Commentaries on physics education by David Hestenes

Who needs Physics Education Research!?

Am. J. Phys. **66:** 465-467 (1998)

David Griffiths[1] has used the occasion of his well-deserved Millikan award to raise serious questions about the reform movement in physics education and the "Hestenes test" in particular. Since my name has been taken in vain, so to speak, I feel compelled to respond.

Along with F. K. Richtmeyer in his inaugural article for the AJP[2], I concur with Griffiths' dour assessment of the *amateurish* state of physics teaching generally. However, I do not believe that substantial improvements can be achieved without a strong program of physics education research (PER). The problems are too difficult and complex to yield to amateurish efforts. Nearly two decades ago I penned a diatribe on the need for a "Science of Teaching."[3] I have since seen PER emerge as a credible discipline in its own right, with a growing body of reliable empirical evidence, clarification of research issues, and, most important of all, an emerging core of able and committed researchers within physics departments across the country. Most of our colleagues have been oblivious to this movement, if not contemptuous of it. Some are beginning to realize that it is more than another "educational fad." It is a serious program to apply to our teaching the same scientific standards that we apply to physics research.

What does the FCI tell us?

I will focus on Griffiths' concerns about the *Force Concept Inventory* (FCI)[4] and its implications, but I wish to place it in the larger context of PER. By the way, the FCI should not be called the "Hestenes test," because its design, construction, analysis and validation was a cooperative effort in which Ibrahim Halloun and Malcolm Wells played crucial roles.

Griffiths is right to be skeptical about the FCI. A healthy skepticism is an indispensable component of the scientific mind; it drives the search for evidence and justified belief. Unfortunately, an unhealthy skepticism is all to common among physicists, especially with respect to education; it seeks to deny and discredit claims that conflict with cherished beliefs.

Griffiths' concerns are important, because they occur to every thoughtful physics teacher confronted by FCI results. I have heard them time and again for years. It would be a poor sort of educational research that failed to address them. The fact is that efforts along this line are far more extensive than casual observers like Griffiths suspect. My personal observations and conclusions are offered here, but teachers must consult the literature and their own classrooms for supporting evidence. Critics of educational reform need to do their homework. Anecdotal evidence is no more adequate in education than it is in science. At best, it suggests directions for productive research. More often, it is misleading or altogether wrong. Some critics doubt that hard evidence is possible in education. The FCI results stand as a counter example.

The data base of results from the FCI is enormous and growing rapidly. I have direct knowledge of data on more than 20,000 students and 300 physics classes spanning the range from high school to graduate school. Judging from the steady stream of FCI reports at AAPT meetings, the FCI has undoubtedly been used in hundreds of other physics classrooms. Richard Hake has compiled, analyzed and published a substantial portion of that data.[5] Halloun and I are analyzing more extensive data on high school students and teachers. Altogether, the data provides overwhelming support for our original conclusions. The data base is now so broad that the unsettling message it brings can no longer be attributed to bias or incompetence of the original investigators.

Having administered the FCI at his own school and seen the dismal results, Griffiths does not doubt the published data or its importance. Rather, he questions the validity of the FCI and the urgency of the results. His doubts are based on a general skepticism of multiple choice tests and his own arm chair analysis of test items. It is precisely to answer such doubts that the FCI has been carefully validated with extensive student interviews. All this has been thoroughly documented in the literature and repeatedly checked by many different people. The FCI is not comparable to the off-the-cuff multiple-choice tests that teachers construct on their own. The carefully constructed distracters for each item are not typical multiple-choice throwaways, but common sense alternatives to Newtonian concepts that amplify the significance of student responses.

It is instructive to see how Griffiths' FCI item analysis falls short. In the published version of his Millikan lecture he omits details from his public presentation — specifically a lengthy analysis of ambiguities in item 19 of the original FCI.[4] That item is about superposition of forces. Griffiths showed how one might construct valid physical arguments in support of any of the given choices — the implication being that good students might thereby be misled into making the wrong choice. Fair enough! However, Griffiths himself had no trouble selecting the "correct response," and our data and student interviews revealed no evidence that the question was "missed" by anyone who understood the Newtonian concepts. In fact, as explained in the article, we found that item 19 is defective for precisely the opposite reason: too many students chose the Newtonian response for nonNewtonian reasons. We retained the item in our published test to emphasize that point, and because the superposition principle is too important to ignore. Unfortunately, we have not devised a satisfactory replacement, though we know that many students have serious misconceptions about superposition. Item 19 has been dropped in a recent minor revision of the FCI,[6] with the ironic consequence of slightly lowering FCI scores.

