AERA-LWhat Might Psychologists Learn from Scholarship of Teaching and Learning in Physics?

Richard R. Hake
Physics Dept., Indiana University – Bloomington

In this article I:

(1) note that psychologists have made outstanding contributions to the Scholarship of Teaching and Learning (SoTL);

(2) bemoan the failure of psychologists to develop a Psychology Concept Inventory;

(3) point to evidence from pre/post testing with the *Force Concept Inventory* for the approximately two-standard-deviation superiority in average pre-to–post-course normalized gains in conceptual understanding for "Interactive Engagement" (IE) over "Traditional" (T) passive-student lecture courses;

(4) provide crucial operational definitions of "average normalized gain," "IE methods," "T methods," and "direct instruction";

(5) list 14 hard won lessons from SoTL in physics that may be of value to psychologists; and

(6) conclude that it is conceivable that psychologists might learn something from SoTL in physics.

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

I. Psychologists Have Made Important Contributions to the Scholarship of Teaching and Learning (SoTL)

It is common knowledge that psychologists have made important contributions to the Scholarship of Teaching and Learning (SoTL); for example, see e.g., 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 paper, physicists Halloun and Hestenes (1985a) paid tribute to the foundational work of McCloskey and his collaborators (Caramazza, McCloskey, & Green, 1981; McCloskey, Caramazza, & Green, 1980).

II. 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 to 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 would appear 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 [AIP (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 employed by psychologists to evaluate their teaching. At the very least, since 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 employed 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 though 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: (1) the widely acknowledged need for reform of higher education – see for example: Arum and Roksa (2011), Delbanco (2012), Hacker and Dreifus (2010), Keeling and Hersh (2011), and Selingo (2013); (2) the evidence for the superiority of interactive engagement over traditional passive-student lecture methods in physics; and (3) accolades for SoTL in physics from (a) biologists [Klymkowsky, Garvin-Doxas, & Zeilik (2003)] and Wood & Gentile (2003)]; (b) economists [Simkins & Maier (2008)]; and (c) 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 - see “Should We Measure Change? Yes!” [Hake (2011) at http://bit.ly/d6WVKO] for the references (highlighted in yellow text) in the next two sentences - (e.g., Rogosa, 1995; Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1983, 1985; Wittmann, 1997; Zimmerman, 1997; Zimmerman & Williams, 1982; 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:

(a) 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 own courses and thus serving as role models for other faculty.

(b) 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.

(c) 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?”

III. 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 “hose 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), i.e., “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 vs 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 (%<post> – %<pre>) to the *maximum* possible average gain (100 – %<pre>). Fourteen "traditional" (T) courses (N = 2084) which made little or no use of interactive-engagement (IE) methods achieved an average gain <g> T-ave = 0.23 ± 0.04 (std dev). In sharp contrast, forty-eight courses (N = 4458) which made substantial use of IE methods achieved an average gain <g> IE-ave = 0.48 ± 0.14 (std dev), almost two standard deviations of <g> 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:

> “. . . . 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 have acquired basic content knowledge, many do not recognize that they can also be used to measure complex cognitive processes. For example, the Force Concept Inventory . . . [Hestenes et al. (1992)] . . . is an assessment that uses multiple-choice items to tap into higher-level cognitive processes (p. 94).”

IV. 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.

Table 1. Fourteen Lessons from the Physics Education Reform Effort

A. Six Lessons on Interactive Engagement

(1) The use of interactive engagement strategies can increase the effectiveness of conceptually difficult courses well beyond that obtained by traditional passive-student lecture methods.

(2 The use of interactive engagement and/or high-tech methods, by themselves, does not ensure superior student learning.

(3) High-quality standardized tests of the cognitive and affective impact of courses are essential to gauge the relative effectiveness of non-traditional educational methods.

(4) 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.

(5) 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.

(6) Although non-traditional 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.

B. Eight Lessons on Implementation

(7) Teachers who possess both content knowledge and “pedagogical content knowledge” are more apt to deliver effective instruction.

