Measuring Teaching and Learning Performance: Interconnected Issues*◊

Richard Hake <rrhake@earthlink.net>, < http://www.physics.indiana.edu/~hake >
Indiana University, Emeritus

From Teaching to Learning

In their landmark wake-up call to American higher education "From Teaching to Learning: A New Paradigm for Undergraduate Education," Barr & Tagg (1995) wrote: "A paradigm shift is occurring in American higher education. Under the traditional, dominant 'Instruction Paradigm' colleges are institutions that exist *to provide instruction*. Subtly but profoundly, however, a 'Learning Paradigm' is taking hold, whereby colleges are institutions that exist *to produce learning*. This shift is both needed and wanted, and it changes everything."

Consistent with the broad theme of this *international* conference "Measurement and Evaluation in Education," I think it might be appropriate to generalize Barr and Tagg's rallying cry as follows:

> "A paradigm shift is desirable in world-wide education. Under the traditional, dominant 'Instruction Paradigm' schools and colleges are institutions that exist *to provide instruction*. It is to be hoped that a 'Learning Paradigm' might eventually take hold, whereby educational institutions in *all* countries would exist *to produce learning*. This shift is both needed and wanted, and it would change everything."

What to Measure and How to Measure

Investigation of the extent to which a paradigm shift from teaching to learning is taking place requires measurement of students' learning in school and college classrooms. But Wilbert McKeachie (1987) has pointed out that the time-honored gauge of student learning – course exams and final grades – typically measures lower-level educational objectives such as memory of facts and definitions rather than higher-level outcomes such as critical thinking and problem solving. Furthermore, the same criticism (Hake, 2002a) as to assessing only lower-level learning applies to Student Evaluations of Teaching (SET's), since, at least in American higher education, their primary justification as measures of student learning appears to lie in the modest correlation with overall ratings of course (+ 0.47) and instructor (+ 0.43) with "achievement" *as measured by course exams or final grades* (Cohen, 1981).

For general characterizations of higher-order learning see Anderson & Krathwohl (2001) and Shavelson & Huang (2003). The latter, in their Chart 1 - "Framework for Cognitive Outcomes," display higher-level learning *within knowledge domains*, as might be measured and enhanced by disciplinary experts: "procedural - knowing how" – see, e.g., Anderson (2004); "schematic – knowing why" ; and "strategic – knowing when certain knowledge applies, where it applies, and how it applies." These contrast with the lower-order "declarative - knowing that."

---

* The reference is: Hake, R.R. 2006. "Measuring Teaching and Learning Performance: Interconnected Issues," in *Proceedings of the Third International Conference on Measurement and Evaluation in Education (ICMEE 2006)*, Penang, Malaysia, 13-15 February. This paper is also online with hot-linked URL's as ref. 38 at < http://www.physics.indiana.edu/~hake > or download directly by clicking on < http://www.physics.indiana.edu/~hake/ICMEEk-2006.pdf > (230 kB). I welcome comments and suggestions directed to <rrhake@earthlink.net>.

◊ Partially supported by NSF Grant DUE/MDR-9253965.

How then can we measure students' higher-level learning? In this paper I shall concentrate on *the measurement of learning in American higher education* because: (a) my experience has been primarily in that area, (b) judging from the paper by Othman et al. (2006) "Accountability in Malaysian Higher Education" reported at this conference, the assessment problems in higher education in Malaysia are similar to those in the U.S. , and (c) the measurement of learning in precollege courses presents problems similar to those in higher education. Much of the material to follow is drawn from "The Physics Education Reform Effort: A Possible Model for Higher Education" [Hake (2005a)].

For American higher education, several *indirect* (and therefore in my view problematic) gauges have been developed; e.g. Reformed Teaching Observation Protocol (RTOP), National Survey Of Student Engagement (NSSE), Student Assessment of Learning Gains (SALG), and Knowledge Surveys (KS's) (Nuhfer & Knipp, 2003). [For a discussion and references for all but the last see Hake (2005b)].

