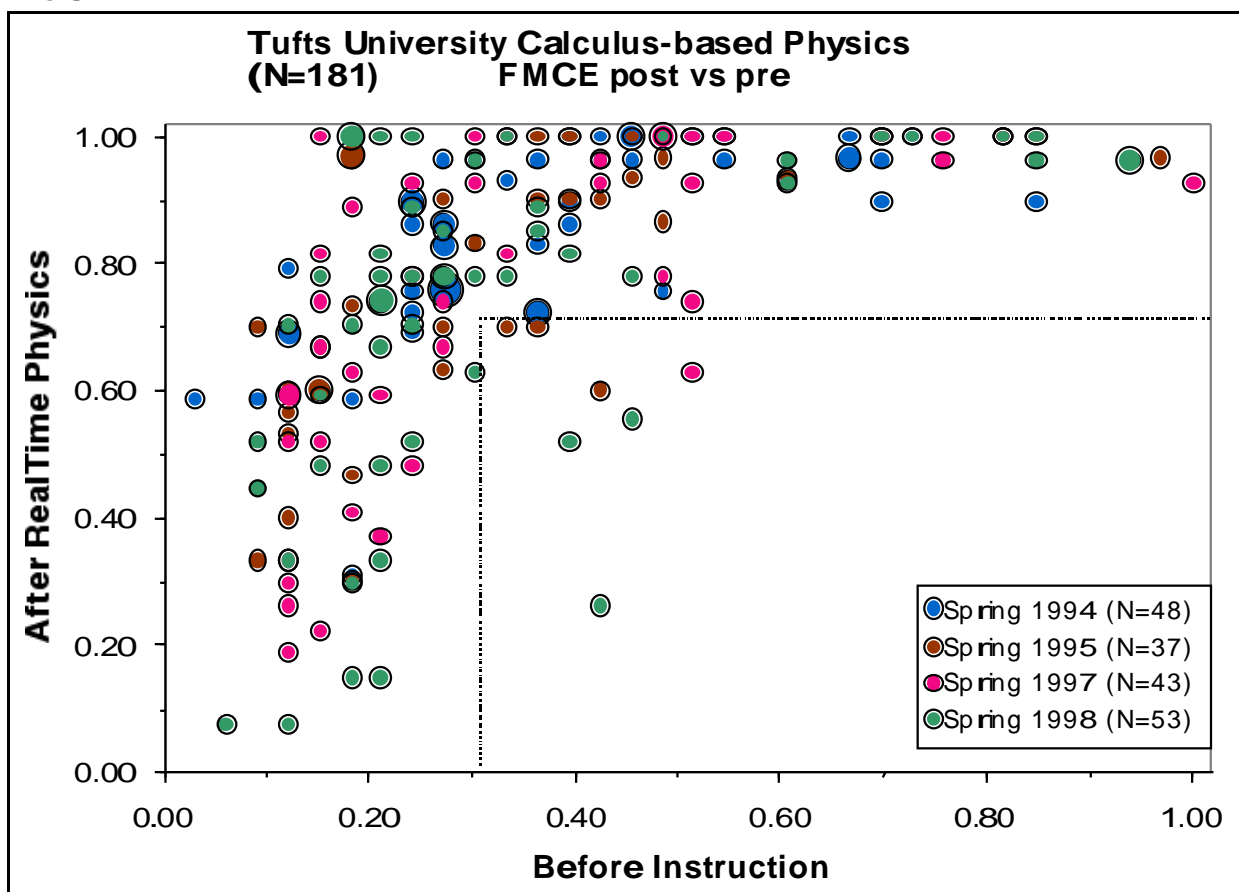
I thought I would give you a specific example to focus the discussion. What I would like to talk about is conceptual and the interpersonal predictors of physics learning. Wouldn't it be nice if a teacher could predict which students were at risk from a conceptual evaluation or from an observation of personal behavior? Can you do it? Some people told me they could, but none of them do research in physics education. The second part is, after finding these students at risk, knowing what do to help the students learn. Other people have mentioned these things. Let me describe a little bit of specific preliminary research in order to focus a discussion on this. Maybe you will see all of the holes and flaws, and maybe also how hard this is to do.

I have collaborators: Allan Risley and Dennis Kuhl. I will show you just a couple things. I want to show you a few results from a conceptual evaluation. I want to tell you a few results from video study—hours and hours and hours of video study—of students, linked to the conceptual evaluation, and I'm not going to spend very much time. If you're not familiar with some of this stuff, this talk won't help set the context very well.

The research methods consist of correlated short answer or evaluation. We do interviewing, we have classroom and laboratory observation and personal interviews. The educational setting of the particular stuff that I'm going to show is Tufts University, and this happens to be off-semester students in a calculus-based class. They are generally students that have had some problems somewhere, so they're not in the main sequence. The problems could be positive but generally they are negative.

In what I will show you as an example (see Figure 1), the students use RealTime Physics for mechanics in the laboratory and experience traditional lectures. They worked in groups, and this is important. The groups were almost always three members—occasionally two or four. They were self-assigned groups, and there was very little rotation among groups except in massive melt down. I am not recommending these things. I am telling you what happened. The professors or the TAs encouraged task rotation to some extent. Most of the groups were successful—that is, most of the people learned in this situation. I will tell you how to measure that and then you can decide whether you believe it or not, but basically this works out very well. Often up to 90% of the students will learn the basic concepts of linear dynamics and kinematics and so on.

