CROSS-FERTILIZATION UPDATE

What Might Psychologists Learn From Scholarship of Teaching and Learning in Physics?

Richard R. Hake
Indiana University – Bloomington

In this article I (a) note that psychologists have made outstanding contributions to the Scholarship of Teaching and Learning, (b) bemoan the failure of psychologists to develop a Psychology Concept Inventory, (c) point to evidence from pre/post testing with the Force Concept Inventory for the approximately 2-SD superiority in average pre- to post-course normalized gains in conceptual understanding for interactive engagement over traditional passive-student lecture courses, (d) provide crucial operational definitions of “average normalized gain,” “interactive engagement methods,” “traditional methods,” and “direct instruction,” (e) list 14 hard-won lessons from SoTL in physics that may be of value to psychologists, and (f) conclude that it is conceivable that psychologists might learn something from SoTL in physics.

Keywords: concept inventory, interactive engagement, passive-student lecture, SoTL

It is common knowledge that psychologists have made important contributions to the Scholarship of Teaching and Learning (SoTL); for example, see the contributions of Halpern, Shavelson, and Sternberg as indicated at the respective Wikipedia entries at http://bit.ly/JMdE6s, http://bit.ly/VvmqKa, and http://bit.ly/1a4GqFB. In addition, McCloskey’s (1983) seminal study of misconceptions regarding “bodies in motion” startled physics educators by showing that even after an introductory course on Newtonian mechanics many students retained naive beliefs about motion. In their pioneering article, physicists Halloun and Hestenes (1985a) paid tribute to the foundational work of McCloskey and his collaborators (Caramazza, McCloskey, & Green, 1981; McCloskey, Caramazza, & Green, 1980).

The Failure of Psychologists to Develop a Psychology Concept Inventory

Despite the above outstanding SoTL contributions by psychologists, contemporary psychology education researchers have evidently failed to develop a Psychology Concept Inventory comparable with the prototype Force Concept Inventory (FCI) of physics (Halloun, Hake, Mosca, & Hestenes, 1995; Hestenes, Wells, & Swackhamer, 1992). The FCI was developed through extensive qualitative and quantitative research on students’ conceptions of “bodies in motion.” It has been administered as a pre- and post-course test to thousands of students in hundreds of introductory undergraduate physics courses so as to gauge the effectiveness of both traditional and reform courses in promoting conceptual understanding.

Considering that Steuer and Ham’s (2008) survey of introductory psychology text sales can be used to estimate that between 1.2 million to 1.6 million undergraduate students complete
an introductory or general psychology course each year, it appears that psychologists are missing a great opportunity to definitively evaluate the effectiveness of their introductory courses by pre- and post-course testing of thousands of students in at least hundreds of courses, as has been accomplished in physics with its approximately 0.61 million undergraduate students per year enrolled in introductory physics courses as indicated by the American Institute of Physics (2013).

My cursory literature search located only two tests of psychology understanding that might be given to introductory course students: Vaughan’s (1977) Test of Common Beliefs and McCutcheon’s (1991) New Test of Misconceptions about Psychology. But these are evidently rarely used by psychologists to evaluate their teaching. At the very least, because psychology is often taken by incoming university students to fulfill a science requirement, it might be interesting for psychologists to administer a Nature of Science Assessment (Lederman, Bartos, & Lederman, 2014) as a pre- and post-course test.

Despite the dearth of formative evaluation of introductory psychology courses, there has been some discussion of such evaluation (Hake, 2007b, 2007c, 2008b; Holton, 2007; Lamal, 1979; Lilienfeld, 2010; Wittmann, 2007). Here “formative evaluation” means evaluation “designed and used to improve an intervention, especially when it is still being developed”—see “Two Different Meanings of ‘Formative Evaluation’ #2” (Hake, 2014). I argue that there is nothing to prevent the development in any discipline of interactive engagement methods similar to those found in physics to yield average normalized gains about two standard deviations greater than those produced by traditional passive-student lecture courses. Holton (2007) suggested that the apparent failure of psychologists to engage in such development, even despite their long history of major contributions to SoTL can be attributed to a lack of funding. But in my opinion, the lack of substantive discussion in the psychology literature of the measurement and enhancement of learning gains in introductory psychology courses suggests that psychologists, with a few notable exceptions, are simply not interested in gauging the effectiveness of their undergraduate courses in enhancing student higher-order learning.

