

Should We Measure Change? Yes! * † ◇

Richard Hake

Indiana University, Emeritus

Formative pre/post testing is being successfully employed to improve the effectiveness of courses in undergraduate astronomy, economics, biology, chemistry, economics, geoscience, engineering, and physics. But such testing is still anathema to many members of the psychology-education-psychometric (PEP) community. I argue that this irrational bias impedes a much needed enhancement of student learning in higher education. I then review the development of diagnostic multiple-choice tests of higher-level learning; normalized gain and ceiling effects; the documented two-sigma superiority of interactive engagement (IE) to traditional passive-student pedagogy in the conceptually difficult subject of Newtonian mechanics; the probable neuronal basis for such superiority; education's lack of a community map; higher education's resistance to change and its related failure to improve the public schools; and, finally, why we should be concerned with student learning.

1. Pre/post Paranoia

. . . gain scores are rarely useful, no matter how they may be adjusted or refined. . . . investigators who ask questions regarding gain scores would ordinarily be better advised to frame their questions in other ways. Cronbach & Furby (1970)

In a recent *Carnegie Perspective*, 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 "How we should measure "change," - or should we?"

* The reference is Hake, R.R. 2007a. "Should We Measure Change? Yes!" the latest version is online as ref. 43 at <<http://www.physics.indiana.edu/~hake>>. To appear as a chapter in *Evaluation of Teaching and Student Learning in Higher Education* [Hake (2007b)], a Monograph of the American Evaluation Association <<http://www.eval.org/>>. I welcome comments and suggestions directed to <rrhake@earthlink.net>.

† A severely truncated version of the present paper titled "Possible Palliatives for the Paralyzing Pre/Post Paranoia that Plagues Some PEP's" [Hake (2006m)] is online at <http://evaluation.wmich.edu/jmde/JMDE_Num006.html> in the November issue of the *Journal of MultiDisciplinary Evaluation*.

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

© Richard R. Hake, 15 January 2007. Permission to copy or disseminate all or part of this material is granted provided that the copies are not made or distributed for commercial advantage, and the copyright and its date appear. To disseminate otherwise, to republish, or to place at another website (instead of linking to <<http://www.physics.indiana.edu/~hake>>) requires written permission.

The dour appraisal of pre/post testing by Cronbach & Furby (1970) has echoed down through the literature to present day texts on assessment such as that by Suskie (2004a)]. 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 (2005a, 2006a)], is one reason for the glacial progress of educational research [Lagemann (2000)] and reform [Bok (2005a,b); Duderstadt (2000, 2001)] in higher education.

As for the unreliability of “change scores,” such charges by Lord (1956, 1958) and Cronbach & Furby (1970) have been called into question by e.g., Rogosa, Brandt, & Zimowski (1982), Zimmerman & Williams (1982), Rogosa & Willett (1983, 1985), Rogosa (1995), Wittmann (1997), Zimmerman (1997), & Zumbo (1999). Furthermore, the measurement of change is an active area of current research by psychometricians such as Collins and Horn (1991), Collins & Sayer (2001), Singer & Willett (2003), Lissitz (2005), and Liu & Boone (2006). All this more recent work should serve as a caution for those who dismiss measurements of change.

Aside from the alleged unreliability of change scores, seven other objections to pre/post testing, have been enumerated by Suskie (2004b)*, and countered by Hake (2004a)* and Scriven (2004)*. Suskie’s fourth objection (as listed by Hake) is:

If we do indeed see a significant gain, we often can't be sure it's due to our courses/program and not to other experiences . . . [“history”]. . . or normal maturation. A student might have a concurrent part-time job, for example, that has improved her oral communication skills far more than her required speech course.

2. The View from U.S. Department of Education

In science education, there is almost nothing of proven efficacy.

Grover Whitehurst, director, Institute of Education Sciences, USDE,
as quoted by Sharon Begley (2004)

“History” and maturation are among the nine threats to internal validity listed in Table 2.4 of Shadish et al. (2002), are discussed on pages 56-57 of that text, and are reiterated by the PEP-dominated “Coalition for Evidence-Based Policy” (CEBP) at the U.S. Dept. of Education [USDE (2003)]:

There is persuasive evidence that the randomized controlled trial, when properly designed and implemented, is superior to other study designs in measuring an intervention’s true effect.