(8) College and university faculty tend to overestimate the effectiveness of their own instructional efforts and thus tend to see little need for educational reform.

(9) Such complacency can sometimes be countered by administering high-quality standardized tests of understanding and by “video snooping.”

(10) 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.

(11) Interdisciplinary cooperation of instructors, departments, institutions, and professional organizations is required for synthesis, integration, and change in the entire chaotic educational system.

(12) Various institutional and political factors, including the culture of research universities slow educational reform

(13) The monumental inertia of the educational system may thwart long-term national reform.

(14) “Education is not rocket science, it's much harder” -- George (Pinky) Nelson, astronaut, astrophysicist, and former director of the AAAS Project 2061, quoted Redish (1999).

Note. The items in the above table were created from Hake (2002a) and Hake (2007a).

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

Acknowledgments

I thank Regan Gurung and Eric Landrum for their help in shortening the original version of this paper and for converting references in the original version to APA Style. This work received partial support from National Science Foundation Grant DUE/MDR9253965. I have benefited from conversations or correspondence with Patricia Cohen, Robert DeHaan, Jerry Epstein, Robert Fuller, Paul Hickman, Brett Magill, David Meltzer, Robert Mislevy, Jeanne Narum, Ed Nuhfer, Antti Savinainen, Lee Sechrest, Werner Wittmann, LarryWoolf, Donald Zimmerman, and Bruno Zumbo. In addition, my ideas on education have benefited from exchanges on discussion lists in diverse disciplines [see e.g., the listing in "A Guide to the ADLsphere: Over Eighty Academic Discussion Lists With Links to Their Archives & Search Engines" (Hake, 2010). Most importantly I thank psychologist Regan A.R. Gurung for suggesting that I write this article and for his valuable comments.

REFERENCES [All URLs accessed on 21 March 2015; most shortened by To.]

Note #1 : A few references are to posts on the CLOSED! archives <http://bit.ly/nG318r> of the physics education discussion list PhysLrnR. To access the archives of PhysLnR one needs to subscribe, but that takes only a few minutes by clicking on http://bit.ly/nG318r and then clicking on "Join or Leave PHYSLRNR-LIST." If you're busy, then subscribe using the "NOMAIL" option under "Miscellaneous." Then, as a subscriber, you may access the archives and/or post messages at any time, while receiving NO MAIL from the list!]

Note #2: The conversion of references from APS to APA style in the above version resulted in the elimination of many URLs in the original APS version. To access the original URLs see the APS version, online as reference #75 at http://bit.ly/a6M5y0 .

AIP. 2013. American Institute of Physics. Recent enrollments in introductory physics courses. *AIP Statistical Research Center*; online at <http://bit.ly/1tTabc6>.

Alexander, P.A. & P.H. Winne, eds. 2006, 2009. *Handbook of Educational Psychology*, 2nd ed. Routledge, publisher's information at http://bit.ly/1ntfFqx. Amazon.com information at http://amzn.to/1Dkl0UZ, note the searchable "Look Inside:" feature. An expurgated Google book preview of the 2006 printing is online at http://bit.ly/1B6edyp.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Berliner, D. 2006. Educational Psychology: Searching For Essence Throughout a Century of Influence, Chapter 1, pp. 3-28 in *Handbook of Educational Research*, 2nd ed. [Alexander & Winne (2006, 2009)].

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment, *Phi Delta Kappan, 80*, 139-144, 146-148.

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

Bressoud, D. (July 2012). Learning from the physicists, *Launchings*; online at http://bit.ly/MrAuyZ .

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in 'sophisticated' subjects: Misconceptions about trajectories of objects, *Cognition, 9*, 117-123.

Cronbach, L. & Furby, L. (1970). How we should measure 'change'- or should we? *Psychological Bulletin, 74,* 68-80.