FIGURE 1



In the group that I picked to show you, somewhat less than 90% of the students learn Newton's Laws. This lower than usual result is useful for our research, although I'm sure it isn't useful to the students. We used the *Force and Motion Conceptual Evaluation*¹ to evaluate conceptual learning. One of the first things I want to say is how you can make a prediction of who may have trouble learning. We've observed something we call a conceptual threshold effect. We don't know exactly why it is here. Let me explain it to you.

The position along the x-axis shows individual student scores, in percent, on the Force and Motion Conceptual Evaluation (FMCE) before all instruction. Many

students are only getting 20% of the questions right. In spite of many low scores, a few students up here (the right side of the graph) were getting 100% correct before instruction.

But mostly, as you can see, there's a large number of students down here (left side of the graph) below 30% to begin with. Now if we look on the y -axis, we can see what happened after instruction. (Students near the top learned a great deal.) If instruction produced no effect, you would have everyone on a 45-degree line. This is the approximate result with traditional instruction. If students knew less after instruction they would be below this line (essentially one student here). Now what actually happens with the RealTime physics instruction is what you see in the graph.

The shape you see shows that essentially everybody who started above 30% learned. There are a few exceptions but it is generally true. We call this the "threshold effect."

Below 30% on the pretest you can't predict what will happen. You notice the post results are spread out, but what's really interesting to me is that many of these people who started so dead low, actually learned very well. And yet we have these people down here (lower left-hand corner of the graph) that didn't learn well at all.

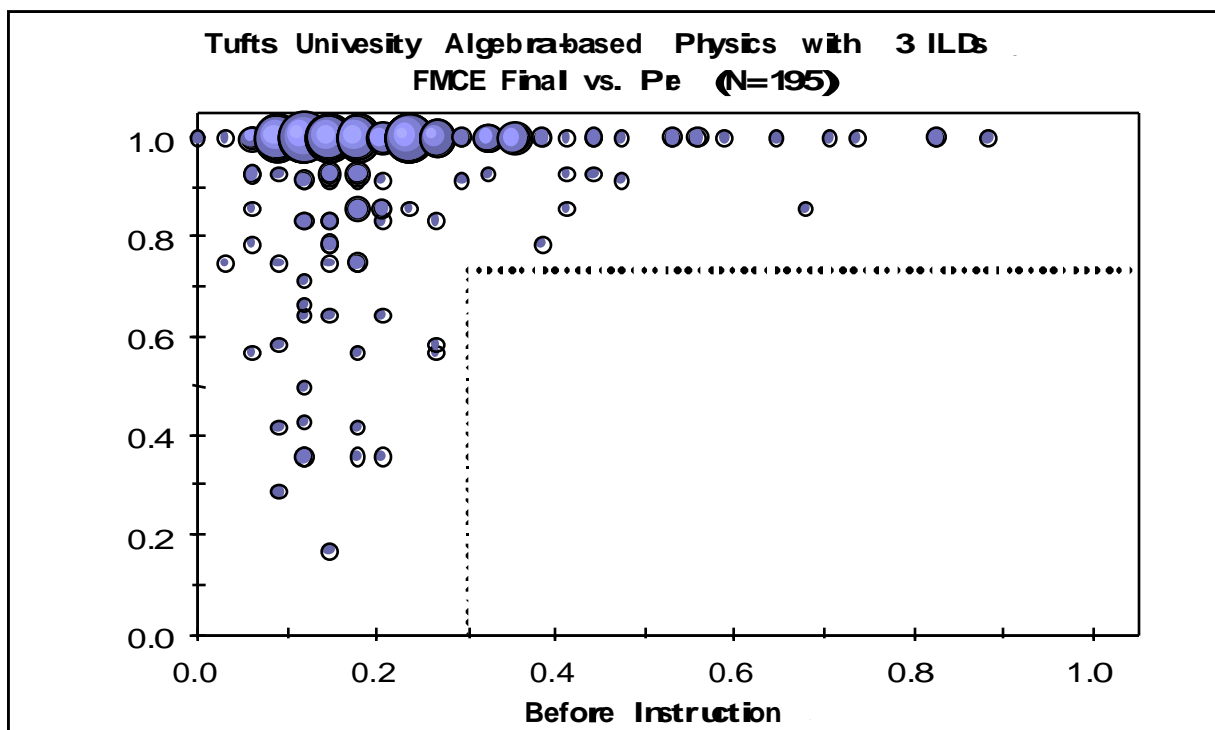
Now how can we account for this? Do they have signs on their foreheads or do they act in particular ways? Keep in mind that these results aren't as good as some of our results since only about 80 to 85% learn the first and second laws and 75 to 80% are able to answer the third law questions.

But anyway, let me just summarize this. Essentially all students who don't learn are below threshold on the pretest and yet some of those students learn very well. Can we tell the difference by observing them? We have hundreds of hours of videotape of these students working. It turns out there are some results and they appear to be reasonable. This is preliminary. We're just looking for something that we can identify ahead of time. Then we will try this in another environment to check.

If a low starter is an initiator in the group he or she will finish high. An initiator is pretty easy to understand—somebody who sort of moves things along. It turns out they can initiate. We're not measuring any skill they have with understanding anything. We have just noticed by careful observation and counting that they are initiators. Initiators are guaranteed to finish high.

If a low starter is not an initiator, it appears that he or she will finish low. This result is a little less sure, but it looks like that may be the case. If someone is above threshold (that's above that 0.3) then he or she will finish high even if overtly passive. This appears to hold as long as the group functions, and there are a few caveats, but generally you can at least appear be passive if you start above 30%. This may or may not make a whole lot of sense in causal terms, but that's what we're investigating.