On the other hand, *direct* measures of student learning have been developed by Hersh (2005) and Klein et al. (2005). Hersh codirects the *Learning Assessment Project* < http://www.cae.org/content/pro_collegiate.htm > that "evaluates students' ability to articulate complex ideas, examine claims and evidence, support ideas with relevant reasons and examples, sustain a coherent discussion, and use standard written English." But Shavelson & Huang (2003) warn that "learning and knowledge are highly domain-specific—as, indeed, is most reasoning. Consequently, *the direct impact of college is most likely to be seen at the lower levels of Chart 1 – domain-specific knowledge and reasoning* [my *italics*]. Klein et al. have devised tests that compare student learning across institutions in both domain-specific and broad-ability areas of the Shavelson & Huang 2003 "Framework of Cognitive Objectives" (SHFCO).

Pre/Post Testing

In sharp contrast to the above mentioned invalid (course exams, final grades, SET's); indirect (RTOP, NSSE, SALG, KS's); or general-ability [Hersh (2005), Klein et al. (2005)] measures discussed above, is the direct measure of students' higher-level *domain-specific* learning through pre/post testing using (a) valid and consistently reliable tests *devised by disciplinary experts*, and (b) traditional courses as controls. It should be realized that domain specific learning is probably coupled to the broad-ability areas of the SHFCO, as suggested for physics by the recent research of Coletta & Phillips (2005).

In my opinion, the physics-education reform model – measurement and improvement of cognitive gains by faculty disciplinary experts *in their own courses* – can provide a crucial complement to the top-down approaches of Hersh (2005) and Klein et al. (2005). Such pre/post testing, pioneered by economists [Paden & Moyer (1969)] and physicists [Halloun & Hestenes (1985a,b)], is rarely employed in higher education, in part because of the tired old canonical objections recently lodged by Suskie (2004) and countered by Hake (2004a) and Scriven (2004). Despite the nay-sayers, pre/post testing is gradually gaining a foothold in introductory astronomy, economics, biology, chemistry, computer science, economics, engineering, and physics courses [see Hake (2004b) for references].

It should be emphasized that such low-stakes *formative* pre/post testing is the polar opposite of the high-stakes *summative* testing mandated by the U.S. Department of Education's *No Child Left Behind Act* (NCLB) for pre-college grades (USDE 2005a). The pre-to-post-test gain data for a course is *only* for the use the teacher of that course, so as to enable the teacher to gauge the degree of student learning in the course and thereby the effectiveness of her or his teaching. The teacher's salary and prestige are not affected the gain data averages. Likewise a student's course grade or prestige are not affected by her or his pre-to-post-test gain. Such formative assessment acts to improve student learning while avoiding the consequences of "Campbell's Law" [Campbell (1975), Nichols & Berliner (2005)]:

> "The more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."

Judging from the paper by Hussan et al. (2006) "Integrity in Public Examinations: A Malaysian Experience" reported at this conference, the Malaysian Ministry of Education is engaged in a continual heroic battle to avoid the consequences of Campbell's Law and thus preserve the integrity of its summative pre-college testing that occurs at the ends of primary education, lower secondary education, upper secondary education, and the pre-tertiary program.

What Physics Has Learned

Physics education researchers (PER's) have employed formative pre/post testing to show that traditional (T) introductory physics courses promote very little change in students' understanding of basic physics concepts; regardless of the experience, enthusiasm, talents, and motivation of their professors. This has driven some physicists to develop novel "interactive engagement" (IE) methods, among them: Microcomputer-based Labs, Concept Tests, Modeling, Active Learning Problem Sets, Overview Case Studies, and Socratic Dialogue Inducing Labs (for references see Hake 2002b). That such IE methods are relatively effective in promoting student higher-level learning has been demonstrated by the nearly two-standard deviation [cf. Bloom's (1984) "two sigma problem") superiority in normalized average learning gains <g> of IE courses over T (traditional) courses [Hake (1998a,b; 2002b,c) and corroborative references therein]. Notable examples are large enrollment courses at Harvard [Crouch & Mazur (2001)], North Carolina State University [Beichner & Saul (2004)], MIT [Dori & Belcher (2004)], the University of Colorado at Boulder [Pollock (2004)], and California Polytechnic State University at San Luis Obispo [Hoellwarth et al. (2005)].

