Why You Should Measure Your Students' Reasoning Ability

Vincent P. Coletta, Jeffrey A. Phillips, Loyola Marymount University, Los Angeles, CA
Jeffrey J. Steinert, Edward Little High School, Auburn, ME

Many teachers administer a force concept test such as the Force Concept Inventory[1,2] (FCI) to their students in an effort to evaluate and improve their instructional practices. It has been commonly assumed that looking at class normalized gains allows teachers to compare their courses with other courses. In this paper we present evidence to suggest that the use of class normalized gains alone may not provide a complete picture. We argue that student reasoning ability should also be assessed before between-course comparisons can be made. Assessment of reasoning ability is also useful in identifying students who are at risk. In the following we shall concentrate on the FCI, but we think our conclusions probably apply to physics concept tests generally.

Normalized FCI Gain as a Measure of Learning

The Force Concept Inventory (FCI) is often used as a measure of students' understanding of basic concepts in Newtonian mechanics and also to assess teaching effectiveness in an introductory mechanics class. This 30-question multiple-choice exam is research-based, with incorrect answers reflecting common student misconceptions. When the test is given both at the beginning and end of an introductory mechanics high school or college course, a student's pre- and post-instruction scores can be used as a measure of conceptual learning achieved during the course. Pre-instruction scores vary widely among the students in a typical introductory class. In order to compare the learning achieved for students with quite different pre-instruction scores, a useful measure is the normalized gain[3-5] G:

$$G = \frac{postscore\% - prescore\%}{100 - prescore\%}.$$

Thus G is just the actual change divided by the maximum possible gain. For example, using this measure, the normalized conceptual gains of students with pre → post scores of 20% → 60%, 40% → 70%, and 80% → 90% all correspond to $G = 0.5$. Loosely speaking, G is the fraction of the concepts that a student learns that were not already known at the start of the course. It should be emphasized that G is the *single student* normalized gain and is *not* the same as Hake's[4] *normalized gain* $<g>$, obtained from the class *averages* of pre-test and post-test scores. Hake's footnote 46 discusses the mathematical relationship of $<g>$ to the class average of individual students' G's, and states that the two are usually within 5%.

The interpretation of G as a measure of learning, independent of a student's initial state of knowledge, is justified by the fact that when other important factors such as reasoning ability are either accounted for or averaged over, students' normalized gains are not correlated with pre-instruction scores. For example, in a study of 12,000 high school students' FCI scores, Hestenes[6] found that there was no significant correlation between G and FCI prescore (correlation coefficient $r = 0.00$). However, in college introductory mechanics courses, G is often positively correlated with prescore.[7] We think that this is not because higher