1. “Pre-post” study designs often produce erroneous results.

Definition: A “pre-post” study examines whether participants in an intervention improve or regress during the course of the intervention, and then attributes any such improvement or regression to the intervention.

The problem with this type of study is that, without reference to a control group, it cannot answer whether the participants’ improvement or decline would have occurred anyway, even without the intervention. This often leads to erroneous conclusions about the effectiveness of the intervention.

*Throughout this paper I make frequent reference to discussion-list posts that contain serious academic discourse and references to the peer-reviewed literature. Consistent with “In Defense of Cross Posting” [Hake (2005c)], in a quixotic and generally futile attempt to tunnel through disciplinary barriers, I often transmit posts to several lists (sometimes including AERA-D, AERA-L, ASSESS, EvalTalk, ITForum, PhysLrnR, POD, and various chemistry, biology, mathematics, philosophy, and psychology lists), but for convenience I usually reference only the posts appearing on the *open* archives <<http://listserv.nd.edu/archives/pod.html>> of POD (Professional and Organizational Development).

But CEBP's criticism of pre/post testing is irrelevant for the recent pre/post studies in physics. The reason is that control groups *have* been utilized - they are the introductory courses taught by the traditional method. The matching is due to the fact that (a) within any one institution the test [Interactive Engagement (IE)] and control [Traditional (T)] groups are drawn from the same generic introductory course taken by relatively homogeneous groups of students, and (b) IE-course teachers in all institutions are drawn from the same generic pool of introductory course teachers who, judging from uniformly poor average normalized gains $\langle g \rangle$ they obtain in teaching traditional (T) courses, do not vary greatly in their ability to enhance student learning.

In fact, I suspect that the pre/post testing in physics, while not the so-called “gold standard” revered by the PEP-dominated CEBP [see e.g., Cook (2006) and the response by Scriven (2006)]. . . might be classified by the USDE’s “What Works Clearing House” <http://www.w-w-c.org/> as “Meets Evidence Standards with Reservations - strong *quasi-experimental studies*. . . [Shadish et al. (2002)]. . . that have comparison groups and meet other WWC Evidence Standards [see <http://www.w-w-c.org/reviewprocess/standards.html>].

3. Towards Valid and Consistently Reliable Diagnostic Tests

*What we assess is what we value. We get what we assess,
and if we don't assess it, we won't get it.*

Lauren Resnick [quoted by Grant Wiggins (1991)]

Of course, pre/post testing is only as good as the tests employed. In some fields, *disciplinary experts* have engaged, or are engaging, in the arduous quantitative and qualitative research required to develop valid and consistently reliable tests that probe for understanding of the basic concepts. A model for such effort is the pioneering but under-appreciated work of Halloun & Hestenes (1985a,b) in developing the *Mechanics Diagnostic Test*, precursor to the widely used *Force Concept Inventory* [Hestenes et al. (1992)]. Such test development was among the themes of a recent NSF (2006) Conference *Assessing Student Achievement (ASA)* <http://www.drury.edu/multinl/story.cfm?ID=17783&NLID=306>; for a report see PKAL (2006b)].

Examples of areas in which student learning gains are now being assessed* by means of pre/post testing are:

- (a) Newtonian mechanics [Halloun & Hestenes (1985a,b), Hestenes et al. (1992), and Thornton & Sokoloff (1998)];
- (b) other physics subjects as indicated at NCSU (2006), FLAG (2006), and Meltzer (2006b);
- (c) Astronomy, Economics, Biology, Chemistry, Engineering [see the references in Hake (2004b) & FLAG (2006)]; Geoscience [Libarkin & Anderson (2005, 2006a,b)]; and Calculus [Epstein (2006)].

*Throughout this paper I make no distinction between the terms *assessment* and *evaluation* as discussed in “Re: A taxonomy” [Hake (2003a)]. For comparison, the Joint Committee on Standards for Educational Evaluation [JCSEE (1994)] gives the following definitions:

assessment: “The act of determining the standing of an object on some variable of interest, for example, testing students, and reporting scores,” and

evaluation: “Systematic investigation of the worth or merit of an object; e.g., a program, project, or instructional material.”

