Learning Science

Content—With Reason

By Paul R. Gross

If you are an experienced and loving teacher, you probably have felt the mixed pleasure and pain brought on by students' struggles to display their content knowledge and ability to reason. Surely, you've seen more than a few exam answers like these:[1]

> Nero was a cruel tyranny who tortured his subjects by playing the fiddle to them.
>
> The sun never set on the British Empire because the British Empire is in the East and the sun sets in the West.
>
> Gravity was invented by Issac Walton. It is chiefly noticeable in the autumn when the apples are falling off the trees.

Such answers tickle us because of the mismatch between the test-takers' logic and sentence structure—both of which are normal—and one or more preposterous details of their assertions. The faulty detail can be as simple as a misspelled or misused word, or as flagrant as complete failure to relate cause to effect. Clearly, ordinary competence in language and logic are not enough to keep us from coming up with howlers—if we don't know, or we simply misunderstand, important details of a subject we address.

This is as true in science education as elsewhere in life. And so, in the course of a long career as a biologist and teacher of science, I have often been troubled by the endless debate about whether we should focus on teaching scientific *reasoning* instead of science *content*, or at least more reasoning and less content. But to comprehend science as a responsible citizen, and certainly to succeed in any science-related career, *both* content and reasoning are essential. The absence of one or the other may produce laughter, but not good science.

AURORA BOREALIS, AP PHOTO/ANCHORAGE DAILY NEWS, EVAN R. STEINHAUSER

AURORA BOREALIS, AP PHOTO/WATERLOO COURIER, BRANDON POLLOCK

As the author writes, "To comprehend science as a responsible citizen, and certainly to succeed in any science-related career, *both* content and reasoning are essential." Both are also essential to comprehending the beauty of the world around us, as shown in these photos.

Arguments for much more reasoning and less content (a necessary tradeoff, given time constraints) in K–12 science began decades ago. Eventually, the idea became a catch phrase. "Content" was redefined to function as a synonym for "facts" (or "mere facts") independent of reasoning. But defining content that way is nothing more than a rhetorical move. No honest study of science textbooks and lessons nationwide, not even from the benighted decades preceding the launch of Sputnik, could conclude that just memorizable facts were required, with no reasoning. Facts were (and are) taught, and facts must be learned if any intellectual discipline is to be understood and practiced. The rhetorical flourishes of those arguing for more scientific reasoning have affected some people's perceptions, but they have not changed the reality that, in general, science curricula have never been *exclusively*

Paul R. Gross is University Professor of Life Sciences, emeritus, and former vice president and provost, University of Virginia. He has taught science at New York University, Brown University, the Massachusetts Institute of Technology, and the University of Rochester, and is former director and president of the Marine Biological Laboratory, Woods Hole, Massachusetts. He has written numerous articles and books on topics ranging from molecular biology of development to the intersection of science and culture.

lists of facts to be memorized, devoid of the means by which those facts are discovered and gain acceptance in the scientific community.

Before we go any further then, let's pause for a moment to consider just what scientific reasoning is. What differentiates scientific from, say, historical reasoning? Other than the content being reasoned about, I can't think of anything. So, I turn to the distinguished philosopher of science and epistemologist Susan Haack to discover that the notion of a species of reasoning unique to science is unfounded. Haack writes:[2]

> Scientific inquiry is continuous with the most ordinary of everyday empirical inquiry. There is no mode of inference, no "scientific method," exclusive to the sciences and guaranteed to produce true, more nearly true, or more empirically adequate results. . . . And, as far as [science] is a method, it is what historians or detectives or investigative journalists or the rest of us do when we really want to find something out: make an informed conjecture about the possible explanations of a puzzling phenomenon, check how it stands up to the best evidence we can get, and then use our judgment whether to accept it, more or less tentatively, or modify, refine, or replace it.

The practices of good science are distinguished by that "informed conjecture"—by a special dependence upon technology (e.g., instruments that broaden the human range of perception), and by especially strong and well-enforced rules having to do with scrutiny and testing of claims and reproducibility of results. But they are not distinguished by an array of clearly identifiable, *cognitively unique* forms of reasoning.

What, then, is to be understood by scientific reasoning? The answer cannot be very deep because the question isn't. Scientific reasoning is using, within a framework of scientific content, certain general cognitive abilities that develop over time or can be encouraged in most learners. So, there is not much that is exclusively scientific about such reasoning other than the fact that one is thinking about scientific content. Scientific reasoning is a sibling to, if not perfectly congruent with, historical reasoning, which is the use of rather similar cognitive basics in the context of records and commentary on the past. Scientific reasoning is deployed with hypotheses and observations about nature. It has other siblings as well: social, artistic, and literary reasoning for example.