FIGURE 2



There are several questions to ask. Are initiators born or made? Nature or nurture? Is somebody an initiator in one group and not in another? Very possibly, but perhaps not. Maybe the people who are really passive and getting those low scores cannot be made to initiate. Will different group structures help? We haven't done much experimenting with that. Will teacher intervention help? Can you do anything even if you can identify these students? Now, in a slightly better class, we would have a finite number of students that we could watch to see if they were getting into trouble. We can watch them and see, and try some things.

Another thing to ask is how easy is it for the practicing teacher to identify a class of behavior. We did this very carefully—counting interactions and so forth—but it looks as if you can identify many of these people if you just follow some rules. We are going to experiment with that, maybe with some of you, perhaps by showing videotapes and asking whether you think specific people are initiators. If you can identify them easily, that means you have another handle on your students right away, as long as you have an interactive session in the first few labs.

Finally, is the distribution of a student's knowledge important for students near threshold? I'm using a sum score that's some sort of average. It turns out that we can look in detail to see why they got the score they got. This may actually make a difference in whether they are on one side of the threshold or another.

We have also looked, not in the laboratory, but in a classroom of a non-calculus physics class where we do the standard mechanics Interactive Lecture Demonstrations (ILD's)², and we get a similar shaped curve (see Figure 2). The plot of post-test versus pre-test has 195 students in it. Some of the dots are multiple and

so are larger. You can see that you get the same behavior here except the pre-test results are lower (no engineers and physicists) and the post-test results are higher. How you determine whether somebody is an initiator or not, in the group discussions that go along with interactive lecture demonstrations, I don't know. The real question may be "Is initiation really the criterion that makes a difference?"

Some of my presentation has been fairly specific, but back to the major question. The major question is "Are there easily identifiable characteristics of a student that will predict that student's success in learning physics?" For me learning physics means understanding concepts—I'm not talking about learning algebra only. Are there characteristics that we can measure conceptually or identify by looking at interpersonal behavior? Given that, can we make changes in that student's experience to help him or her learn?

This is a much less grand vision, perhaps, than some of what we heard, but I would love to know these answers. I see an awful lot of research to do before anyone understands even the simple questions I have raised.

References:

1. Thornton, R. K. and Sokoloff, D.R., "Assessing Student Learning of Newton's Laws: The Force and Motion Conceptual Evaluation of Active Learning Laboratory and Lecture Curricula", *Am. J. Phys.*, 66, pp. 338-352 (1998)
2. Sokoloff, D.R. and Thornton, R. K., "Using Interactive Lecture Demonstrations to Create an Active Learning Environment." *The Physics Teacher*, Vol. 35, pp. 340-346 (1997).

Alan Van Heuvelen

I have three ideas to add to the physics education research topics suggested by others. The first concerns the role of representations in physics learning. Jiajie Zhang, in a recent paper, says, "external representations are so intrinsic to so many cognitive tasks that they guide, constrain and even determine cognitive behavior." He goes on to say, "relatively little research has been directed toward the nature of external representations and cognition."

Internal representations involve some unknown brain language for things like audible and visual images and for conscious thinking. External representations are things such as diagrams, sketches and equations on a paper and signs beside a highway. Xueli Zou, one of our graduate students, prepared a nice figure (see Figure 3) that shows four different types of representations of a physics process. The different representations contain more or less equivalent information. The sketch helps our brains visualize the process quickly but is not useful for calculating things. The equations are better for that but it is difficult to visualize the process by looking only at the equations.

Historically in physics education we have relied on words and equations, both very abstract representations of the world. Perhaps this is part of the reason that students do poorly on conceptual tests. We need to integrate more intuitive representations into our instruction and perhaps invent new representations for difficult areas—representations that help students use the concepts in those areas more effectively.

FIGURE 3

INFORMATIONALLY EQUIVALENT REPRESENTATIONS

A miner pulls a wagon of supplies to the top of a hill where he lives by hanging from a rope attached to the wagon. As the cart moves at increasing speed up the hill, the miner moves with increasing downward speed.

$$T - f - W_c \cos\theta = m_c a_{cx}$$

$$N - W_c \sin\theta = 0$$

$$x_c = x_{0c} + v_{0xc}t + \frac{1}{2}a_{cx}t^2$$

$$0 = a_{mx}$$

$$T - W_m = m_m a_{my}$$

$$a_{cx} = -a_{my}$$

$$y_m = y_{0m} + v_{0ym}t + \frac{1}{2}a_{my}t^2$$

An example might be in geometrical optics where we use three rays in diagrams to locate images. Researchers find that students think that if a card blocks half of a lens, half of the image disappears. What if we use multiple rays in the diagram instead of three rays? Will students realize that the image does not travel intact from the object to a screen?

My favorite example about the importance of representations is from quantum electrodynamics (see Figure 4). I remember toward the end of my years in graduate school sitting in a course on S-matrix theory. I did not understand anything. The S-matrix equations lacked intuitive imagery. Feynman diagrams had been invented but we did not use them. These diagrams are more intuitive representations of QED scattering processes. There are rules for converting the diagrams into the equations—much the way we use free-body diagrams to help construct the component equations for Newton's second law.