The above mentioned *Force Concept Inventory* and *Mechanics Diagnostic* tests were used as pre/post tests in a survey Hake (1998a,b; 2002a,b,c) of 62 university, college, and high-school courses with a total enrollment of 6542 students. That study demonstrated that “Interactive Engagement” (IE) methods of instruction *can* yield average *normalized* gains $\langle g \rangle$ in conceptual understanding about two standard deviations [cf. Bloom’s “Two Sigma Problem” (1984)] greater than in courses using “traditional” (T) methods. In terms of *effect sizes*, the data of Hake (1998a,b) yield a Cohen effect size “d” = 2.4 [Hake (2002a)] and a Glass effect size “D” = 6.2 [for a discussion of D vs d see Thompson (2006, p. 189-192)].

Although ignored by most PEP’s, and even by some physicists [for example, those contributing to McCray, DeHaan, & Schuck (2003)], the above indicated research [Hake (1998a,b)] has been noted positively by workers in many different disciplines [astronomy, biology, chemistry, cognitive science, communication, economics, engineering, geoscience, mathematics, medicine, physics, and even psychology !]. See e.g. : Marchese (1997); Swartz (1999); Heller (1999); Zeilik et al. (1999); Breslow (1999, 2000); Rothman & Narum (1999); Nelson (2000); Albacete & VanLehn (2000); Stokstad (2001); Morote & Pritchard (2002); Savinainen & Scott (2002a,b); Dancy & Beichner (2002); Powell (2003); Elliott (2003); Klymkowsky et al. (2003); Wood & Gentile (2003); McConnell et al. (2003); Evans et al., (2003); Hegedus & Kaput (2004); Handelsman et al. (2004); Pavelich et al. (2004); Khodor et al. (2004); DeHaan (2005); Buck & Wage (2005); Smith et al. (2005); Hilborn (2005); Moore (2005); Wieman & Perkins (2005); Heron & Meltzer (2005); Kluck (2005); Bardar et al. (2006); Froyd et al. (2006); Nuhfer (2006a,b); and Michael (2006).

In addition my meta-analysis and subsequent confirmatory pre/post studies have at least partially stimulated the reform of a tiny fraction of introductory physics courses in the U.S., including large enrollment courses at Harvard [Crouch & Mazur (2001)], North Carolina State University [Beichner & Saul (2004)], MIT [Dori & Belcher (2004)], University of Colorado at Boulder [Pollock (2004)], and California Polytechnic State University at San Luis Obispo [Hoellwarth et al. (2005)].

4. Some Definitions and Explanations: Normalized Gain, Multiple Choice Tests (MCT's), Ceiling Effects, Interactive Engagement (IE) & Traditional (T) Courses

The true meaning of a term is to be found by observing what a man does with it, not what he says about it. Percy Bridgman (1927)

a. The average *normalized gain* $\langle g \rangle$ is the average *actual* gain $[\langle \%post \rangle - \langle \%pre \rangle]$ divided by the *maximum possible average gain* $[100\% - \langle \%pre \rangle]^* \dagger$, where the angle brackets indicate the class averages. This half-century-old gain parameter was *independently* employed by Hovland et al. (1949), who called g the “effectiveness index”; Gery (1972), who called g the “gap-closing parameter”; Hake (1998a,b), who called g the “normalized gain”; and Cohen et al. (1999), who had the good sense to call g what it *is*, namely “POMP” (Percentage Of Maximum Possible). In Hake (1998a,b), the use of the average *normalized gain* $\langle g \rangle$, rather than the average *actual* gain $\langle \%G \rangle = [\langle \%post \rangle - \langle \%pre \rangle]$ or the average posttest score $[\langle \%post \rangle]$ allowed meaningful comparison of *course effectiveness* for the 62 courses of the study. This was associated with the fact that the correlation of $\langle g \rangle$ with $\langle \%pre \rangle$ for the 62 courses was a very low +0.02. In contrast, the correlation of $\langle \%G \rangle$ with $\langle \%pre \rangle$ for the 62 courses was -0.49, and the correlation of $\langle \%post \rangle$ with $\langle \%pre \rangle$ for the 62 courses was +0.55. For the 62 courses in my survey $\langle \%pre \rangle$ ranged from 18% for a Dutch high school (below the random guessing level of 20%) to 71% for two Harvard courses [see Tables I(a,c) of Hake (1998b)]. Hence, I think it is probably generally true that $\langle \%G \rangle$, a Cohen or Glass “effect size” proportional to $\langle \%G \rangle$, and $\langle \%post \rangle$ are *not suitable for comparing course effectiveness over diverse groups with widely varying average pretest scores*. It can be shown that valid $\langle g \rangle$ comparison of course effectiveness requires that the test pose a performance ceiling effect (PCE) rather than an instrumental ceiling effect (ICE) as discussed in “f” below and in Hake (2006d,e,f,g,h).