Some definitions are in order. In the above paragraph (a) the average *normalized* gain <g> is the *actual* gain [<%post> - <%pre>] divided by the *maximum possible gain* [100% - <%pre>], where the angle brackets indicate the class averages; (b) "traditional" (T) courses are operationally defined courses as those reported by instructors to make little or no use of "interactive engagement" (IE) methods, relying primarily on passive-student lectures, recipe labs, and algorithmic problem exams; (c) IE courses are operationally defined as those designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield *immediate* feedback through discussion with peers and/or instructors.

With regard to the Malaysian Ministry of Education's exemplary program to enhance school-based formative assessment [Hassan (2006)], it should be emphasized that IE-methods with their immediate feedback to students also entail immediate feedback to teachers on the knowledge state of the students, thereby facilitating continual formative assessment. Two examples are the continual Socratic questioning employed in "Socratic Dialogue Inducing Labs" [Hake (1992, 2002d)], and Mazur's (1997) continual Concept Testing by means of an electronic classroom communication system [see e.g. Burnstein & Lederman (2003)].

For links to over 50 U.S. PER groups, over 200 PER papers published in the American Journal of Physics since 1972, and tests of cognitive and affective conditions see, respectively, Meltzer (2005a), Meltzer (2005b), and NCSU (2005). The very active PER discussion list PhysLrnR < http://listserv.boisestate.edu/archives/physlrnr.html > logged over 750 posts in 2005. As far as I know, no other discipline is so actively researching undergraduate student learning. For reviews see McDermott & Redish (1999), Redish (1999), Thacker (2003), Heron & Meltzer (2005), and Wieman & Perkins (2005).

Synapse Stimulation

The fact that IE methods are far more effective in promoting conceptual understanding than traditional passive-student methods is probably related to the "enhanced synapse addition and modification" induced by those methods. Bransford et al. (2000) wrote: ". . . synapse addition and modification are lifelong processes, driven by experience. In essence, the quality of information to which one is exposed and the amount of information one acquires is reflected throughout life in the structure of the brain. This process is probably not the only way that information is stored in the brain, but it is a very important way that provides insight into how people learn." Leamnson (1999, 2000) has also stressed the relationship of biological brain change to student learning. In his Chapter 5 "Teaching and Pedagogy," Leamnson (1999) wrote, "Teaching must involve telling, but learning will only start when something persuades students to engage their minds and do what it takes to learn." Another reminder that the affective and the cognitive are inextricably linked, as recently emphasized by Nuhfer (2005).

The Challenge

I see no reason that student learning gains far larger than those in traditional courses could not eventually be achieved and documented in other disciplines from arts through philosophy to zoology *IF* their practitioners would: (a) reach a consensus on the *crucial* concepts that all beginning students should be brought to understand, (b) undertake the lengthy qualitative and quantitative research [see, e.g., Halloun & Hestenes (1985a,b)], required to develop multiple-choice tests (MCT's) of higher-level learning of those concepts, and (c) develop Interactive Engagement methods suitable to their disciplines. Why MCT's? So that the tests can be given to thousands of students in hundreds of courses under varying conditions in such a manner that meta-analyses can be performed, thus establishing general causal relationships in a convincing manner.

But can multiple-choice tests measure *higher-order* learning? Wilson & Bertenthal (2005) think so, writing (p. 94): "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 – available in Malaysian!] . . . is an assessment that uses multiple-choice items to tap into higher-level cognitive processes."

Lessons Learned
Can nearly all university disciplines develop synapse-stimulating interactive engagement methods, and also valid and reliable multiple-choice tests of affective and cognitive conditions to measure their effectiveness? I would bet "Yes," provided they care enough about student learning to mount the necessary research and development effort.

Aside from the advantages of pre/post testing, perhaps physics education researchers' most important lessons (Hake 2002b) for higher education are Lessons #1, 3, and 4:

> L1: The use of Interactive Engagement strategies can increase the effectiveness of conceptually difficult courses well beyond that obtained with traditional methods.
>
> L3: High-quality standardized tests of the cognitive and affective impact of courses are essential for gauging the relative effectiveness of non-traditional and traditional educational methods. For examples of such physics tests see the listing at NCSU (2005).
>
> L4: Education Research and Development by disciplinary experts (DEs), 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 DEs should take advantage of the insights of DEs engaged in education R&D in other disciplines, cognitive scientists, faculty and graduates of education schools, and classroom teachers.