FIGURE 4



This leads to another related representation project—how do we get students to use these diagrams most effectively? I find that students tend to think of constructing diagrammatic and graphical representations as a separate task from constructing the equations to describe a process. For us, free-body diagrams and qualitative work-energy bar charts are tools for helping to visualize a process and for helping to construct mathematical descriptions of processes.

In summary, there is considerable work that I think we can do in helping students learn to represent processes with more intuitive representations, in using these representations to help construct the math representations of processes and in

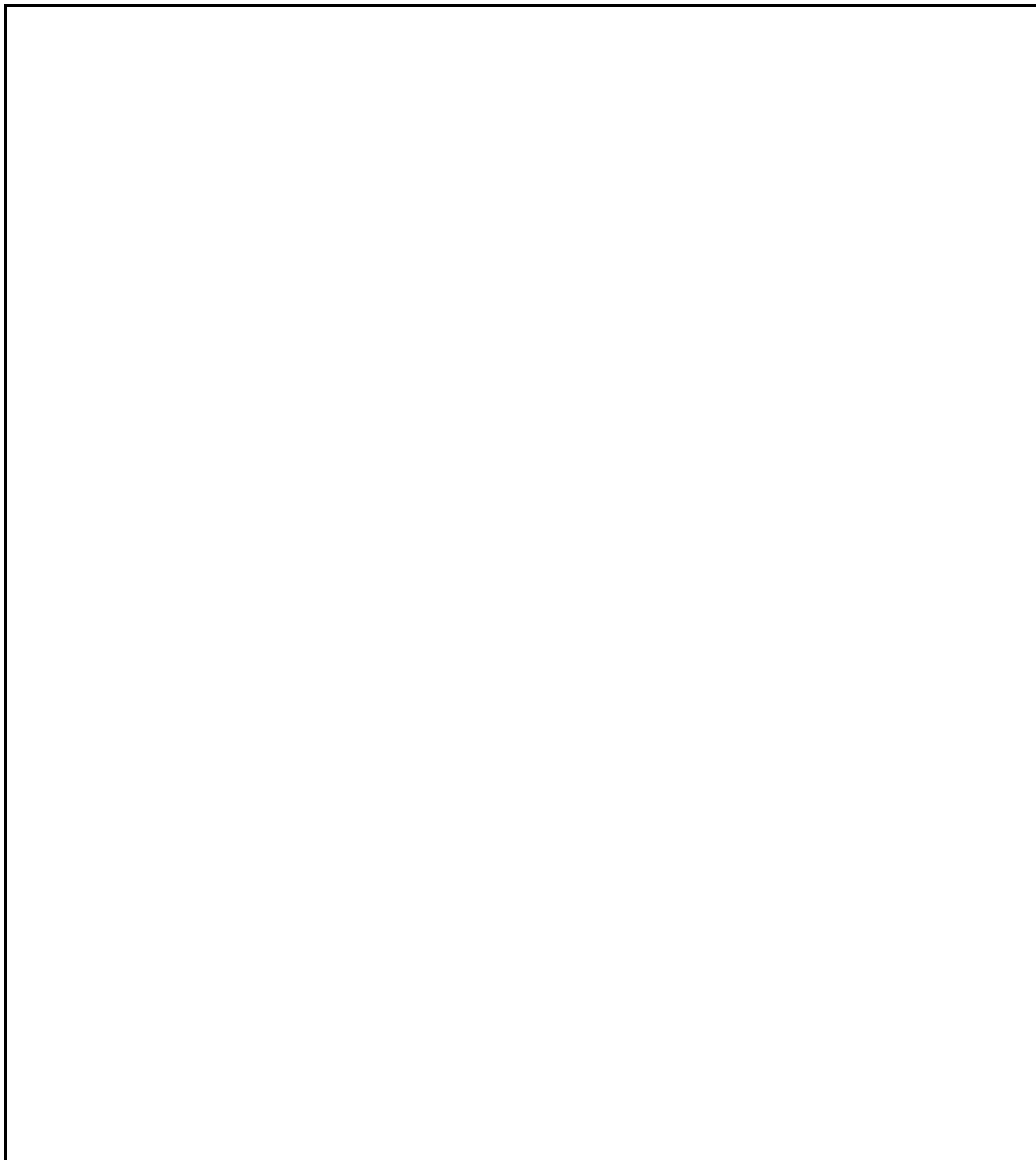
even inventing new representations that enhance conceptual and problem-solving learning.

A second idea for physics education research projects concerns workplace skills. In recent years, there have been several studies concerning the skills that people find most useful in the workplace. A US Department of Labor SCANS Report in 1991 indicated that business and industrial firms are looking for people who have learned how to learn, have communications and problem-solving expertise and are effective in groups (see Figure 5). How do we help students improve their abilities to learn how to learn? An AIP 1994 workplace skill study indicated that physicists value these same skills in industrial and government lab workplaces (see Figure 6). Eighty percent of physicists in the workplace either work in a group or supervise a group. Physics knowledge was the least needed "skill" for physicists working in industry or government labs, yet we spend almost all of our research efforts helping people learn physics concepts. Certainly we can't do problem solving without knowledge, but a balance between emphasis on research about acquiring knowledge and acquiring some of the other skills that are very important for physicist seems in order. With the rapidly changing knowledge base, learning how to learn may be the most important thing students could get out of their time in college.