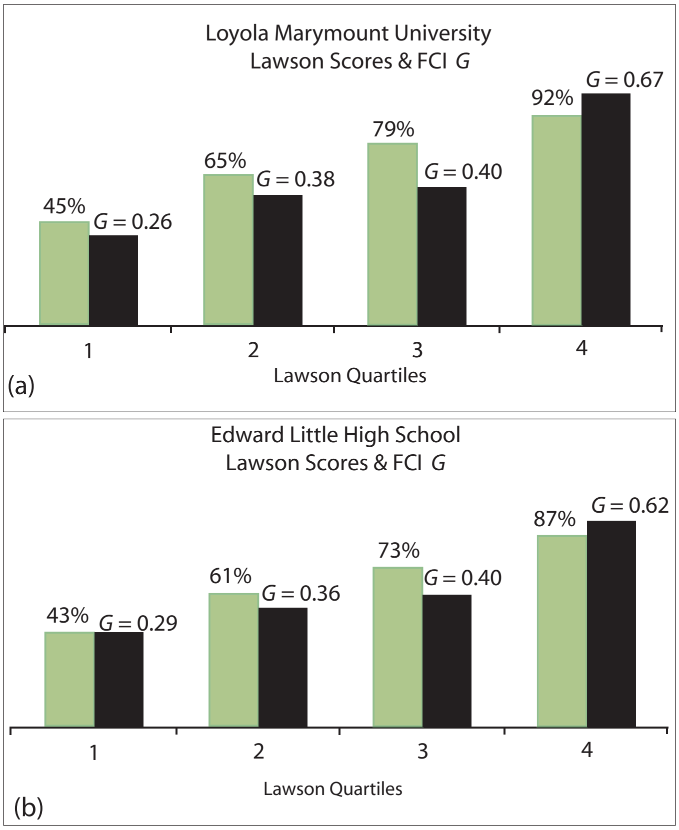

Fig. 1. a) Average normalized gains on the Force Concept Inventory for different populations of LMU students. Standard errors in Lawson scores in each quartile ranged from 1% to 3%. b) Average normalized gains on the Force Concept Inventory for different populations of Edward Little High School students. Standard errors in the *G*'s for each quartile ranged from 0.02 to 0.03. Standard errors in Lawson scores in each quartile ranged from 1% to 2%.

prescores tend to *cause* higher *G*'s but rather because in college classes both high prescores and high *G*'s tend to be achieved by those students with the strongest reasoning abilities. Higher prescores may be a reflection of the greater conceptual learning achieved by stronger reasoners in their high school physics courses, and higher *G*'s may be achieved by stronger reasoners in their college courses. Thus we conjecture that the correlation between *G* and prescore in many college classes is a product of a correlation between conceptual learning and reasoning skills. Our more detailed interpretation of the relationship between prescore, normalized gain, and reasoning ability may be found in our *American Journal of Physics* article.[7]

Interactive Engagement and Traditional Courses

Interactive engagement (IE) is the name used to describe a broad array of nontraditional instructional methods for teaching introductory physics that have been developed in recent years. These include Peer Instruction,[8] Washington Tutorials in Physics,[9] Modeling,[10] Interactive Lecture Demonstrations,[11] Workshop Physics,[12] and many others. The common feature of all these methods is that they require the active participation of students in class, rather than the traditional approach of listening to a lecture and taking notes. Research demonstrates that IE methods are consistently more effective than traditional methods.[3,13] Traditional courses consistently result in class average *G*'s of only about 0.2, whereas IE classes produce consistently higher average *G*'s, typically in the range 0.3 to 0.6.

Normalized FCI Gain and Reasoning Ability

Why the wide range of class average *G*'s for IE classes? One obvious reason is that some interactive methods may be more effective than others. Another possibility is that some individual instructors may be more effective than others. However, there remains a third possibility, one that has been largely unrecognized, that we think may often be the most significant reason for the observed range of values of *G*, namely the nature of the population of students in a given class in introductory mechanics.

Reasoning ability varies widely among individual students, and the average reasoning ability of entire classes can also vary significantly. A few others, including Epstein,[14] Henderson et al.,[15] Clement,[16] Meltzer,[5] and Hake[4] have suggested that population effects may be important. In particular: Jerry Epstein suggested that the "0.7 barrier" in <*g*> observed by Hake[3] could be due to students in physics classes "who have nowhere near the basic skill and cognition levels to benefit from a sound 'Interactive-Engagement' program"; John Clement reported observing a correlation between reasoning level and conceptual gain in physics; David Meltzer suggested reasoning ability as a hidden variable that could affect *normalized* gains on the Conceptual Survey in Electricity and Magnetism; and Charles Henderson et al. and Richard Hake[4] reported correlations between FCI *G*'s and prescore.

Our research[7] suggests that the reasoning ability of students, as measured by Lawson's Classroom Test of Scientific Reasoning Ability,[17] may account for much of the observed variation in *G* among individuals in any particular class and also for some of the variation in class average *G*'s among different IE classes. Figures 1(a) and 1(b) show the results of our study of normalized FCI gains and Lawson Test scores for 98 students in various IE introductory mechanics classes at Loyola Marymount and for 199 students in physics classes at Edward Little High School. Of the 98 LMU students, 42 were taught by one of us (Coletta) using a method in which each chapter is covered first in a "concepts" class, in a Socratic style very similar to Peer Instruction, and then again in a "problems" class. Another author (Phillips) taught 23 students in lectures, interspersed with small group activities, using conceptual worksheets, short experiments, and context-rich problems. The other 33 students were taught by LMU professors Bulman and Sanny, who both lecture with a strong conceptual component and with frequent class dialogue. Some of the classes were calculus based, primarily composed of engineering majors; others were noncalculus based with mostly biology and natural science majors. All of the 199 Edward Little High School students were taught by one of us (Steinert) in algebra-based regular or honors physics classes using Modeling Instruction.