For those concerned with school science, however, the issue is *scientific reasoning*, and the goal is to encourage better-informed rationality about nature, to bring about significant improvements in students' scientific literacy and problem-solving skills. Of course, there is an enormous literature on the question of how to do this. At least among cognitive scientists, the consensus seems to be that, "Just as it makes no sense to try to teach factual content without giving students opportunities to practice using it, it also makes no sense to try to teach critical thinking devoid of factual content."[3] Here, for "critical thinking," we may substitute "scientific reasoning." In the relevant contexts, they mean almost the same thing: scientific reasoning in the absence of scientific content doesn't make sense. Reasoning and content are not practically and neatly separable.

So, why isn't this old debate over? Why, in fact, is there a debate at all? Unfortunately, it seems that ongoing, important, and often laudable research on how to increase students' science learning continues to stumble, from time to time, over these questions. This is understandable: any researcher will tell you that gathering data about complex processes is the easy part; making sense of those data, and drawing sound conclusions from them, is the hard part. So it's important that all of us, not just researchers but teachers too, question studies that reach puzzling conclusions. Not because we, individually, will thereby come up with the "right" conclusion, but because such questioning is essential to ensuring that the research enterprise as a whole advances both intellectually and in its eventual usefulness.

LAVA, PHOTO BY J.D. GRIGGS, 02/25/83, JG928, COURTESY OF THE U.S. GEOLOGICAL SURVEY

Scientific Reasoning in *Science* Magazine

Let's examine a recent article on scientific content and scientific reasoning that has received a good bit of coverage in the popular media. A few months ago, *Science*—one of the two most selective international science journals (the other one is *Nature*)—published an important article on a study of learning and scientific reasoning.[4] This fascinating paper has some perplexing features. *Science*'s summary of the study declares that "comparisons of Chinese and U.S. students show that content knowledge and reasoning skills diverge." Now, such a showing ought not be in the least surprising to the journal's readers. "Divergence" is both innocuous and ambiguous; and as we have suggested, the claim that content and reasoning can be separated has been afloat for many years.

Nevertheless, however commonplace the statement, such a divergence would be very important if it were (1) anything more than a simple acknowledgement that content knowledge and basic reasoning skills are in some respects different things, and (2) demonstrated unequivocally to exist, with rigor typical of most *Science* articles. It would be very important not only for K–12 science, but for all education. But as noted, the article, titled "Learning and Scientific Reasoning," offers some puzzles. They need to be considered before the study's conclusions are taken as grounds for action. Among the firmest—and yet most questionable—conclusions offered in the text is this:[5]

LAVA, AP PHOTO/DAVID JORDAN

WATER DROPLET, ©ISTOCKPHOTO.COM/DA_PHOTOS

> The current style of content-rich STEM [science, technology, engineering, and mathematics] education, even when carried out on a rigorous level, has little impact on the development of students' scientific reasoning abilities. It seems that it is not what we teach, but rather how we teach, that makes a difference in student learning of higher-order abilities in scientific reasoning.

To restate: "Higher-order scientific reasoning" cannot be achieved by science learners if they are offered only "content-rich" science courses and programs. Something different must be added or substituted. That something, according to the authors, is the explicit teaching of scientific reasoning, here (as commonly elsewhere) identified with inquiry learning. Within the enforced economies and terseness of the writing characteristic of *Science* and *Nature*, a claim such as that is usually taken very seriously. Should this one be so taken? To find out, we must examine the data provided and (using scientific reasoning and relevant content from cognitive science) judge the conclusions drawn from them.

Data for this study come from three tests—two of physics knowledge and one of general scientific reasoning—administered to freshmen college students in the United States and China. All the students were science or engineering majors, enrolling in college-level, calculus-based physics—but the tests were given before instruction began. The authors, Lei Bao and a dozen colleagues, specify carefully the differences between these two cohorts. The most striking is their precollege preparation in physics. Bao et al. explain that "Chinese students go through rigorous problem-solving instruction in all STEM subject areas throughout most of their K–12 school years and become skillful at solving content-based problems." This is, as they note, in sharp contrast with K–12 science education for U.S. students, who probably spend less time in science study of any kind and, obviously, less time doing physics. As the authors observe, "The amount of instructional time and the amount of emphasis on conceptual physics understanding and problem-solving skills are very different in the two countries." This, they claim, provides what is, in effect, a controlled experiment, an opportunity to see if these variations in content learning—*intensive*, as in China, versus (relatively) *superficial*, as in the United States—have an impact on scientific reasoning ability.