FIGURE 5

Workplace Needs
<p>Businesses and industrial firms want future employees who:</p> <ul style="list-style-type: none"> • have learned how to learn; • possess listening and oral communication skills; • are adaptable because of creative thinking and problem-solving expertise; and • are effective in groups using interpersonal skills, negotiation skills and teamwork. <p>* SCANS (Secretary's Commission on Achieving Necessary Skills), "What Work Requires of Schools" A SCANS Report for America 2000," Washington, DC: US Department of Labor, 1991.</p>

My third and final research idea concerns the use of computers and technology. So many people are now starting to make Java simulations and other software activities. There are several questions. How do we best use these products to enhance student learning? What are the characteristics of good software—for example, is seeing a process and a simultaneous representation of that process important for learning? I believe that there is a huge potential for using interactive software products in all kinds of learning environments. This is a subject area with considerable research potential.

Figure 6**Jim Stith**

It appears to me that we talk a lot about process versus content. Back in the IUPP days we talked an awful lot about “less is more”. We have said “less is more” for so long that it has almost taken on the role of fact. What is the research basis for that particular assertion that we have been making for several years? If we are going to

keep saying that, and if we are going to act as if we believe it is true, do we owe it to our community to at least get some experimental basis supporting that?

Now I'll ask the panelists to react to that.

Bruce Sherwood

I could offer one comment. When Ruth Chabay and I were developing a new approach to E&M we consciously took a principle of minimalism, by which we meant if there was not a good task where students would actively use the concept, then it was better not to teach that concept because it wouldn't be learned. And there would be a double whammy because not only would you have spent time teaching this thing which didn't get learned because it was not exercised, but you would also have paid an opportunity cost in not having spent that time on something that would have been useful.

So our curriculum is minimalist in that sense of the word. We made a measurement, which we reported at last year's Denver meeting, that students in this curriculum scored a letter grade higher than students in the traditional curriculum taught by excellent teachers. Unfortunately, while we think that the cleanliness of the curriculum contributed, it's like what other panelists have said—it's very difficult to separate out whether it was that or it was something else we did, because naturally you throw everything you can at the problem. Sorting out which of the elements was the crucial one is almost impossible, and there might not even be one single crucial element. It might be that you get the one letter grade better from having done this and that and the other thing, and each of them contributed their little widow's mite, which eventually added up to something.

On the one hand, I don't see how you would ever, in an actual field situation, tease out which of the things really made the difference, because you don't have that much control. That makes it really hard to answer the kinds of questions that are asked.

Ron Thornton, however, has just given us a stunning example of looking at a very messy situation and essentially, in a kind of anthropological approach, picking out a feature which actually had a strong correlation with some very interesting data. I was thrilled by what you showed us.

Ron Thornton

I would like to make a comment that's related to something some of the other people said. Particularly in high schools where the classes are smaller and the teachers know their students very well, they definitely report (we've asked teachers to look for it) the discouragement students feel when they teach something so that it's not understood, such as Newton's Laws or kinematics. When the students go through that and they don't understand, they get very discouraged. First of all, this looks like sort of a simple thing. They thought they understood it and yet they're unable to give the answers that are wanted, and they don't understand why their answers aren't right. That's very discouraging. Teachers report that when they go on to other things it continues to be discouraging.

If, however, they spend more time and actually teach something so that the majority of students understand things, they feel a large surge in interest in the class, because they understand something. It might be that they're learning what it means to understand something. This is conjecture—the fact is that these students then pay a lot more attention to physics and seem to do better later on.

How does this relate to less is more? Unless you take the time to teach something well, then students won't have an experience like that, or at least seldom will they have such an experience. That's not a research result, but it might be an argument for less is more, and it's one that you (Bruce) evidently said as well. You were only going to put things in your curriculum that students seem to process and understand.

Alan Van Heuvelen

If I have to roar through one chapter a week I just don't have time to get students to the point where they can use the mathematical language of physics with understanding, which also involves being able to break a complex problem into parts and then put them back together. But I find it really fun if I can cut out some of the knowledge, and it seems as if our former physics majors aren't going to mind if I do that. If I cut out some of the concepts then I can slow down on the parts that I've identified as most important. I can take time to build that foundation, learn to use the math and then get into some complex problems. I can also make group work a huge part of that and maybe even work a little bit on communication skills. I think this AIP study really supports this less is more philosophy very nicely.

Jim Stith

Let's shift to questions. I've got several in this vein, about dropping free body diagrams, for Bruce. If our major clientele is engineers and we drop free body diagrams from the course, what will be the impact? What about those of us who teach pre-meds?

Bruce Sherwood

I welcome the opportunity to comment on the supposed tyranny of the engineers, because I know that in many physics departments, conservative physicists who do not want to change what they've been doing hide behind the pretense, "But the engineers won't stand for it." When you actually talk with engineers you get a somewhat different story. Tom Foster actually did this at the University of Minnesota, and what he heard from the engineers is quite different from the physicists' presentation of what the engineers would say were you to ask them something.

At the University of Michigan the chemist Ege wanted to reform the introductory chemistry curriculum to make it more like chemistry, as opposed to freshman chemistry. Her chemistry colleagues said, "The engineers won't stand for it. You can't do that." She went and talked to the engineers, who said, "Well, you're the chemist. You ought to think through what would be a good chemistry course." In our specific case, the engineers who will use something like free body diagrams always have their own engineering courses that teach that in spades. They'll have a

whole semester of statics, they'll have whole semester of dynamics, they'll have a whole semester of deformable media, which we don't touch.