In “Should We Measure Change? Yes!” (Hake, 2011), I stated that formative pre/post testing is being successfully used to improve the effectiveness of courses in undergraduate astronomy, biology, chemistry, economics, engineering, geoscience, and physics. But such testing is still anathema to many members of the psychology education community. I argue that this irrational bias impedes a much needed enhancement of student learning in higher education. Lloyd Bond (2005), a senior scholar at the Carnegie Foundation, wrote:

If one wished to know what knowledge or skill Johnny has acquired over the course of a semester, it would seem a straightforward matter to assess what Johnny knew at the beginning of the semester and reassess him with the same or equivalent instrument at the end of the semester. It may come as a surprise to many that measurement specialists have long advised against this eminently sensible idea. Psychometricians don’t like ‘change’ or ‘difference’ scores in statistical analyses because, among other things, they tend to have lower reliability than the original measures themselves. Their objection to change scores is embodied in the very title of a famous paper by Cronbach and Furby (1970), “How Should We Measure Change, or Should We?”

Cronbach and Furby’s (1970) dour appraisal of pre/post testing has echoed down through the literature to present-day texts on assessment such as that by Suskie (2009). In my opinion, such pre/post paranoia and its attendant rejection of pre/post testing in evaluation, as used so successfully in physics education reform (Hake, 2005, 2008b), is one reason for the glacial progress of educational research (Berliner, 2006; Lagemann, 2000). [For an antidote to the anti-pre/post testing opinions of Linda Suskie see the pro-pre/post testing opinions of Peggy Maki (2010).]

Psychologists have, for the most part, ignored SoTL in physics (for an exception, see Lattery, 2008). This despite the following: (a) the widely acknowledged need for reform of higher education—see for example Arum and Roksa (2011), Delbanco (2012), Hacker and Dreyfus (2010), Keeling and Hersh (2011), and Selingo (2013); (b) the evidence for the superiority of interactive engagement over traditional passive-student lecture methods in physics; and (c) accolades for SoTL in physics from biologists (Klymkowsky, Garvin-Doxas, & Zeilik, 2003; Wood & Gentile, 2003), economists (Simkins & Maier, 2008), and mathematicians (Bressoud, 2012).
As for the unreliability of change scores, such charges by Lord (1956, 1958) and Cronbach and Furby (1970) have been called into question by Hake (2011) as well as Rogosa (1995), Rogosa, Brandt, & Zimowski (1982), Rogosa & Willett (1983, 1985), Wittmann (2007), Zimmerman (1997), Zimmerman & Williams (1982), and Zumbo (1999). Furthermore, the measurement of change is an active area of research by psychometricians such as Collins and Horn (1991), Collins and Sayer (2001), Lissitz (2005), Liu and Boone (2006), and Singer and Willett (2003). All of this work should serve as a caution for those who dismiss measurements of change.

In my opinion, without mass pre/post testing there can be little understanding of the need for, or the results of, various types of reform pedagogies or curricula. I regard the apparent failure of psychologists to research the effectiveness of their own introductory courses as an important issue in education research because, among other things:

1. One might expect psychologists with their long history of education research (Berliner, 2006; Freedheim & Weiner, 2004; Lagemann, 2000) and their leading role in classroom-oriented ‘design-based research’ (Hake, 2008a; Kelly, 2003; Kelly, Lesh, & Baek, 2008) to be in the vanguard of those actively researching the effectiveness of their own courses and thus serving as role models for other faculty.