*It is possible to calculate an average normalized gain for a course of N students by averaging the single-student normalized gains $g_i = [\%post_i - \%pre_i] / [100\% - \%pre_i]$ for all students in the course as $g(\text{ave}) = [S_{i=1 \text{ to } N} g_i] / N$. For N greater than about 20, $g(\text{ave})$ will usually be within about 5% of $\langle g \rangle$. As shown in footnote 46 of Hake (1998a), the close agreement is related to the fact that $[g(\text{ave}) - \langle g \rangle]$ is proportional to the correlation between g_i and pre_i , and can be either positive (positive correlation of g_i with pre_i) or negative (negative correlation of g_i with pre_i). In practice, the absolute correlation of g_i with pre_i is generally relatively low, just as the correlation between $\langle g \rangle$ and $\langle \%pre \rangle$ is relatively low for the 62 courses surveyed in Hake (1998a,b). For reasons discussed in Hake (2002b), it's better to use $\langle g \rangle$ rather than $g(\text{ave})$ in comparing the effectiveness of courses.

\dagger Two recent papers are relevant to the normalized gain:

(1) Bao (2006) discusses some mathematical properties of g , but omits mention of footnote 46 of Hake (1998a) - see above footnote - and gives an incomplete history of the use of normalized gain.

(2) Marx & Cummings (2007) advocate replacement of the single-student “normalized gain” g_i with single-student “normalized change” c_i , involving “the ratio of the gain to the maximum possible gain or loss to the maximum possible loss.” Although c_i might be useful in single-student gain analyses such as those by Meltzer (2002) and Hake (2002i), I think that replacing a class average $\langle g \rangle$ with a class average $\langle c_i \rangle$, as advocated by Marx & Cummings, has little, if any, advantage.

For a discussion of $\langle g \rangle$ in the context of the more psychometrically standard “Item Response Theory” (IRT) see Mislevy (2006). In my opinion, Mislevy’s objection that $\langle g \rangle$ is not “grounded in the framework of probability-based reasoning,” must be balanced against the:

- (1) *empirical* justification of $\langle g \rangle$ as an easy-to-use gauge of course effectiveness in hundreds of studies of classroom teaching in widely varying types of courses and institutions with widely varying types of instructors and student populations,
- (2) evidently unsolved problem of how to employ IRT *to compare the effectiveness of courses* in which the initial average knowledge state of students is highly variable,
- (3) difficulties that average faculty members might experience in using IRT to improve the effectiveness of their courses, and
- (4) dearth of examples of the constructive employment of IRT in *higher* [as opposed to K-12 (Pellegrino et al., 2001)] education research on classroom teaching - although Libarkin & Anderson’s (2005, 2006a,b) work affords as example IRT applied to pre/post testing in higher education.

Incidentally, IRT resources are available online in the book by Baker (2001) and the excellent IRT resource and tutorial page by Rudner (2001). See also physicist/astronomer Phil Sadler’s (1999) discussion of IRT in his chapter “The Relevance of Multiple Choice Testing in Assessing Science Understanding.” A recent article “Testing the test: Item response curves and test quality” [Morris et al. (2006)] employs “item response curve” (IRC) analysis - a simplified version of IRT - to evaluate multiple-choice questions.

b. 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.

c. Can MCT’s measure conceptual understanding and higher-order learning? [For a cogent discussion of higher-order learning see Shavelson & Huang (2003).] 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] . . . is an assessment that uses multiple-choice items to tap into higher-level cognitive processes