Griffiths' analysis raises an important point about precision and ambiguity in test questions. Physicists have developed a technical language for precise expression of scientific concepts and unambiguous description of physical situations. Unfortunately, until students are privy to their special meanings, technical terms can be a barrier rather than a help to understanding. Consequently, on examinations students often respond to the form of the technical language rather than its meaning. For example, for a typical University Physics course we found that nearly 80% of the students could *state* Newton's Third Law at the beginning of the course, while FCI data showed that less than 15% of them fully understood it at the end. In designing FCI questions we tried to avoid technical language in order to get closer to what students really think. We reasoned that Newtonian thinkers would be able to resolve the consequent imprecision and ambiguities. Our validation interviews confirmed this. A few FCI questions have been revised[5] to remove ambiguities that bothered other physicists besides Griffiths; however, as Hake reports, this has not significantly affected test scores.

As Griffiths notes, FCI data has been widely cited in the current reform movement as compelling evidence that there are serious problems with physics instruction. It is too often overlooked that this is far from the only evidence. There is a huge PER literature on student misconceptions which supports the same conclusion. Most physics teachers are oblivious to the *huge gap* between what teachers think they are teaching and what students are actually learning.[7] More than anyone else, Lillian McDermott[8] and her coworkers have been carefully documenting this gap in one topic after another, reporting results in a steady stream of AAPT talks and AJP papers. Not only is physics instruction frequently failing to address student misconceptions, it often inadvertently strengthens them and induces new ones. Much more PER is needed to sort all this out. Readers who want specifics are advised to attend PER sessions at the semiannual AAPT meetings.

Perhaps the most important function of the FCI is that it sets a *minimal standard* for effectiveness of instruction in Newtonian mechanics. It is a discrimination test, requiring only that students make a forced choice between basic Newtonian concepts and naive alternatives. Griffiths wonders whether we should expect students to meet this standard in a first course. Surely, he argues, the student has “learned” the material at some level, and real understanding takes a long time to mature. Unfortunately, what the student with even an average FCI score has learned is more likely to be wrong and misleading than enlightening. The Newtonian *concept of force* is complex, with six major components which are systematically probed by the FCI. To the extent that students have not mastered this complex concept, they will systematically misinterpret what they hear and read in the physics course; they will treat the technical language of physics as muddled jargon; and they will be forced to resort to rote methods in learning and problem solving. Hake’s data[5] supports our evidence that problem solving skill really does depend on the concepts assessed by the FCI. (Would we expect it to be otherwise?) Therefore, an emphasis on problem solving without due attention to FCI concepts will be counterproductive. It will reinforce mindless plug-and-chug by rewarding it. *Practice makes permanent*, but not necessarily *perfect!!*

After one is convinced that physics instruction cannot be very effective without addressing student misconceptions, the question is “how?” Somehow Griffiths has come to the mistaken belief that reformers advocate “teaching to the test” to raise FCI scores. In fact, that approach fails badly. Just telling the students the answers induces only rote learning, which has a half life on the order of a few days. The problem is not to teach “right answers” but to develop cognitive skills that generate right answers.

The Good News is that gains in FCI scores can be improved considerably by “interactive engagement” teaching methods, as documented by Hake.[5] The Bad News is that this is not easy, and there is plenty of room form improvement. Using a “modeling method of instruction,” Malcolm Wells was the first to achieve relatively large FCI gains.[4] Recently, a whole cohort of high school teachers has achieved comparable gains after learning the method in NSF-supported workshops.[9] Unfortunately, that version of “modeling instruction” is not readily adapted to the large lecture halls in universities. However, Hake reports success by other approaches that are so adapted.

What are lectures for?

Many people are convinced that FCI data shows conclusively that lectures are (perhaps, totally) ineffective in teaching the basic concepts of physics, even apart from other evidence pointing to the same conclusion. The FCI gains that are found with traditional instruction can be attributed to the students’ own efforts. There is no evidence that students who attend lectures learn more than those who don’t. This sobering conclusion has been interpreted by some as a call to banish lectures. Griffiths has risen to the defense. What we need most, though is an objective assessment of what can and cannot be accomplished by lectures — another job for PER.

I believe that when the dust settles it will be perfectly obvious to everyone where the traditional lecture method fails. The lecture is an efficient means for transmitting information in a motivating context. However, the message cannot be received unless it is understood, and understanding cannot be transmitted. If the recipient is ill prepared to understand, the information will be lost or misconstrued. The complex cognitive skills required to understand physics cannot be developed by listening to lectures any more than one can learn to play tennis by watching tennis matches.

Griffiths counsels caution in teaching reform lest the benefits of traditional instruction be lost. But he fails to note the dangers. It may be that a good lecture is the best way to communicate a coherent vision and the highest values of physics. But we should be mindful of intellectual casualties among students who lack the skills to match the inspiration they get from lectures with performance on examinations. Even the best of students are at risk. A telling anecdote will help make the point. *Feynman's Lectures on Physics* are widely hailed as a masterwork of physics pedagogy. They have inspired and educated a generation of budding physicists. The tapes of his lectures portray the pinnacle of classroom performance. Recall that the lectures were expressly prepared for first year physics students at Cal Tech. On that score, Feynman himself regarded the lectures as a failure. Only a small fraction of the students were really able to cope with the course. Shortly thereafter I got to know one of those bright students who had come to Cal Tech on a scholarship. He was so devastated by the experience that he dropped out of Cal Tech and came to ASU, where it took some time to recover his self confidence.