Delbanco, A. (2012). *College: What it was, is, and should be.* Princeton, NJ: Princeton University Press. Freedheim, D.K., & Weiner, I. B. (2004). *Handbook of psychology, history of psychology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Hacker, A., & Dreifus, C. (2010). *Higher education? How colleges are wasting our money and failing our kids and what we can do about it*. New York, NY: Times Books

Hake, R.R. (1998a). Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *American Journal of Physics, 66*, 64-74; online as an 84 kB pdf at http://bit.ly/9484DG . See also the crucial companion paper Hake (1998b).

Hake, R. R. (1998b). Interactive-engagement methods in introductory mechanics courses; online at http://bit.ly/aH2JQN .

Hake, R. R. 2002a. Lessons from the Physics Education Reform Effort. *Ecology and Society*. *2*, 28; online at http://bit.ly/aL87VT. For an update on the six lessons on "interactive engagement" see Hake (2007a).

Hake, R. R. (2002b). Re: Problems with student evaluations: Is assessment the remedy? Online at http://bit.ly/hjt9ll .

Hake, R. R. (2005). The physics education reform effort: A possible model for higher education; online at http://bit.ly/9aicfh .

Hake, R. R. (2007a). Six Lessons From the Physics Education Reform Effort, *Latin American Journal of Physics Education*, *1(1*), 24-31; online with AIP style numbered references as initially published at http://bit.ly//12eL2LS (scroll down to page 24). Also available with APA style references as a 684 kB pdf at http://bit.ly/96FWmE .

Hake, R. R. (2007b). Re: Pretest and post-test instruments for an introduction to psychology; online at http://bit.ly/1y3pzA4 ; for access see the note under "References" above.

Hake, R. R. (2007c). Why don't psychologists research the effectiveness of introductory courses? Online at http://bit.ly/10kL9F6 .

Hake, R. R. (2008a). Design-based research in physics education research: A review; online at http://bit.ly/9kORMZ .

Hake, R. R. (2008b). Can Distance and Classroom Learning Be Increased? *IJ-SoTL 2(1*) ; online at http://bit.ly/HCcQjm . The *International Journal of Scholarship of Teaching and Learning* (IJ-SoTL) http://bit.ly/1j7ircX is an open, double blind, peer reviewed, international electronic journal containing articles, essays, and discussions about the scholarship of teaching and learning (SoTL) and its applications in higher/tertiary education today.

Hake, R.R. (2010). Over Eighty Academic Discussion Lists With Links To Their Archives & Search Engines, online as a 3.7 MB pdf at http://bit.ly/970OZr.

Hake, R.R. (2011). Should we measure change? Yes! Online at http://bit.ly/d6WVKO .

Hake, R.R. (2014). Two Different Meanings of "Formative Evaluation #2," online on the OPEN! AERA-L archives at http://bit.ly/1FP2xlY. Post of 29 Jan 2014 19:06:18-0800 to AERA-L and Net-Gold. The abstract and link to the complete post were transmitted to several discussion lists and are on my blog "Hake'sEdStuff" at http://bit.ly/MxnMBt with a provision for comments.

Halloun, I., & Hestenes, D. (1985a). The initial knowledge state of college physics, *American Journal of Physics, 53,* 1043-1055.

Halloun, I., & Hestenes, D. (1985b). Common sense concepts about motion, *American Journal of Physics, 53,* 1056-1065.

Halloun, I., Hake, R. R., Mosca, E. P., & Hestenes, D. (1995). Force concept inventory; online at http://bit.ly/b1488v .

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *Physics Teacher*, *30*, 141-158.

Holton, D. (2007). Re: Pretest and post-test instruments for an introduction to psychology; online at http://bit.ly/1rQCYWl; for access see the note under "References" above.

Karplus, R. (1964). One physicist experiments with science education, *American Journal of Physics, 32,* 837-839.

Keeling, R.P., & Hersh, R. H. (2011). *We're losing our minds: Rethinking American higher education.* London, England: Macmillan Publishers.

Kelly, A.E. (2003). Research as design. *Educational Researcher, 32,* 3-4.

Kelly, A.E., Lesh, R. A., & Baek, J. Y. (2008). *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching*. New York, NY: Routledge.