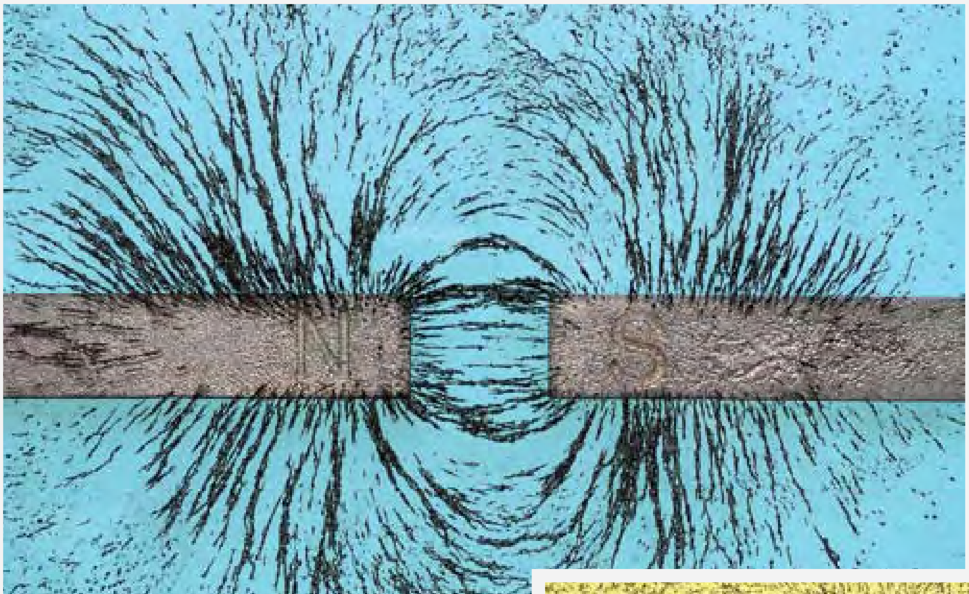
Here, however, the first puzzle of the study appears. The description of content learning in the United States indicates correctly that it is less intense and more varied than in China. But then it claims incorrectly that "scientific reasoning is not explicitly taught in schools in either country."*

Had this paper, with its generous online supplementation and other publications from the lead author's research group, failed to show awareness of the current research literature in K–12 science education, their claim that scientific reasoning is not being taught would have been understandable. And, thus understood by us, the study would simply have been ... dismissible. Why? Because it is not true that scientific reasoning is not taught in U.S. schools.

Scientific reasoning goes by different names, one of the most favored being "inquiry," as in "inquiry-based learning." This type of science study is so well established in the United States that a book-length *retrospective* and prospective account of inquiry-based science standards was published by the U.S. National Research Council nearly a decade ago.[6] One need only skim the most recent Fordham Institute study on state science standards to discover that scientific reasoning and "science process" skills, which focus on reasoning, are key elements of the expectations for student proficiency in nearly all of the 50 state standards reviewed.[7] The current Science Framework

---

*In the article, scientific reasoning is not simply subsumed under content; the authors' use of "content" implies that, for them, the word means something like *just the facts, ma'am*—with perhaps some very ad hoc concept juggling and problem solving.



MAGNETIC FIELD, © VISUALS UNLIMITED/CORBIS

MAGNETIC FIELD, © VISUALS UNLIMITED/CORBIS

for the National Assessment of Educational Progress reflects that preoccupation by dividing attention between science content and science practices. Of the latter, there are four, each preceded by an action verb: "identifying" or "using." The "using" statements are explicit reasoning skills.[8]

However, Bao et al.'s own print and online bibliographies *do* cite appropriate contemporary resources, indicating that they have at least come in contact with the evidence of inquiry-based learning in U.S. science classrooms. Hence, their statement that scientific reasoning is not taught in U.S. schools is not due to ignorance. It is just a misconception of current, standards-based science curricula nationwide, and of the associated literature.