In the U.S., calls for the accountability of higher education in promoting student learning are becoming more forceful, both from inside the university, e.g., Duderstadt (2000), Weber & Duderstadt (2004), Hersh (2005), Hersh & Merrow (2005), Bok (2005a,b,c); and outside the university, e.g., by the U.S. Dept. of Education's new "Commission on the Future of Higher Education" (USDE 2005b). For reports on the Commission's first two meetings and commissioner's comments on the possibility of NCLB-like testing in higher education, and on the declining literacy of college graduates (NAAL 2005), see Lederman (2005a, b).

As Hersh (2005) observes: ". . . in an era when the importance of a college diploma is increasing while public support for universities is diminishing, [assessment of student learning] is desperately needed. The real question is who will control it. Legislators are prepared to force the issue: Congress raised the question of quality during its recent hearings on the reauthorization of the Higher Education Act; all regional accrediting agencies and more than forty states now require evidence of student learning from their colleges and universities; and pressure is rising to extend a *No Child Left Behind*-style testing regime to higher education" [see USDE (2005a,b)].

Thus it would appear to be high time for American faculty members to turn more of their attention to shifting the higher education paradigm from teaching to learning, both because *it's the right thing to do*, and because not doing so may invite stifling oversight by state and national bureaucrats. Judging from the paper by Othman et al. (2006) "Accountability in Malaysian Higher Education" reported at this conference, a similar suggestion might be appropriate for Malyasian university faculty.

References and Footnotes [Tiny URL's courtesy < http://tinyurl.com/create.php >.]

Anderson, L.W. & L.A. Sosniak, eds. 1994. *Bloom's Taxonomy: A Forty-Year Retrospective,* Ninety-Third Yearbook of The National Society for the Study of Education, Univ. of Chicago Press. Amazon.com information at < http://tinyurl.com/7bcnm >.

Anderson, L.W. & D. Krathwohl, eds. 2001. *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Addison Wesley Longman. See also Anderson & Sosniak (1994). The original 1956 Bloom et al. cognitive domain taxonomy has been updated to include important post-1956 advances in cognitive science – see especially Chapters 4 & 5 on the Knowledge and Cognitive Process Dimensions. See also the companion *affective* taxonomy by Krathwohl et al. (1990). Amazon.com information at < http://tinyurl.com/dh229 >.

Anderson, J.R. 2004. *Cognitive Psychology and Its Implications*. Worth, 6th ed. Amazon.com information at < http://tinyurl.com/cerow >.

Barr, R.B. & J. Tagg. 1995. "From Teaching to Learning: A New Paradigm for Undergraduate Education," *Change* **27**(6); 13-25, November/December. Reprinted in D. Dezure, *Learning from Change: Landmarks in Teaching and Learning in Higher Education from Change 1969-1999*. American Association for Higher Education, pp. 198-200. Also online at < http://tinyurl.com/8g6r4 >.

Beichner, R.J & J.M. Saul. 2004. "Introduction to the SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) Project," in *Proceedings of the International School of Physics "Enrico Fermi" Course CLVI* in Varenna, Italy, M. Vicentini and E.F. Redish, eds. IOS Press. Amazon.com information at < http://tinyurl.com/cwgrk >; online at at < http://www.ncsu.edu/per/Articles/Varenna_SCALEUP_Paper.pdf > (1MB).

Bloom, B.S. 1984. "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring," *Educational Researcher* **13**(6), 4-16 (1984). Bloom wrote: "Using the standard deviation (sigma) of the control (conventional) class, it was typically found that the average student under tutoring was about two standard deviations above the average of the control class. . . The tutoring process demonstrates that *most* of the students do have the potential to reach this high level of learning. I believe an important task of research and instruction is to seek ways of accomplishing this under more practical and realistic conditions than the one-to-one tutoring, which is too costly for most societies to bear on a large scale. This is the '2 sigma' problem."

Bok, D. 2005a. *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More*. Princeton University Press. Amazon.com information is at < http://tinyurl.com/bnn8c >.

Bok, D. 2005b. "The Critical Role of Trustees in Enhancing Student Learning," *Chronicle of Higher Education*, 13 December, online to subscribers at < http://chronicle.com/chronicle/v52/5217guide.htm > .