Figures 1(a) and 1(b) are remarkably similar. Both show that students in the top Lawson quartile have much higher normalized gains than those in the lowest quartile, 0.67 versus 0.26 for LMU students and 0.62 versus 0.29 for Edward Little High School students. The correlation coefficients between Lawson scores and normalized FCI gain are 0.51 and 0.53, respectively for the LMU and Edward Little populations. The LMU data show an equally strong correlation between Lawson score and FCI prescore (correlation coefficient $r = 0.53$), supporting our conjecture that the observed correlation between FCI prescore and FCI *G* ($r = 0.33$) is a result of students with stronger reasoning ability achieving greater learning both in high school and college classes.

Based on Fig. 1, if a class had an average Lawson Test score below 50%, it would be reasonable to expect a class average normalized FCI gain of 0.3 or less, but if the class had an average Lawson Test score of 90% or more, one might expect a class average normalized gain of 0.6 or more. Because reasoning ability appears to be such an important factor in determining *G*, we believe that it makes no sense to measure *G* alone without somehow taking into account the reasoning ability of a class. It would be a far greater accomplishment to achieve a class average *G* of 0.5 in a class in which Lawson test scores were nearly all below 50% than to get a class average *G* of 0.7 in a class in which Lawson's test scores were nearly all above 90%.

Perhaps the most important reason for assessing reasoning ability is that it enables us to identify students who are at greatest risk in introductory physics. To the extent we can do this, we open up the possibility of focusing special instructional efforts on a subset of the class. We are currently engaged in designing instructional materials to develop reasoning ability, based on the work of Adey and Shayer,[18,19] Feuerstein et al.,[20,21] Karplus and colleagues,[22,23] and others.

A question apparently not addressed by physics education researchers is whether or not an introductory physics class can substantially affect students' reasoning ability as measured by pre/post Lawson testing, as has been reported by Wyckoff (2001) for a biology course.

As physicists we describe the conditions of laboratory measurements as completely as possible. Our measurements of student learning should similarly be accompanied by as complete as possible description of the conditions accompanying those measurements. Ignoring the reasoning ability of our students when we try to measure the learning they achieve in our courses means that we are ignoring a variable that can have a huge effect on our data, which in turn can mislead us about how effective we are in teaching our students.

References

1. Physics Education R&D Group, "Assessment instrument information page" (North Carolina State University, 2005); online at http://www.ncsu.edu/per/TestInfo.html.
2. D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory," *Phys. Teach.* **30**, 141–158 (March 1992). The 1995 revision by Halloun, Hake, Mosca, & Hestenes, which was used in this study, is available online [password protected] at http://modeling.asu.edu/R&E/Research.html.