d. Instrumental Ceiling Effect (ICE) and a Performance Ceiling Effect (PCE) - what's the difference? An example of an ICE would be an adult high-jump training camp where the range of jump-height measurement is restricted to 0-5 ft by antiquated equipment, so that an ICE would exist at 5 ft. An example of a PCE would be an adult high-jump training camp where range of jump-height measurement is 0-10 ft. Then a PCE would exist at about 8 ft, the Olympic performance record. Use of the average normalized gain $\langle g \rangle$ in *gauging the relative effectiveness* of training camps in increasing the average jump height of their trainees (even though the average pre-camp jump height varies drastically among camps) would be reasonable *if* a PCE prevailed in all camps, but less reasonable if an ICE prevailed. Similarly, use of the average normalized gain $\langle g \rangle$ in *gauging the relative effectiveness* of introductory mechanics classes in increasing students' conceptual understanding (even though the average pretest score varies drastically among courses) is reasonable if the test employed (e.g., the FCI) poses a PCE. That this is, in fact, the case is argued in Hake (2006d,e,f,g,h). My own cursory literature search suggested that the important distinction between an ICE and a PCE has not previously been made in the psychometric literature.

e. "Traditional" (T) courses are *operationally* defined 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.

f. Interactive Engagement (IE) courses are *operationally* defined as those designed at least in part to promote conceptual understanding through *continual* interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield *immediate* feedback through discussion with peers and/or instructors. Among the IE methods used in the courses surveyed by Hake (1998a,b) were *Microcomputer-based Labs*, *Concept Tests*, *Modeling*, *Active Learning Problem Sets*, *Overview Case Studies*, and *Socratic Dialogue Inducing Lab*, all developed by physicists [for references to these methods see Hake (1998a,b). For specification of learning theories relevant to each (following Ken Heller, 1999)] see Hake (2002a)]. The primary theories are [for the references see Hake (2002a)]:

- (1) "Developmental Theory," originating with Piaget [see, e.g., Inhelder and Piaget (1958); more recently see Gardner (1985), Inhelder et al. (1987), Phillips and Soltis (1998)];
- (2) "Cognitive Apprenticeship" [Collins et al. (1989), Brown et al. (1989)];
- (3) "Social Interaction" [Vygotsky (1978), Lave and Wenger (1991), Dewey (1997), Phillips and Soltis (1998)].

Because the intensity of the interaction that occurs in IE physics courses does not seem to be widely appreciated, it may be worthwhile to quote the description of a fairly typical IE method [Hake (1992) - with updated references]:

Socratic Dialogue Inducing (SDI) labs have been shown [Hake (1998a, 1998b - Table Ic)] to be relatively effective in guiding students to construct a coherent conceptual understanding of Newtonian mechanics. The SDI method might be characterized as “*guided construction*,” rather than “guided discovery” or “inquiry.” We think the efficacy of SDI labs is primarily due to the following essential features:

- (1) interactive engagement of students who are induced to think constructively about simple Newtonian experiments which produce conflict with their commonsense understandings;
- (2) the Socratic method [e.g., Arons (1973, 1974, 1990, 1993, 1997); Hake (1992, 2002 f,g,h,j)] of the *historical* Socrates [Vlastos (1990, 1991, 1993)] (not Plato’s alter ego in the *Meno*!, as mistakenly assumed by many - even some physicists) utilized by experienced instructors who have a good understanding of the material and are aware of common student preconceptions and failings;
- (3) considerable interaction between students and instructors and thus a degree of individualized instruction;
- (4) extensive use of multiple representations (verbal, written, pictorial, diagrammatic, graphical, and mathematical) to model physical systems;
- (5) real world situations and kinesthetic sensations (which promote student interest and intensify cognitive conflict when students’ direct sensory experience does not conform to their conceptions);
- (6) cooperative group effort and peer discussions;
- (7) repeated exposure to the coherent Newtonian explanation in many different contexts.

Of course, the SDI method, as other IE methods surveyed in Hake (1998a,b), bears near zero resemblance to the extreme “discovery learning,” researched by Klahr & Nigam (2004) and widely interpreted [most recently by Kirschner et al. (2006)] as demonstrating the superiority of “direct instruction.”

But wait! What do Klahr & Nigam and Kirschner et al. *mean* by “discovery learning,” and what do they *mean* by “direct instruction”? In discussion list posts Hake (2004f,g) criticizing the California Curriculum Commission’s (CCC’s) anti-hands-on* “Criteria For Evaluating K-8 Science Instructional Materials In Preparation for the 2006 Adoption,” I opined that popular pedagogic terms such as “discovery learning,” “direct instruction,” “hands-on activities,”

*Since then the CCC, after “a session with the executive director of the state school board, representatives of Gov. Arnold Schwarzenegger, business leaders, legislators, higher education officials, teachers, and officers of the California Science Teachers Association” [Galley (2004b)], amazingly flip-flopped from being vehemently *anti*-hands-on science instruction [Galley (2004a)] to being nominally *pro*-hands-on science instruction [Galley (2004b); CSTA (2006a,b)]. But, unfortunately, the switch may be more cosmetic than real, as indicated in “Direct Science Instruction Suffers a Setback in California - Or Does It?” [Hake (2004e)] and “Will the No Child Left Behind Act Promote Direct Instruction of Science?” [Hake (2005e)].