Bok, D. 2005c. "Are colleges failing? Higher ed needs new lesson plans" *Boston Globe*, 18 December, freely online (probably only for a short time) at < http://tinyurl.com/da5v2 >, and to educators at < http://tinyurl.com/aj95w > (scroll to the APPENDIX). Bok wrote: ". . . studies indicate that problem-based discussion, group study, and other forms of active learning produce greater gains in critical thinking than lectures, yet the lecture format is still the standard in most college classes, especially in large universities."

Bransford, J.D., A.L. Brown, R.R. Cocking, eds. 2000. *How people learn: brain, mind, experience, and school*. Nat. Acad. Press; online at < http://tinyurl.com/apbgf >.

Burnstein, R.A. & L.M. Lederman. 2003. "Comparison of Different Commercial Wireless Keypad Systems," *Phys. Teach.* **41**(5): 272-275; online at < http://ojps.aip.org/dbt/dbt.jsp?KEY=PHTEAH&Volume=41&Issue=5 >.

Campbell, D. T. 1975. "Assessing the impact of planned social change," in G. Lyons, ed. , *Social research and public policies: The Dartmouth/OECD Conference*, Chapter 1, pp. 3-45. Dartmouth College Public Affairs Center, p. 35; online at < http://www.wmich.edu/evalctr/pubs/ops/ops08.pdf > (196 kB).

Cohen, P.A. 1981. "Student ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies," *Review of Educational Research* **51**: 281. For references to Cohen's 1986 and 1987 updates see Feldman (1989).

Coletta, V.P. and J.A. Phillips. 2005. "Interpreting FCI Scores: Normalized Gain, Preinstruction Scores, & Scientific Reasoning Ability," *Am. J. Phys*. **73**(12): 1172-1182; online at < http://tinyurl.com/8eozb >.

Crouch, C.H. & E. Mazur. 2001. "Peer Instruction: Ten years of experience and results," *Am. J. Phys*. **69**: 970-977; online at < http://tinyurl.com/d35z4 >.

Dori, Y.J. & J. Belcher, J. 2004. "How Does Technology-Enabled Active Learning Affect Undergraduate Students' Understanding of Electromagnetism Concepts?" *The Journal of the Learning Sciences* **14**(2), online as a 1 MB pdf at < http://tinyurl.com/cqoqt >.

Duderstadt, J.J. 2000. *A University for the 21st Century.* Univ. of Michigan Press; for a description see < http://tinyurl.com/9lhpl >. Duderstadt writes: "Few faculty members have any awareness of the expanding knowledge about learning from psychology and cognitive science. Almost no one in the academy has mastered or used this knowledge base. One of my colleagues observed that if doctors used science the way college teachers do, they would still be trying to heal with leeches."

Feldman, K.A. 1989. “The Association Between Student Ratings of Specific Instructional Dimensions and Student Achievement: Refining and Extending the Synthesis of Data from Multisection Validity Studies,” *Research on Higher Education* **30**: 583.

Hake, R.R. 1992. "Socratic pedagogy in the introductory physics lab." *Phys. Teach.* **30**: 546-552; an updated version (4/27/98) is online at < http://www.physics.indiana.edu/~sdi/SocPed1.pdf > (88 kB).

Hake, R.R. 1998a. “Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,” *Am. J. Phys*. **66**: 64-74; online at < http://www.physics.indiana.edu/~sdi/ajpv3i.pdf > (84 kB).

Hake, R.R. 1998b. “Interactive-engagement methods in introductory mechanics courses,” online at < http://www.physics.indiana.edu/~sdi/IEM-2b.pdf > (108 kB) - a crucial companion paper to Hake (1998a).

Hake, R.R. 2002a. “Re: Problems with Student Evaluations: Is Assessment the Remedy?” online at < http://www.physics.indiana.edu/~hake/AssessTheRem1.pdf > (72 kB).

Hake, R.R. 2002b. “Lessons from the physics education reform effort,” *Ecology and Society* **2**: 28; online at < http://www.ecologyandsociety.org/vol5/iss2/art28/ >.

Hake, R.R. 2002c. “Assessment of Physics Teaching Methods,” *Proceedings of the UNESCO-ASPEN Workshop on Active Learning in Physics*, Univ. of Peradeniya, Sri Lanka, 2-4 Dec. 2002; online at < http://www.physics.indiana.edu/~hake/Hake-SriLanka-Assessb.pdf > (84 kB).