2. Educational psychologists often staff the “Teaching and Learning Centers” of U.S. universities and thus might (but generally do not) influence faculty to research the cognitive effectiveness of their courses through valid and consistently reliable diagnostic tests developed by disciplinary experts, rather than through the usual problematic (Hake, 2002b) student evaluations of teaching.

3. Psychologists and psychometricians seem to be in control of the U.S. Department of Education’s “What Works Clearinghouse” and “No Child Left Behind.” Why should they be the arbiters of the “What Works Clearinghouse” when, as far as I know, they have not even bothered to research what works in their own courses?

Evidence for the Superiority of Interactive Engagement Over Traditional Passive-Student Lecture Methods in Physics

Some definitions (Hake, 1998a, 1998b) are in order:

1. “Interactive Engagement” (IE) methods are defined operationally as “those designed at least in part to promote conceptual understanding through active engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback to both students and instructors through student discussion with peers and/or instructors.” The feedback to instructors facilitates formative evaluation in the sense used by Black and Wiliam (1998) and Shavelson (2008); that is, “all those activities undertaken to provide information to be used as feedback so as to adapt the teaching to meet student needs” (see “Two Different Meanings of ‘Formative Evaluation’ #2”; Hake, 2014).

2. “Traditional” methods are defined operationally as “those which make little or no use of interactive engagement methods, relying primarily on passive-student lectures, recipe laboratories (in which detailed and explicit procedures must be followed), and algorithmic problem examinations”—this is what is known to most physicists (but not to most cognitive scientists) as “direct instruction.”

For Newtonian mechanics, physics education researchers have demonstrated that interactive engagement methods can produce a roughly two-standard-deviation superiority in average normalized pre- to post-course learning gains \(g\) over traditional passive-student lecture methods (Hake, 1998a, 1998b). Similar differences in \(g\) between IE and traditional courses had been reported in at least 25 other peer-reviewed publications (Hake, 2008a). That research involves the measurement of pre- to post-course learning gains on valid and consistently reliable multiple-choice concept inventories developed by disciplinary experts (Halloun & Hestenes, 1985a, 1985b; Hestenes et al., 1992; Thornton & Sokoloff, 1998)—and the use of reasonably well-matched control groups provided by traditional introductory courses. The
abstract of “Interactive-Engagement Versus Traditional Methods: A Six-Thousand Student Survey of Mechanics Test Data for Introductory Physics Courses” (Hake, 1998a) reads:

A survey of pre/post test data using the Halloun-Hestenes Mechanics Diagnostic test or more recent Force Concept Inventory is reported for 62 introductory physics courses enrolling a total number of students \( N = 6542 \). A consistent analysis over diverse student populations in high schools, colleges, and universities is obtained if a rough measure of the average effectiveness of a course in promoting conceptual understanding is taken to be the average normalized gain \( g \). The latter is defined as the ratio of the actual average gain \( \langle g \rangle = \langle \% \text{post} \rangle - \langle \% \text{pre} \rangle \rangle \) to the maximum possible average gain \( (100 - \langle \% \text{pre} \rangle) \). Fourteen "traditional" (T) courses (\( N = 2084 \)) which made little or no use of interactive-engagement (IE) methods achieved an average gain \( \langle g \rangle = 0.23 \pm 0.04 \) (std dev). In sharp contrast, 48 courses (\( N = 4458 \)) which made substantial use of IE methods achieved an average gain \( \langle g \rangle = 0.48 \pm 0.14 \) (std dev), almost two standard deviations of \( \langle g \rangle \) IE-ave above that of the traditional courses.

Results for 30 (\( N = 3259 \)) of the above 62 courses on the problem-solving Mechanics Baseline test of Hestenes & Wells imply that IE strategies enhance problem-solving ability. The conceptual and problem-solving test results strongly suggest that the classroom use of IE methods can increase mechanics-course effectiveness well beyond that obtained in traditional practice.