“active learning,” “cooperative learning,” “inquiry,” and “interactive engagement,” should be *operationally defined* [see, e.g. Holton & Brush (2001), Phillips (2000)], even despite the “anti-positivist vigilantes” [Phillips (2000), Phillips & Burbules (2000)]. More generally, rigorous operations should be defined for distinguishing pedagogic method X from other methods Y, Z, A, B, C, . . .

Although operational definitions are uncommon in the educational literature, in “Re: Back to Basics vs. Hands-On Instruction” [Hake (2004f) - *see that post for the references below*] I indicated my own guesses as to what various authors have meant by “direct instruction,” writing:

I suspect that “direct instruction” means to:

(a) *Mathematically Correct* <<http://mathematicallycorrect.com/science.htm>>: “drill and practice,” “non-hands-on,” “teach ‘em the facts” [Metzenberg (1998)], and “non-discovery-learning,” where “discovery learning” means setting students adrift either in aimless play or ostensibly to discover on their own, say, Archimedes’ principle or Newton’s Second Law.

(b) *Physics Education Researchers* (PER’s): traditional *passive student* lectures, recipe labs, and algorithmic problem sets.

(c) *Klahr & Nigam* (2004): . . . instruction in which “the goals, the materials, the examples, the explanations, and the pace or instruction are all teacher controlled,” but in which hands-on activities *are* featured. At least this is Klahr & Nigam’s (KN’s) definition of what they call “*extreme direct instruction*” (extreme DI), possibly having in mind the reasonable idea of a continuum of methods from extreme DI to extreme “discovery learning” (DL). In extreme DL, according to KN, there is “no teacher intervention beyond the suggestion of a learning objective: no guiding questions, and no feedback about the quality of the child’s selection of materials, explorations, or self-assessments.” I suspect that KN might classify “interactive engagement” methods [Hake (1998a,b)] and inquiry methods [NRC (1997, 1999, 2000)]; Bransford et al. (1999); Donovan et al. (1999)] as somewhere along a continuum ranging from extreme DI to extreme DL, since “interactive engagement” and “inquiry” methods both involve various degrees of judicious teacher intervention so as to guide students’ conceptual understanding, problem solving abilities, and process skills towards those of professionals in the field.

(d) *Association of Direct Instruction* (ADI 2004):

(1) teaching by telling (as contrasted by teaching by implying), or

(2) instructional techniques based on choral responses, homogeneous grouping, signals, and other proven instructional techniques, or

(3) specific programs designed by Siegfried Engelmann and his staff.

Direct Instruction programs incorporate the above “2” coupled with carefully designed sequences, lesson scripting, as well as responses to anticipated children’s questions as expounded in Englemann & Carnine (1982).

And the abstract of “Best Practices in Science Education #2 ” [Hake (2006n)] reads:

I agree with David Brookes [2006] that Klahr & Nigam’s (KN) (2004) heavily publicized study represents a “double straw man” in that (a) the “extreme discovery learning” that they reported to be inferior to “extreme direct instruction” is almost never used in the classroom, and (b) “extreme direct instruction” means to KN pedagogy rather similar to the “interactive engagement” methods shown to be relatively effective by physics education researchers.

Thus the *interpretation* of Klahr and Nigam (2004) that “direct instruction” (as defined by KN) is superior to “discovery learning” (as defined by KN), while consistent with KN’s research, appears to be a *misinterpretation* to physics education researchers (PER’s) if they use the PER definition of “direct instruction,” and are unaware of the KN definitions of “direct instruction” and “discovery learning.” Thus there appears to be a communication failure involving different meanings for these terms.