Hake,R.R. 2002d. "Socratic Dialogue Inducing Laboratory Workshop," *Proceedings of the UNESCO-ASPEN Workshop on Active Learning in Physics*, Univ. of Peradeniya, Sri Lanka, 2-4 Dec. 2002; online at < http://www.physics.indiana.edu/~hake/Hake-SriLanka-SDIb.pdf > (44 kB).

Hake, R.R. 2004a. “Re: pre-post testing in assessment,” online at < http://listserv.nd.edu/cgi-bin/wa?A2=ind0408&L=pod&P=R9135&I=-3 >. Post of 19 Aug 2004 13:56:07-0700 to POD.

Hake, R.R. 2004b. “Re: Measuring Content Knowledge,” POD posts of 14 &15 Mar 2004, online at < http://listserv.nd.edu/cgi-bin/wa?A2=ind0403&L=pod&P=R13279&I=-3 > and < http://listserv.nd.edu/cgi-bin/wa?A2=ind0403&L=pod&P=R13963&I=-3 >.

Hake, R. R. 2005a. “The Physics Education Reform Effort: A Possible Model for Higher Education,” online at < http://www.physics.indiana.edu/~hake/NTLF42.pdf > (100 kB), a slightly edited version of an article that was (a) published in the *National Teaching and Learning Forum* **15**(1), December 2005, online to subscribers at < http://www.ntlf.com >, and (b) disseminated by the Tomorrow’s Professor list < http://ctl.stanford.edu/Tomprof/postings.html > as Msg. 698 on 14 Feb 2006.

Hake, R.R. 2005b. “Re: Measuring Teaching Performance,” POD post of 13 May 2005; online at < http://listserv.nd.edu/cgi-bin/wa?A2=ind0505&L=pod&P=R9303&I=-3 >.

Halloun, I. & D. Hestenes. 1985a. “The initial knowledge state of college physics students,” *Am. J. Phys*. **53**: 1043-1055; online at < http://modeling.asu.edu/R&E/Research.html >. Contains the “Mechanics Diagnostic” test (omitted from the online version), precursor to the widely used “Force Concept Inventory” [Hestenes et al. (1992)].

Halloun, I. & D. Hestenes. 1985b. “Common sense concepts about motion,” *Am. J. Phys*. **53**: 1056-1065; online at < http://modeling.asu.edu/R&E/Research.html >.

Hassan, S. 2006. “Integrity in Public Examinations: A Malaysian Experience,” in *Proceedings of the Third International Conference on Measurement and Evaluation in Education (ICMEE 2006)*, Penang, Malaysia, 13-15 February.

Hersh, R.H. 2005. “What Does College Teach? It's time to put an end to 'faith-based' acceptance of higher education's quality,” *Atlantic Monthly* **296**(4): 140-143, November; freely online to (a) subscribers of the *Atlantic Monthly* at < http://tinyurl.com/dwss8 >, and (b) (with hot-linked academic references) to educators at < http://tinyurl.com/9nqon > (scroll to the APPENDIX).

Hersh R.H. & J. Merrow, eds. 2005. *Declining by Degrees: Higher Education at Risk*. Palgrave Macmillan. Amazon.com information at < http://tinyurl.com/bvcf4 >.

Hestenes, D., M. Wells, & G. Swackhamer, 1992. “Force Concept Inventory,” *Phys. Teach.* **30**: 141-158; online (except for the test itself) at < http://modeling.asu.edu/R&E/Research.html >. The 1995 revision by Halloun, Hake, Mosca, & Hestenes is online (password protected) at the same URL, and is available in English, Spanish, German, ***Malaysian***, Chinese, Finnish, French, Turkish, Swedish, and Russian.

Heron, P.R.L. & D. Meltzer. 2005. “The future of physics education research: Intellectual challenges and practical concerns,” *Am. J. Phys.* **73**(5): 459-462; online at < http://www.physicseducation.net/docs/Heron-Meltzer.pdf > (56 kB).

Hoellwarth, C., M. J. Moelter, and R.D. Knight. 2005. “ A direct comparison of conceptual learning and problem solving ability in traditional and studio style classrooms,” *Am. J. Phys.* **73**(5): 459-463; online at < http://tinyurl.com/br88n >.