It is not necessarily the case that the engineers want what we have been giving them. They have been taking what we have been giving them, but what they mostly need from us is deep principled understanding of some basic physics principles. It is not up to us to teach circuit theory or bridge construction. That is not our job, and we wouldn't even do it very well if we thought that was our job. The engineers will do that themselves. We have something else to offer—something students won't get in the engineering courses.

In some places it's going to be a hard sell—in some places a very soft sell—but we shouldn't delude ourselves into thinking that we are overly constrained by engineering requirements, because I assert that we sometimes are talking to ourselves rather than to them.

I didn't address the bio-med issues because that is a harder one for me to think through.

Jim Stith

To the group. Do you need a Ph.D. in physics to do physics education research? Some of these agenda items seem to be content independent.

Ron Thornton

I don't know if you need a Ph.D. in physics to do physics education research, but I think you don't. Some of us that have Ph.D.s in physics still can't do physics education research. If the question means, "Will you be allowed to do it?" that's another question. I don't think, quite frankly, that these things are content independent. I think they depend very strongly on the content. The principles that we invoke, within the content, are content independent, but understanding what goes on in a particular field for teaching students is highly contextual. I know quite a bit of mathematics, but I still don't know how I would teach mathematics without doing a real study of what goes on in the mathematics classroom and understanding all the various pieces.

I think that you have to know physics very well to do physics education research. That doesn't mean you have to have a Ph.D. Having a Ph.D. is not even a guarantee that you know physics very well. You may know some small area, but it's no guarantee you know the overall layout of the field, and it certainly has nothing to do with knowing how to teach physics, unless that's what you are concentrating on.

Q

I think the question is more related to getting a Ph.D. in physics with a dissertation involving research in some of the topics mentioned. I don't think that would go in our physics department. It's not a question of whether the topics are interesting, perhaps somebody should do them. Part of the intent of this was to help graduate students in the field. To make suggestions that they pursue research that is not going to get them a Ph.D. in physics, I think, is a problem.

Jose Mestre

That may be a problem in some physics departments, but may not be in others. One of the issues that I mention about assessment is something that I don't see how a physics department could possibly say is not important.

Q

This is the second year in a row we've brought grad students to this meeting from all over the country. Many of those students are in physics departments hoping to get a Ph.D. in physics in a field that is respected by the traditional physics community, and we are not addressing issues in intellectual content of physics at all. I find it dismaying and distressing. We're discussing how to increase funding, which is important. But it's not important in the first few years of being a graduate student. It's important that they learn, that they address intellectual issues so they can get a direction for their research. Some of these other things are interesting but peripheral. And it's not a comment about assessment. Assessment is very important. There's assessment of attitudes, assessment of understanding, but at least in our department most of these interesting issues that have been brought up would not get a Ph.D. in physics.

If one of our graduate students was to write a dissertation in physics education research in our department that did not indicate that student needed to have deep understanding of the subject matter and the problems it presents to students as physics subject matter it would not be approved. That's what I'm trying to say.

Q

I wrote the question. Our physics department is relatively conservative, but at least it has taken the step of supporting physics education research. There are a lot of departments out there that have not taken that step. In order to justify my presence in that department I have to do something, I believe, that shows that it does indeed belong in physics, that requires a deep knowledge of some physics.

Jose Mestre

First, I'm not sure I'm willing to grant you that some of the things we've discussed don't require deep knowledge of physics. Second of all, maybe that's a problem specific to your school. Joe Redish just sponsored a dissertation on research in some of these things, which was accepted by the department at Maryland.

Ron Thornton

I'm a little bit confused about this, and I'll tell you why. For instance, suppose you're studying Atwood's Machines and you know all there is to know about Atwood's Machines. Does that show a deep knowledge of Physics? No, probably not.

I'm trying to understand exactly what the thesis topic would have to be to satisfy your physics department. That's what I'm asking. The richness has to do with the educational surround, about the learning of physics, not with just the content.

Q

Nothing's black and white. Every single physics education research dissertation in our department, by the nature of what it was, showed a deep understanding of the subject matter. Let me try to contrast that with the different topics from this session.

Suppose you wanted to do research on what workplace attitudes were fostered by physics classes. I'm just making that one up. That kind of topic which is more sociological, but undeniably important. That would not have a home in many physics departments. I'm terribly concerned about these issues, and we hope we're doing some of those on the side. But no student in our department who made that his or her primary dissertation topic would be acceptable to the department.

We could probably get a student through on one of these. But the respect that the department would have for us and for our endeavor would not be maintained. I'm trying to make it clear to the young people in here that if your aspiration is to be in a physics department—and if you're going to affect the teaching of physics you have to be in a physics department, because nobody else can teach physics—you have to earn a place in a physics department. I've had enough experience to recognize that people are not inclined to look at this kind of research as adequate for their faculty. I don't think these aren't important. I think they're more important than a lot of other things.

Jim Stith

I'm going to move on to another question. and I'm going to invite us to continue discussion during the break and at lunch because this is a very important one.

Jose, how do you know that your assessment tools really evaluate what you want to know? How do we get to trust our results?

Jose Mestre

How do we know we're measuring what we think we're measuring? I'm not sure of the answer. That's a tough one.

Jim Stith

How do we convince our colleagues?

Jose Mestre

Say you develop an assessment item or an assessment technique. I don't think you stop at what the results are. You have to keep going back and measuring things from different perspectives to see if in fact the results that you are getting from your assessment are reflecting what you think you are measuring. You might want to do in-depth interviews after the student answers a question to see what it is that they think they are answering. It is not a one shot thing. You have to study something from different angles, and then if all the angles give you the same answer you start to become convinced that maybe it's measuring the thing you want to measure.