Now the second puzzler appears. Although they refer to the physics courses, especially those taken by the Chinese students, as emphasizing "conceptual physics understanding and problem-solving skills," the researchers do *not*, apparently, include conceptual understanding and problem-solving skills within scientific reasoning ability. For this old subscriber to *Science*, such an exclusion is incomprehensible.

These lapses are regrettable because they create a flaw in the experimental design, the clarity of which depends upon the assumption that neither American nor Chinese K–12 science students receive special instruction in scientific reasoning. In reality, all available evidence indicates that both U.S. and Chinese students receive at least some instruction in scientific reasoning.

The authors believe that in these otherwise matched student groups, the students have a clear, large difference in exposure to and study of physics *content*. If this is so (and there is no reason to doubt it), and assuming that a good test of scientific reasoning *not* tied to content—i.e., not domain specific—is available, then it is possible to test for the impact of that difference in studying content on scientific reasoning ability. More specifically, it is possible to test for that holy grail of instruction, transferability. Transferability would mean that students become good scientific thinkers generally—that their reasoning transfers smoothly across all scientific subjects—instead of being limited to the specific areas they have studied. In the authors' words, they are interested in "domain-general reasoning skills such as the abilities to systematically explore a problem, to formulate and test hypotheses, to manipulate and isolate variables, and to observe and evaluate the consequences." So the ultimate question that Bao et al. undertake to answer is whether the Chinese students, recipients of prolonged and intense content instruction, are rendered thereby more adept at general scientific reasoning than the Americans, whose study of physics and other science content has been slight by comparison.

Performance in Physics and in Scientific Reasoning

To answer this question, Bao et al. employed three good tests: the Force Concept Inventory (FCI),[9] which assesses knowledge of introductory Newtonian mechanics; the Brief Electricity and Magnetism Assessment (BEMA), which assesses understanding of electricity (including circuits) and magnetism; and the Lawson Classroom Test of Scientific Reasoning (LCTSR), which is supposed to assess capacity for general scientific reasoning (that is, with minimal domain dependence). To the authors' credit, the *Science* print article and its online supplements together provide adequate detail on the tests, the testing, and their results.

Outcomes of the tests are clear enough in the article and supplements. On the FCI, Chinese students performed very well, with a narrow distribution of scores centered on an impressive mean of 86 percent. The American scores were much more broadly distributed around a mean of 49 percent—distinctly failing. On the BEMA, Chinese students scored at a mean of 66 percent, but the Americans scored at a mean of 27 percent—not much better, Bao et al. note, than would have been produced by randomly choosing answers to the test questions. These tests distinguished the two populations of test takers, one well prepared in physics, the other not.*

So far, no surprises. These results look like those of recent international assessments in science and mathematics, in which the performance of U.S. students, especially in the higher grades, is at best undistinguished and sometimes awful.

The results of testing scientific reasoning with the LCTSR, however, were surprising (to me). Both groups showed a mean of 74 percent and their score distributions were effectively identical.† Such results *should* be surprising, at least to many *Science* readers; but the authors, instead of being surprised and questioning the results, conclude that they have a substantive finding regarding scientific reasoning,

*Numbers of test takers in all cases were large enough for there to be no doubt that the calculated means are properly representative.

worthy of their most important comments and recommendations. They believe the results indicate that content instruction, in physics anyway, *cannot* inculcate good scientific reasoning abilities and habits. More study of content leads only to more "content knowledge," not to that higher-level, general competence in science that is so eagerly sought.

I am sorry that the authors were not surprised by their findings. Had they been surprised, they might have questioned their immediate response to the data and considered alternative conclusions. The job of considering alternatives, then, is left to others.

What Do the Scores Mean?

Let's set aside, for the moment, our earlier concern about whether U.S. and Chinese students are actually taught scientific reasoning in an explicit way, and take the information presented by the authors at face value. It indicates that the training Chinese students receive before coming to college includes much practice with important concepts of physics and with skills needed to solve physics problems. Tested at the end of this period for knowledge of two central physics topics, the Chinese students perform handsomely. Not only are they ready for calculus-based college physics, but they can be said, in all justice, to know physics, at least the physics taught in high school. For students in the United States, the situation is essentially the opposite. Only a third or so of them take high school physics. The rest learn physics, if at all, from the general science of grades K–8 and via the (derived) physics components of other science disciplines, such as biology, chemistry, Earth science, or environmental studies. These students perform poorly on the physics tests. They cannot be said to know physics.