Klein, S.P., G.D. Kuh, M.Chun, L. Hamilton, & R. Shavelson. 2005. "An Approach to Measuring Cognitive Outcomes Across Higher Education Institutions." *Research in Higher Education* **46**(3): 251-276; online at < http://www.stanford.edu/dept/SUSE/SEAL/ > // “Reports/Papers” scroll to “Higher Education,” where “//” means “click on.”

Krathwohl, D.R., B.B. Masia, with B.S. Bloom. 1990. *Taxonomy of Objectives Book 2; Affective Domain*. Longman. Amazon.com information at < http://tinyurl.com/bh6tc >.

Leamnson, R. 1999. “*Thinking About Teaching and Learning: Developing Habits of Learning with First Year College and University Students*.” Stylus. Amazon.com information at < http://tinyurl.com/d38ar >.

Leamnson, R. 2000. “Learning as Biological Brain Change,” *Change*, November/December; online at < http://tinyurl.com/9gc6a >.

Lederman, D. 2005a. “Tough Love for Colleges,” *Inside Higher Ed*, 9 December, online at < http://insidehighered.com/news/2005/12/09/commission >.

Lederman, D. 2005b. “Graduated but Not Literate,” *Inside Higher Ed*, 16 December, online at < http://insidehighered.com/news/2005/12/16/literacy >. Lederman writes (my *italics*): “Not only does [the report on literacy NAAL (2005)] find that the average literacy of college educated Americans declined significantly from 1992 to 2003, but it also reveals that just 25 percent of college graduates — and only 31 percent of those with at least some graduate studies — scored high enough on the tests to be deemed ‘proficient’ from a literacy standpoint, which the government defines as ‘using printed and written information to function in society, to achieve one’s goals, and to develop one’s knowledge and potential.’ *‘This seems like another piece of hard evidence, a fairly clear indication, that the ‘value added’ that higher education gave to students didn’t improve, and maybe declined, over this period,’ said Charles Miller, the former University of Texas regent who is heading the U.S. education secretary’s Commission on the Future of Higher Education.* ‘You have the possibility of people going through schools, getting a piece of paper for sitting in class a certain amount, and we don’t know whether they’re getting what they need. This is a fair sign that there are some problems here.’ ”

Mazur, E. 1997. "Peer instruction: a user's manual." Prentice Hall; online at < http://galileo.harvard.edu/ >, click on "Peer Instruction" in the second paragraph.

McDermott, L.C. & E.F. Redish. 1999. “RL-PER1: Resource letter on physics education research,” *Am. J. Phys.* **67**(9): 755-767; online at < http://www.physics.umd.edu/rgroups/ripe/perg/cpt.html >.

McKeachie, W.J. 1987. “Instructional evaluation: Current issues and possible improvements,” *Journal of Higher Education* **58**(3): 344-350.

Meltzer, D. 2005a. “Links to United States Physics Education Research Groups,” online at < http://www.physicseducation.net/links/index.html >.

Meltzer, D. 2005b. “Listing of Physics Education Research papers published in the American Journal of Physics since 1972,” online at < http://www.physicseducation.net/current/index.html >. Scroll to “Ongoing Projects.”

NAAL. 2005. *National Assessment of Adult Literacy: A First Look at the Literacy of America’s Adults in the 21st Century*,” National Center for Educational Statistics, online at < http://nces.ed.gov/NAAL/PDF/2006470_1.PDF > (724 kB).

NCSU. 2005. “Assessment Instrument Information Page,” Physics Education R & D Group, North Carolina State University; online at < http://www.ncsu.edu/per/TestInfo.html >.

Nichols, S.L & D.C. Berliner. 2005. “The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing,” Arizona State Univ. Education Policy Studies Laboratory, online at < http://tinyurl.com/7butg > (1.7 MB).

Nuhfer, E. 2005. “DeBono’s Red Hat on Krathwohl’s Head: Irrational Means to Rational Ends-More Fractal Thoughts on the Forbidden Affective: Educating in Fractal Patterns XIII.” *National Teaching and Learning Forum* **14** (5), online to subscribers at < http://www.ntlf.com/FTPSite/issues/v14n5/diary.htm >.