3. R.R. Hake, “Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,” *Am. J. Phys.* **66**, 64–74 (Jan. 1998); online at http://www.physics.indiana.edu/~sdi/ajpv3i.pdf.
4. R.R. Hake, “Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization,” submitted to the Physics Education Research Conference (Boise, ID, Aug. 2002); online at http://www.physics.indiana.edu/~hake/PERC2002h-Hake.pdf.
5. D.E. Meltzer, “The relationship between mathematics preparation and conceptual learning gains in physics: A possible ‘hidden variable’ in diagnostic pretest scores,” *Am. J. Phys.* **70**, 1259–1268 (Dec. 2002); online at http://www.physicseducation.net/docs/AJP-Dec-2002-Vol.70-1259-1268.pdf.
6. D. Hestenes, private communication, August 2005.
7. V.P. Coletta and J.A. Phillips, “Interpreting FCI scores: Normalized gain, pre-instruction scores, and scientific reasoning ability,” *Am. J. Phys.* **73**, 1172–1182 (Dec. 2005); online at http://scitation.aip.org/dbt/dbt.jsp?KEY=AJPIAS&Volume=73&Issue=12.
8. E. Mazur, *Peer Instruction: A User’s Manual* (Prentice Hall, Upper Saddle River, NJ, 1997); online as a 433-kB pdf at http://mazur-www.harvard.edu/publications.php?function=search&topic=8.
9. L.C. McDermott and P.S. Shaffer, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 2002).
10. M. Wells, D. Hestenes, and G. Swackhamer, “A modeling method for high school physics instruction,” *Am. J. Phys*. **63**, 606–619 (July 1995); online at http://modeling.asu.edu/R&E/Research.html.
11. D.R. Sokoloff and R.K.Thornton, *Interactive Lecture Demonstrations in Introductory Physics* (Wiley, New York, 2004).
12. P.W. Laws, *Workshop Physics Activity Guide Modules* (Wiley, New York, 2004).
13. E.F. Redish and R.N. Steinberg, “Teaching physics: Figuring out what works,” *Phys. Today* **52**, 24–30 (Jan. 1999); online at http://www.physics.umd.edu/perg/papers/redish/pt.htm.
14. J. Epstein, “The ‘0.7 Barrier’ on the FCI—A Suggestion of the underlying problem and a proposal for further research,” Physics Education Research Conference 2000: Teacher Education (Univ. of Guelph, August 2–3, 2000), online at http://www.sci.ccny.cuny.edu/~rstein/percpaps/epstein.pdf.
15. C.R. Henderson, K. Heller, and P. Heller, “Common concerns about the FCI,” AAPT *Announcer* **29**, 99 (1999); online at http://groups.physics.umn.edu/physed/Talks/talks.html.
16. J. Clement, “The correlation between scientific thinking skill and gain in physics conceptual understanding,” AAPT *Announcer* **31**, 82 (2001).
17. A.E. Lawson, “The development and validation of a classroom test of formal reasoning,” *J. Res. Sci. Teach*. **15**, 11–24 (1978). An updated multiple-choice version of the test is in the appendix of Ref. 7.
18. *Really Raising Standards: Cognitive Interventions and Academic Achievement,* edited by P. Adey and M. Shayer (Routledge, London, 1994).
19. *Learning Intelligence: Cognitive Acceleration Across the Curriculum from 5 to 15 Years*, edited by M. Shayer and P. Adey (Open University Press, 2002).
20. R. Feuerstein, *Instrumental Enrichment: An Intervention Programme for Cognitive Modifiability* (University Park Press, Baltimore, MD, 1980).
21. *Mediated Learning Experience (MLE):Theoretical, Psychosocial and Learning Implications,* edited by R. Feuerstein, P.S. Klein, and A.J. Tannenbaum (Freund Publishing House, London, 1994).
22. B. Kurtz and R. Karplus, “Intellectual development beyond elementary school VII: Teaching for proportional reasoning,” *School Sci. Math.* **79**(5), 387-398 (May–June 1979).
23. “Central Role of Students’ Reasoning,” in *A Love of Discovery: Science Education—The Second Career of Robert Karplus,* edited by R.G. Fuller (Kluwer, 2002), Chap. 6.

PACS codes: 01.40.F, 01.40.Gm

Vincent Coletta *is professor of physics and department chair at Loyola Marymount University. He is author of* College Physics*, published in 1995, and soon to be re-published electronically by Physics Curriculum and Instruction. His doctoral thesis at the University of Notre Dame was in the area of statistical mechanics. Currently he is active in Physics Education Research.*

Loyola Marymount University, Los Angeles, CA 90045-2659; vcoletta@lmu.edu

Jeffrey A. Phillips *is an assistant professor of physics at Loyola Marymount University. His condensed matter interests include nanotribology and liquid helium films. He currently is studying how students' scientific reasoning skills and views about reasoning impact their success in introductory physics.*

Loyola Marymount University, Los Angeles, CA 90045-2659; jphillips@lmu.edu

Jeff Steinert *teaches chemistry and physics at Arizona School for the Arts. His interest in Physics Education Research stems from his involvement in the Modeling Workshop project at Arizona State University. He has led numerous summer workshops for high school and college physics instructors in Maine and Florida*

steinert@goasa.org