Now, both cohorts are tested with the LCTSR for their ability to think about very simple natural (i.e., scientific) situations, for example: explaining the results of filling graduated cylinders of differing diameter with the same volume of liquid, and vice versa. The test questions have mainly to do with logic and efficient thinking. On such a test, both cohorts perform at a solid average level; and what's more, the population score distributions are essentially the same. What is going on?

Bao et al. conclude that even though the Chinese students know physics content, their scientific reasoning is *no better than* that of the American students. As for scientific reasoning that is transferable and immediately usable in real-world problems, the authors evidently believe, Chinese students are no better equipped than those content-challenged[10] U.S. students.

But this is not a necessary, or even the most likely, conclusion. A more likely one is that the LCTSR is testing the students' reasoning about certain simple *but unfamiliar* natural situations. So, it requires all the test takers, Chinese and American, to rely on the same relatively slow, relatively inefficient kind of thinking.

The findings of cognitive science tell us that domain knowledge strongly affects the quality of thinking. Specifically, its accuracy, speed, and efficiency—manipulating information in working memory—are much improved when relevant, quickly recoverable knowledge (procedural as well as factual) is stored in long-term memory. So, if you want to solve physics problems quickly and efficiently, you'll need a good bit of factual and procedural physics knowledge stored in your long-term memory. How is such knowledge stored in long-term memory? By solving physics problems! Bit by bit, you tackle more and more complex problems, and eventually you have in long-term memory a rich domain of physics facts, procedures, and tricks of thought about concepts of physics and physicslike problems.

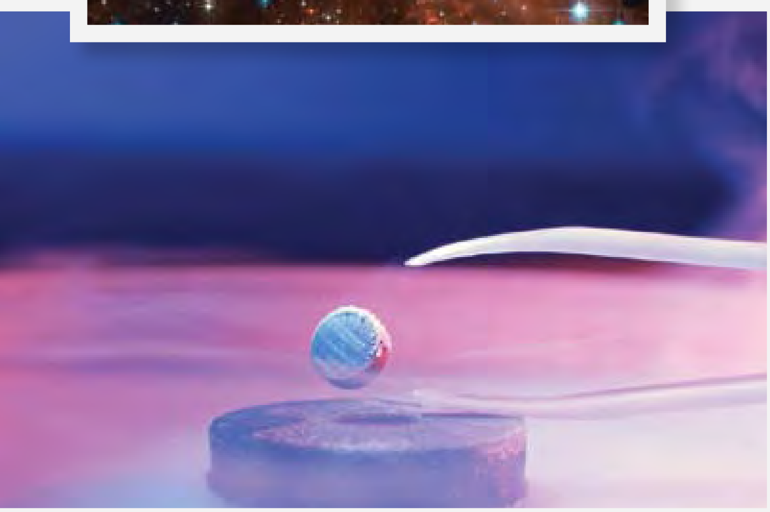
Faced, then, with a new problem in physics, you ordinarily will retrieve examples of correct solutions to similar problems encountered earlier—not the primitive steps of the required solution (which from practice have become automatic for you, like number facts in arithmetic). You will automatically import from long-term memory into working memory whole chunks of problem types and solutions, and thus will be able to grab quickly the appropriate one for application to the new problem. Your thinking (or reasoning) will be efficient. You will do well in a written or

PLASMA BALL, ©ISTOCKPHOTO.COM/SDOURADO

PLASMA BALL, ©ISTOCKPHOTO.COM/BORSHEIM

†The possibility that this result was due to a ceiling effect occurs immediately to any test-hardened teacher, and these energetic authors did not fail to consider it. Online supplementary material includes their independent investigation of that possibility, with the result that no ceiling effect is a likely cause of the near-identical, 74 percent mean scores for Chinese and American college freshmen who are preparing to major in science or engineering.

NEBULA, PHOTO COURTESY OF NASA, ESA, AND THE HUBBLE HERITAGE (STSCI/AURA)-ESA/HUBBLE COLLABORATION



SUPERCONDUCTIVITY, © LAWRENCE MANNING/CORBIS

a real-world test.[11] Reasoning works with content!