Nuhfer, E. & D. Knipp. 2003. “The Knowledge Survey: A Tool for All Reasons,” in *To Improve the Academy* **21**: 59-78; online at < http://www.isu.edu/ctl/facultydev/KnowS_files/KnowS.htm >.

Othman, A.R., R.J. Shavelson, & M.A. Ruiz-Primo. 2006. “Accountability in Malaysian Higher Education,” in *Proceedings of the Third International Conference on Measurement and Evaluation in Education (ICMEE 2006)*, Penang, Malaysia, 13-15 February.

Paden, D.W. & M.E. Moyer. 1969. “The Relative Effectiveness of Teaching Principles of Economics,” *Journal of Economic Education* **1**: 33-45.

Pollock, S. 2004. “No Single Cause: Learning Gains, Student Attitudes, and the Impacts of Multiple Effective Reforms,” *2004 Physics Education Research Conference: AIP Conference Proceeding, vol. 790*; J. Marx, P. Heron, & S. Franklin, eds., pp. 137-140, online as a 316 kB pdf at < http://tinyurl.com/9tfk4 >.

Redish, E.F. 1999. “Millikan lecture 1998: building a science of teaching physics,” *Am. J. Phys*. **67**(7): 562-573, online at < http://www.physics.umd.edu/rgroups/ripe/perg/cpt.html >.

Scriven, M. 2004. “Re: pre- post testing in assessment,” AERA-D post of 15 Sept 2004 19:27:14-0400; online at < http://tinyurl.com/942u8 >.

Shavelson, R.J. & L. Huang. 2003. “Responding Responsibly To the Frenzy to Assess Learning in Higher Education,” *Change Magazine*, January/February; online at < http://www.stanford.edu/dept/SUSE/SEAL/ > // "Reports/Papers" scroll to "Higher Education," where “//” means “click on.”

Suskie, L. 2004. “Re: pre- post testing in assessment,” ASSESS post 19 Aug 2004 08:19:53-0400; online at < http://tinyurl.com/akz23 >.

Thacker, B.A. 2003. “Recent advances in classroom physics,” *Rep. Prog. Phys*. **66:** 1833–1864, online at < http://www.iop.org/EJ/abstract/0034-4885/66/10/R07 >.

USDE. 2005. U.S. Dept. of Education, “Secretary Spellings Announces New Commission on the Future of Higher Education,” press release online at < http://tinyurl.com/cxgfz >: “Spellings noted that the achievement gap is closing and test scores are rising among our nation's younger students, due largely to the high standards and accountability measures called for by the *No Child Left Behind Act*. More and more students are going to graduate ready for the challenges of college, she said, and we must make sure our higher education system is accessible and affordable for all these students.”

USDE. 2006. U.S. Department of Education, *No Child Left Behind Act*, online at < http://www.ed.gov/nclb/landing.jhtml?src=pb >.

Weber, L.E. & J.J. Duderstadt, eds. 2004. *Reinventing the Research University*, published by Economica. Amazon.com information is at < http://tinyurl.com/89kew >. In his perceptive review, Wolff (2005) points out that Roger Downer is the only contributor to explore teaching and pedagogy. Downer states on p. 64 that "the ultimate goal of education is not excellent teaching, but, rather, excellence in student learning."

Wieman, C. & K. Perkins. 2005. "Transforming Physics Education," *Phys. Today* **58**(11): 36-41; online at < http://www.colorado.edu/physics/EducationIssues/ > / "Papers" (where "/" means "click on").

Wilson, M.R. & M.W. Bertenthal, eds. 2005. *Systems for State Science Assessment*, Nat. Acad. Press; online at < http://www.nap.edu/catalog.php?record_id=11312 >. For an executive summary see < http://www.nap.edu/execsumm_pdf/11312.pdf >

Wolff. B. 2005. Review of Weber & Duderstadt (2004), *Planning for Higher Education* **33**(4): 37-39; online at < http://tinyurl.com/82ppn >. Wolff wrote: "In each of the essays there is a written and implied understanding that a university of the future can only be successful 'by adapting to market forces' (p. 208)." But in a nation beset with life-threatening problems (political, social, environmental, and economic) should not the university adapt to Barr and Tagg's 1995 student-learning paradigm rather than to "market forces" ?