Q

It takes a long time. It takes about three or four years of intensive work with one assessment tool to really have validated it in all the ways we talked about yesterday.

Bruce Sherwood

I'll offer a quick case study on the use to which these results are put. I alluded to a measurement that we carried out, comparing two groups of two different E&M curricula, and there were very large effects. For many years our colleagues in the physics department had asked us for hard data that would assess what we were

doing. It's almost impossible to make that measurement early in the project at a time when everybody wants you to make this measurement. At the beginning, it makes no sense—you're just trying to get the thing to work at all. But after five or six or seven years of this we were finally at the stage where we could make a real measurement. We looked at this and that and the other and there was a robust, large effect showing the difference between the two ways of teaching. The results of this measurement had almost no impact. It is very easy to request the measurement. It is very difficult to make this measurement. And it is very easy to ignore the results.

In particular, it had been well known for many years that the engineering students at Carnegie Mellon are much stronger than the science students. It was the science students who had our E&M curriculum and it was the engineers who had the traditional one. After this measurement it was now well known that the engineering students aren't as strong as the science students. Another quite typical reaction was, from very good traditional teachers, saying, "I will work harder; I will work harder."

The validity isn't even in question here, but that other issue that came up certainly is. What does it mean to anybody else? Often very little, even with very hard data.

Jim Stith

"What should we teach physics or engineering majors?" is a different question than, "What should we teach to pre-service, pre-med, humanities, etc majors?" Question. How can our community help determine the course objectives?

Bruce Sherwood

I guess I was just suggesting that it ought to be a more visible agenda than our discussions usually have. We talk more about how we can improve the current instruction. We have been talking less about what we should be teaching. And it should probably be different things to different audiences. I would just like to hear more discussion.

Q

At the University of Minnesota we surveyed the engineers who had sent their students to the calculus-based course, and asked them why they bothered to send their students to us. We interviewed, in a separate project, those departments that send students to our algebra-based course. Technical fields: agricultural students, architecture students, physical therapy majors and so forth. We asked them the same question. We got the same answer from all departments: basic concepts; qualitative and quantitative problem solving. They want the same basic stuff. Maybe the level at which the stuff is taught is different, but they want the same stuff.

Jim Stith

Is the idea of misconceptions too limiting a way to think about what students do in physics? How do we evaluate skills and process knowledge?

Ron Thornton

I don't think anybody in the group up here is overly focused on so-called misconceptions, and most of you in the audience don't either. We like to use activity-based methods because it is a process and we like to use those methods to help the students learn about the world and to do some of the synthesis, and so on, on their own. We're not talking about open environments here, we're talking about rather highly guided environments, because of the time savings associated with guided environments. We require a lot more of the students in courses that have been changed in the ways we like. They are doing a lot more thinking than they do in traditional environments. That particular focus is definitely on process. When you do group work, you are also focusing on communication and on transmitting your results to other people within a small group, and then sometimes you do group construction for the group as a whole. I would say that all of these are process things. They're the pieces that, we've found, work for getting students to understand on a conceptual basis.

Q

I think that's this question is getting at something different. Don't we need an assessment that could assess things other than conceptual learning. A lot goes on in these courses besides conceptual learning.

Alan Van Heuvelen

We're doing a little bit of work in that. David Van Domelen is trying to make an assessment that can evaluate a student's abilities to, for example, take a complex multi-part problem and break it into pieces and, more or less, find a solution. We're using more or less standard techniques. We use think aloud. We give the student an activity and do a think aloud thing and listen to what they're saying. We have tried several different formats for those kinds of tests, and it's a fairly difficult task. What we find quite often, too, is that students coming out of a traditional class are very unsuccessful at doing those kinds of things.

Jim Stith

Last question to all the panelists. What is really known in cognitive science that is or could be useful to us in helping us to put what we do into a theoretical framework?

Bruce Sherwood

I live at Carnegie Mellon which is a hot bed of cognitive science with the world's highest ranked cognitive psychology department. I certainly feel that that my views and my efforts have been significantly stimulated by contact with that community. But it's too much to say that they somehow can tell us what we need to do. It's much more complicated than that.

As Ron says, much of what we have to do is very content specific, even though general principles are being applied. The general principles make sense in a specific content. Even at a place like ours—heavily influenced and surrounded by this—it's not that somehow they can just tell us what we need to know. Cognitive science has

provided a vocabulary, a conceptual framework, a lot of things, but it's not a forcing function that can tell us what we can and should do.

Ron Thornton

A very simple example that came up is, "Representations are important." Cognitive science tells us that representations are important. It doesn't tell us whether the representations we're using in physics are effective or not, and to some extent we already know that some of them aren't. So, now that we know they are important, what do we have to do? We have to figure out what the representations are used for, and what representations would be better, and how to determine that they are better, and so on. I don't think cognitive science is going to do that. We will have to do that. But they've told us that it's important.

Jose Mestre

All these questions you are asking take forever to answer, but let me just say that in the spring of '99 there will be a report coming out of the National Research Council that I recommend that everybody read. The report will be called "How People Learn: Brain, Mind, Experience and School" and it will go into a lot of these issues about what cognitive science is good for. Just as a quick summary, there is a lot in the expert/novice literature that can be applied to what we do. For example, how do experts organize their knowledge so that it is so efficient for recall in solving problems. What are the conditions for transferring knowledge to problems? All these things are important. The report goes into that, plus a lot more. Another example is uses of technology for instruction, in the right kind of way rather than just developing a program and saying, "Here it is." You're asking questions that cannot be answered in two or three sentences, so I'm throwing the suggestion out that you might get good ideas by reading that report.