The Moral of all this is not that we should dispense with lectures, but that we have much to learn about how to use them as effective pedagogical tools. Enough has been observed about their deficiencies and dangers to shift the burden of proof to their proponents. Happily, some have taken the challenge. Physicists Eric Mazur[10] and Alan Van Heuvelen[11] and engineer Richard Felder[12] are among the leaders in modifying the lecture format to get students actively engaged in the classroom. And they have assumed responsibility for systematically evaluating the effectiveness of their own methods.

What makes a good teacher?

In my experience, the best teachers are the most receptive to FCI results. They are aware that something is amiss in physics instruction, and they are eager for hints on how to do better. Even so, it comes as a shock to learn that the problem is much more serious than anyone had previously guessed. Eric Mazur has publicly testified to his own shock at the FCI scores of his Harvard students.[13] I am happy to say that he followed our advice[4] and immediately interviewed his (more than one hundred) students to confirm that they really did have conceptual problems with the FCI questions that they missed. After recovering from his funk over the results, Mazur went on to study the issues in more depth and to thoroughly revise his approach to teaching. He has continued since to use the FCI to verify that his new methods really are more effective, and he has cross-checked the results with other tests. Mazur is making his methods and materials available to the rest of us.[10] In my opinion his pedagogical success is real and worthy of study if not emulation. But who can duplicate his classroom performance?

Griffiths believes that "*any* pedagogical method requires a good teacher, and good teachers are extremely rare." But why are they so rare? Because good teaching requires a special talent? Or because it requires complex skills that are difficult to acquire?

Recently I have been in a unique position to study teaching competence on a broad scale. For the better part of a decade I have been PI on NSF teacher enhancement grants for inservice high school physics teachers. This has generated extensive data on the teaching of nearly 150 teachers. A thorough analysis will be published when the study is complete, but here is a preview of some pertinent conclusions:

(1) *Subject competence* is essential to teacher effectiveness. Teachers with low FCI scores are unable to raise student scores above their own.

(2) Proficiency in *scientific inquiry* is more important than specific content knowledge. Beyond a minimal background of a few physics courses, teaching effectiveness depends only weakly on the extent of academic physics training. The best teachers love the challenge of learning something new and are eager to share the experience with students.

(3) Managing *the quality of classroom discourse* is the single most important factor in teaching with interactive engagement methods. This factor accounts for wide differences in class FCI score among teachers using the same curriculum materials and purportedly the same teaching methods. Effective discourse management requires careful planning and preparation as well as skill and experience.

(4) Teachers create an environment wherein students *construct their own understanding* of the subject. The quality of the constructions depends crucially on the conceptual tools available to the students and facilitation by the teacher.

> (5) *Effective teaching requires complex skills* which take years to develop.[9] Technical knowledge about teaching and learning is as essential as subject content knowledge. Few teachers can acquire it without participating in a strong program of professional development. However, most are capable of achieving a high level of teaching proficiency, and even the best need the stimulus of peers to keep improving.

Though *good teaching* may be rare, as Griffiths says, I think it is a skill that can be learned by most physics teachers. It may be that *great teaching* requires a special talent. I think Griffiths would agree that great teachers are great learners who love to share the sources of their inspiration.

References

1. D. Griffiths, "Millikan Lecture 1997: Is there a text in this class?" *Am. J. Phys.* **65**: 1141-1143 (1997).

2. F. K. Richtmeyer, "Physics is Physics," *Am. J. Phys.* **1,** 1–5 (1933).

3. D. Hestenes, "Wherefore a Science of Teaching?", *The Physics Teacher* **17**, 235-242 (1979).

4. D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory," *The Physics Teacher* **30**: 141-158 (1992).

5. 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).

6. Updated FCI available at: http://modeling.asu.edu.

7. A. Lightman & P. Sadler, "Teacher predictions versus actual student gains," *The Physics Teacher* **31**, 162-167 (1993).

8. L. McDermott, "Millikan Lecture 1990: What we teach and what is learned: Closing the gap,"*Am. J. Phys.*, **59**, 301-315 (1991).

9. M. Wells, D. Hestenes, and G. Swackhamer, "A Modeling Method for High School Physics Instruction," *Am. J. Phys.* **63:** 606–619 (1995).

10. E. Mazur, *Peer Instruction*, Prentice Hall, New Jersey (1997).

11. A. Van Heuvelen, “Overview, Case Study Physics,” *Am. J. Phys.* **59**,898-907 (1991).

12. R. Felder, “A Longitudinal Study of Engineering Student Performance and Retention. IV. Instructional Methods,” *J. Engineering* **84**, 361-367 (1995). Information about his *Effective Teaching Workshop* can be found at his home page: http://www2.ncsu.edu/unity/lockers/users/f/felder/public/

13. Eric Mazur, Qualitative vs. Quantitative Thinking: Are we teaching the right thing? *Optics and Photonic News*, February 1992, p. 38.