But wait—can multiple-choice tests measure conceptual understanding and higher-order learning? (For a cogent discussion of higher-order learning, see Shavelson & Huang, 2003). Wilson and Bertenthal (2005) think so, writing the following:

Performance assessment is an approach that offers great potential for assessing complex thinking and learning abilities, but multiple choice items also have their strengths. For example, although many people recognize that multiple-choice items are an efficient and effective way of determining how well students

---

Table 1
Lessons From the Physics Education Reform Effort

Six lessons on interactive engagement
- The use of interactive engagement strategies can increase the effectiveness of conceptually difficult courses well beyond that obtained by traditional passive-student lecture methods.
- The use of interactive engagement and/or high-tech methods, by themselves, does not ensure superior student learning.
- High-quality standardized tests of the cognitive and affective impact of courses are essential to gauge the relative effectiveness of non-traditional educational methods.
- Education research and development (R&D) by disciplinary experts, and of the same quality and nature as traditional science/engineering R&D, is needed to develop potentially effective educational methods within each discipline. But the disciplinary experts should take advantage of the insights of disciplinary experts doing education R&D in other disciplines, cognitive scientists, faculty and graduates of education schools, and classroom teachers.
- The development of effective educational methods within each discipline requires a redesign process of continuous long-term classroom use, feedback, assessment, research analysis, and revision.
- Although nontraditional interactive engagement methods appear to be much more effective than traditional methods, there is need for more research to develop better strategies for enhancing student learning.

Eight lessons on implementation
- Teachers who possess both content knowledge and “pedagogical content knowledge” are more apt to deliver effective instruction.
- College and university faculty tend to overestimate the effectiveness of their own instructional efforts and thus tend to see little need for educational reform.
- Such complacency can sometimes be countered by administering high-quality standardized tests of understanding and by “video snooping.”
- A major problem for undergraduate education in the United States is the inadequate preparation of incoming students, in part due to the inadequate university education of K–12 teachers.
- Interdisciplinary cooperation of instructors, departments, institutions, and professional organizations is required for synthesis, integration, and change in the entire chaotic educational system.
- Various institutional and political factors, including the culture of research universities, slow educational reform.
- The monumental inertia of the educational system may thwart long-term national reform.
- “Education is not rocket science, it’s much harder” — George (Pinky) Nelson, astronaut, astrophysicist, and former director of the AAAS Project 2061, quoted in Redish (1999).

Note. The items in this table were created from Hake (2002a) and Hake (2007a).
Lessons From SoTL in Physics

For more than five decades, physics education researchers have repeatedly shown that traditional introductory physics courses with passive student lectures, recipe labs, and algorithmic problem exams are of limited value in enhancing students’ conceptual understanding of the subject (Karplus, 1964; McKinnon, 1971; McDermott & Redish, 1999). Unfortunately, this work was largely ignored by the physics and education communities until Halloun and Hestenes (1985a, 1985b) devised the Mechanics Diagnostic (MD) test of conceptual understanding of Newtonian mechanics. Among many other virtues, the MD and the subsequent Force Concept Inventory (FCI; Hestenes et al., 1992; Halloun et al., 1995) tests have two major advantages: (a) the multiple-choice format facilitates relatively easy administration of the tests to thousands of students; and (b) the questions probe for a conceptual understanding of the basic concepts of Newtonian mechanics in a way that is understandable to the novice who has never taken a physics course (and thus can be given as an introductory course pretest), yet at the same time are rigorous enough for the initiate (Hake, 2002a).

Here are 14 hard won lessons derived from lessons from the physics education reform effort, as presented in Table 1. I suspect that these lessons might be beneficial to some in the psychology community.

Based on all the above, I think it is conceivable that psychologists (as biologists, economists, and mathematicians) might learn something from the Scholarship of Teaching and Learning in physics.1

References


Bond, L. (2005, March). A different way to think about teaching and learning: Who has the lowest prices. Carnegie Perspectives, 1–3. Available online at http://1.usa.gov/1wPSg1i


Received October 29, 2014
Revision received January 14, 2015
Accepted January 25, 2015