Q

It's very interesting to me that we've been talking about what cognitive psychology can tell us, and so on. There is a field that has some answers for some of the questions we've been discussing. I find it interesting that we're not saying, "What is it that we already know from the field of education that can inform what we are doing?" as well as, "What do we know about how students learn?" Some of the questions that were up there have actually been researched and some is known. Is everything known? No. But it's a place to start, and I would strongly advise that as a group we don't ignore it.

Additional questions

During this panel presentation, audience members were asked to write questions on pieces of paper and pass them to the front. Only a few of the many questions were addressed during the question-answer period. Below is a list of all questions collected, in no particular order.

1. What is really known in cognitive science that is or could be useful to us in helping us put what we do into a theoretical frame? Is the idea of "misconceptions" too limiting a way to

- think about what students do in physics? How do we evaluate skills and process knowledge?
2. A general question: When we talk about issues of “what works” we predominantly make use of data such as a class average on a particular test score. This is great and helps us begin to understand what works. However, this still doesn’t address the question of what works for all students. How can we do a better job of focusing more on assessment issues related to individual learning gains - rather than just group data? Not all instructional or curricular tools will work for all students. So even when learning gains are achieved as evidence by class data, we still need to address how to deal with those students for whom the tools didn’t work.
 3. What should we teach to physics and engineering majors is a different questions than: What should we teach pre-service, pre-med., humanities, etc. majors. How can our community help determine the course objectives?
 4. Do you need a Ph.D. in physics to do physics ed. research? Some of these agenda items seem to be content-independent.
 5. This is more of a comment than a question: To extend Alan Van Heuvelen’s comments on representations and simulation. We need research on teaching students to draw scientific diagrams (which we currently don’t teach them). We also need research on the construction/use of visualizations of abstract concepts (e.g. fields).
 6. My wish is to invent (a black box called) instruction method to go from initial state of knowledge of zero to final state of knowledge 100%.
 7. Doesn’t “what needs to be in the curr.” depend on the student population? I teach engineers. Don’t they need to understand forces and free body diagrams? You teach pre-meds or non-science/engineering majors. They may not.
 8. If one goal is “interpersonal skills” how do we develop this in the typical university professor?
 9. Jose: Specifically, how do you propose tracking students in reform classes as an ethnography? Would you plant a spy-student like Tobias? Or will you intentionally alter the dynamics with an observer?
 10. What are the courses of faculty “inertia” toward changing their instruction? Can it be reduced?
 11. How do we convince physics departments to hire someone in physics education as opposed to high energy, atomic, etc?
 12. To Alan: If we don’t focus on physics content, why do we belong in a physics department?
 13. Many new approaches to physics instruction - e.g. interactive engagement, collaborative groups - are coming out of PER. Assessment of the overall effectiveness of these approaches is largely limited to conceptual tests, such as the FCI. Although gains are high, these results are generally not impressive to our colleagues who are skeptical of these new approaches. For example, we are accused of “teaching to the test.” What reliable assessment tools can we develop for quantitative reasoning and problem-solving? These are the skills by which many of our colleagues will judge our success or failure.
 14. Are there any factors which can predict at-risk students? For example, is GPA a good predictor?
 15. What do you think will be the effect of history of physics?
 16. What role do you see science education research (K-12) playing in these research areas?

17. Does the distribution of before and after scores reported by Ron change in a course without the group work component? Do low starters stay low starters when they can't initiate?
18. How does an instructor decide what is important to teach and how much time should be spent on each area. Who should decide this?
19. Should we not expand or invite general education educators to our research groups? I think the answer is yes. It will/may (1) expand our influence to more high school science/physics teachers. HS certification requires general education course. If we can influence these educators, we can influence HS teachers. (2) expand our access to funds. These are education funds available and new alliances would open these doors. (3) In terms of recognized education groups, we are the "new kind on the block." We would learn the old timers vocabulary so we can change them.
20. Dropping free body analysis - if a major clientele is engineers, what would the impact of dropping free body analysis from our courses? Can we afford the loss of engineers from our courses? Will it happen? Is the problem what we teach or how?
21. To Jose: How to evaluate students' problem solving skill? How did you know your assessment tools really evaluate. What you wanted to know? How did you trust your results?
22. Bruce Sherwood: "Improving buggy whips." Will there ever be enough consensus on what a buggy whip is and what it's not.
23. Research on institutional contexts as they affect instruction.
24. Jose & Alan: Can't we focus more on hands-on assessments, where we develop methods of evaluating students' ability to solve practical problems involving manipulation of equipment and design of experiment.
25. What role do you see science education research (K-12) playing in these research areas?
26. K-12. What is it about physics that students find so much more difficult than other subjects. Is it non-intuitive? Is it non taxonomy based? Can we learn from comparative studies with: math, chemistry, biology. Are there meta-issues?
27. Bruce Sherwood: What should be in the introductory physics curriculum? What evidence can you offer that suggests that free-body diagrams are more difficult to teach than topics in modern physics. Comments: On-going research on student understanding in advanced topics hydrostatics, thermal physics and engineering statics suggests that difficulties with free-body diagrams and Newton's laws persist in new contexts. These results have been reported at national AAPT meetings and will be in several Ph.D. dissertations in physics education research at U. Washington.