Klymkowsky, M.W., Garvin-Doxas, K., & Zeilik, M. (2003). Bioliteracy and teaching efficiency: What biologists can learn from physicists, *Cell Biology Education, 2,* 155-161.

Lagemann, E.C. (2000). *An elusive science: The troubling history of education research*. Chicago, IL: University of Chicago Press.

Lamal, P.A. (1979). College students' common beliefs about psychology. *Teaching of Psychology, 6,* 155-158.

Lattery, M. J. (2008). Signature pedagogies in introductory physics. In R. A. R. Gurung, N. L. Chick, & A. Haynie (Eds.), *Exploring signature pedagogies: Approaches to teaching disciplinary habits of mind* (pp. 280-294). Sterling, VA: Stylus.

Lederman, N.G., Bartos, S. A., & Lederman, J. S. (2014). The development, use, and interpretation of nature of science assessments, In M. R. Matthews (Ed.), *Handbook of research in history, philosophy and science teaching* (pp. 971-997). New York, NY: Springer.

Lilienfeld, S.O. (2010). Confronting psychological misconceptions in the classroom: Challenges and rewards, *Observer, 23.*

Lord, F.M. (1956). The measure of growth, *Educational and Psychological Measurement, 16,* 421-437.

Lord, F.M. (1958). Further problems in the measurement of growth, *Educational and Psychological Measurement, 18,* 437-454.

Maki, P.L. (2010). *Assessing for learning: Building a sustainable commitment across the institution* (2nd ed.). Sterling, VA: Stylus. Note especially "Making the Invisible Visible in Physics," pp. 123-129.

McCloskey, M., Caramazza A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects, *Science, 210,* 1139-1141.

McCloskey, M. (1983). Intuitive physics. *Scientific American, 248*, 122-130.

McCutcheon, L.E. (1991). A new test of misconceptions about psychology, *Psychological Reports, 68,* 647-653.

McDermott L. C., & Redish, E. F. (1999). RL-PER1: Resource letter on physics education research, *American Journal of Physics*, *67*, 755-767.

McKinnon, J.W. (1971). Earth science, density, and the college freshman. *Journal of Geology Education, 19*, 218-220.

Redish, E.F. (1999). Millikan lecture 1998: Building a science of teaching physics, *American Journal of Physics, 67,* 562-573.

Selingo, J. (2013). *College unbound: The future of higher education and what it means for students*. New York, NY: New Harvest.

Shavelson, R.J., & Huang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change, 35*, 10-19. Shavelson considers "formative assessment" in the sense of feedback to students.

Simkins, S. P., & Maier, M. H. (2008). *Learning from physics education research: Lessons for economics education*: online at http://bit.ly/1rr962K .

Steuer, F. B., & Ham, K. W. (2008). Psychology textbooks: Examining their accuracy. *Teaching of Psychology, 35*, 160-168. doi:10.1080/00986280802189197

Suskie, L. 2009. *Assessing Student Learning: A Common Sense Guide*, foreword by Trudy Banta. Wiley, 2nd edition, publisher's information at http://amzn.to/fMXH97, note the searchable "Look Inside" feature. Contrast this book by the anti-pre/post Linda Suskie with that of the pro-pre/post Peggy Maki (2010).

Thornton, R.K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, *American Journal of Physics, 66,* 338-352.

Vaughan, E.D. (1977). Misconceptions about psychology among introductory psychology students, *Teaching of Psychology, 4,* 138-141.

Wilson, M.R., & Bertenthal, M. W. (2005). *Systems for state science assessment*. Washington, DC: National Academy Press.

Wittmann, M. (2007). Re: Pretest and Post-test instruments for an Introduction to Psychology; online at http://bit.ly/1sV4APA : for access see the note under "References" above.

Wood, W. B., & Gentile, J. M. (2003). Teaching in a research context, *Science, 302,* 1510; online as a 213 kB pdf at http://bit.ly/SyhOvL .