Here, then, is an alternative view of the Bao et al. results. The Chinese students *know* physics. The American students don't. Now both groups face a *different* challenge—different enough from the standard physics problems so that the Chinese students' superior conceptual and problem-solving skills in physics provide no immediate advantage. The new challenge is to think about problems of a very simple scientific character, but in forms and subject-matter domains that neither group has encountered before. As the authors explain in their online supplementary materials, the LCTSR "measures fundamental reasoning components with simple context scenarios that do not require complex content understanding. This test design can improve the measurement of the basic reasoning abilities by reducing the possible interference from understandings of content knowledge." But if so, both cohorts will handle most of the questions on the LCTSR (or any challenge like it) *the same way*: they will need to think through each question from scratch—to find the best answer starting from elementary principles. That kind of thinking is slower and more error-prone than the thinking available to a physics-savvy Chinese student taking the FCI or the BEMA.

There is one remote possibility to consider. Going back to the first puzzle, suppose that, contrary to a crucial assumption of the authors, the American students *do* receive considerable instruction in what they call scientific reasoning, and (as the authors claim) the Chinese students do not. That could, in principle, account for the Americans performing well enough to match the performance of the Chinese. But any such explanation seems extremely unlikely, given the remarkable congruence of the LCTSR results of both groups. And, if the Chinese students had really received no scientific reasoning instruction, we would expect the Americans, who *have* been taught scientific inquiry, to do much better on the LCTSR than the Chinese. They did not.

That, of course, raises the possibility hinted at by the second puzzle. It could be that *both* the U.S. and Chinese students receive instruction in scientific reasoning. Bao et al. may not define it that way, but an "emphasis on conceptual physics understanding and problem-solving skills," which is how they characterize the Chinese instruction, sounds to me like plenty of emphasis on reasoning about science—and about much else! So it may be that both the Chinese intensive approach and the American nonintensive approach are equally effective—or *equally ineffective*—in teaching the domain-independent reasoning the LCTSR is supposed to test.

One final possibility is a rather unhappy one, but perhaps the most realistic. It could be that, because the students were matched in every relevant characteristic except physics "content" instruction, these two large student groups, Chinese and American, have simply reached the same level of general reasoning ability (or have the same average IQ).[12] So the LCTSR, with its general reasoning questions, is simply establishing a good control (a proper isolation of variables)—that is, these two groups of rather well-matched students are of about the same general cognitive ability. And, of course, that would be comforting in some ways, but also no surprise.

Whichever conclusion(s) may be correct, what we can say with confidence is that these Chinese students learned enough physics in school. The U.S. students—who, having opted already for science, technology, engineering, and mathematics majors in college, are among our best science students—have not learned enough. That should be a big worry, and not only because, as we saw at the outset, reasoning devoid of content can prompt a chuckle or two. ☐

Endnotes

1. For these and other examples of students' misunderstandings (along with explanations), see www.britishcouncil.org/learnenglish-central-stories-exams.htm.

2. Susan Haack, *Defending Science—Within Reason: Between Scientism and Cynicism* (Amherst, NY: Prometheus Books, 2003), 24.

3. Daniel T. Willingham, "Critical Thinking: Why Is It So Hard to Teach?" *American Educator* 31, no. 2 (Summer 2007): 8–19, www.aft.org/pubs-reports/american_ educator/issues/summer07/Crit_Thinking.pdf.

4. Lei Bao et al., "Learning and Scientific Reasoning," *Science* 323 (2009): 586–587. Supporting online materials at: www.sciencemag.org/cgi/content/full/323/5914/586/DC1.

5. Bao et al., "Learning and Scientific Reasoning," 587.

6. National Research Council, *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning* (Washington, DC: National Academy Press, 2000).

7. Paul R. Gross et al., *The State of State Science Standards* (Washington, DC: Thomas B. Fordham Institute, 2005), www.edexcellence.net/doc/Science%20Standards.FinalFinal.pdf.

8. Science Framework for the 2009 National Assessment of Educational Progress (Washington, DC: National Assessment Governing Board, 2009).

9. See this interesting use of the FCI in connection with reasoning ability, scientific and in general: Vincent P. Coletta and Jeffrey A. Phillips, "Interpreting FCI Scores: Normalized Gain, Preinstruction Scores, and Scientific Reasoning Ability," *American Journal of Physics* 73, no. 12 (2005): 1172–1182. This paper includes a copy of the Lawson test of scientific reasoning.

10. Politically correct locution.

11. A recent, accessible treatment of these issues, with sufficient references to the literature of cognitive science, is: Daniel T. Willingham, *Why Don't Students Like School?: A Cognitive Scientist Answers Questions about How the Mind Works and What It Means for the Classroom* (San Francisco: Jossey-Bass, 2009).

12. Coletta and Phillips, "Interpreting FCI Scores," 1177.