A recent example of what I would regard as a communication failure is provided by the previously mentioned paper of Kirschner et al. (2006) with its seemingly (to some - including myself) non-sequitur title “Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching.”[♦] This provocative treatise has been extensively discussed in the 7-post thread “Why Minimal Guidance During Instruction Does Not Work” at <<http://tinyurl.com/yatob9>>[†] on the Math-Learn discussion list of 24-26 October 2006; and the 43-post thread “Article of interest” at <<http://listserv.boisestate.edu/cgi-bin/wa?A1=ind0612&L=physlrnr> - 2>^{*} on the PhysLrnR discussion-list of 5-17 December 2006.

Kirschner et al. wrote:

Klahr and Nigam (2004), in a very important study, not only tested whether science learners learned more via a discovery versus direct instruction route but also, once learning had occurred, whether the quality of learning differed. Specifically, they tested whether those who had learned through discovery were better able to transfer their learning to new contexts. The findings were unambiguous. Direct instruction involving considerable guidance, including examples, resulted in vastly more learning than discovery. Those relatively few students who learned via discovery showed no signs of superior quality of learning.

[♦] A Google <<http://www.google.com/>> search for “Why Minimal Guidance During Instruction Does Not Work” (with the quotes) yielded 256 hits on 15 Jan 2007 21:41:05-0800, indicating that, as might be expected, the provocative title “Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching.” of Kirschner et al. has drawn considerable attention on the internet.

[†] One must subscribe to MathLearn in order to access it’s archives but it takes only a few minutes to subscribe by clicking on the archive URL <<http://groups.yahoo.com/group/math-learn/>>, and then subscribing.

^{*}One must subscribe to PhysLrnR in order to access it’s archives but it takes only a few minutes to subscribe by clicking on the archive URL <<http://listserv.boisestate.edu/archives/physlrnr.html>>, and then clicking on “Join or leave the list (or change settings).” 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!

But here again, as with Klahr and Nigam (2004), “direct instruction” appears to mean to Kirschner et al. (2006), pedagogy rather similar in some respects to the “interactive engagement” methods shown to be relatively effective by physics education researchers, as can be seen from:

a. The contrast of “unguided or minimally guided instruction” with “direct instructional guidance” in the first paragraph of Kirschner et al. [my *italics*, my insert at “. . . [insert]. . .”; see their article for references other than Klahr & Nigam (2004)]:

Disputes about the impact of instructional guidance during teaching have been ongoing for at least the past half-century (Ausubel, 1964; Craig, 1956; Mayer, 2004; Shulman & Keisler, 1966). On one side of this argument are those advocating the hypothesis that *people learn best in an unguided or minimally guided environment, generally defined as one in which learners, rather than being presented with essential information, must discover or construct essential information for themselves* (e.g., Bruner, 1961; Papert, 1980; Steffe & Gale, 1995). On the other side are those suggesting that novice learners should be provided with direct instructional guidance on the concepts and procedures required by a particular discipline and should not be left to discover those procedures by themselves (e.g., Cronbach & Snow, 1977; Klahr & Nigam, 2004; Mayer, 2004; Shulman & Keisler, 1966; Sweller, 2003). *Direct instructional guidance is defined as providing information that fully explains the concepts and procedures students are required to learn. . .* [It would be a great boon to introductory physics instructors if Kirschner et al. could tell them how to “provide information that fully explains the concepts and procedures” relative to e.g., Newton’s Second Law of Motion]. . . *as well as learning strategy support that is compatible with human cognitive architecture*. Learning, in turn, is defined as a change in long-term memory.

b. The Kirschner et al. abstract [my insert at “. . . [insert]. . .”]:

Evidence for the superiority of guided instruction. . . [read “interactive engagement”?]. . . is explained in the context of our knowledge of human cognitive architecture, expert–novice differences, and cognitive load. Although unguided or minimally guided instructional approaches are very popular and intuitively appealing, the point is made that these approaches ignore both the structures that constitute human cognitive architecture and evidence from empirical studies over the past half-century that consistently indicate that minimally guided instruction is less effective and less efficient than instructional approaches that place a strong emphasis on guidance of the student learning process. The advantage of guidance begins to recede only when learners have sufficiently high prior knowledge to provide “internal” guidance. Recent developments in instructional research and instructional design models that support guidance during instruction are briefly described.

That education specialists Kirschner et al. have failed to communicate to most physics education researchers (PER’s) [and probably to most disciplinary education researchers] is suggested by the PhysLrnR post of prominent PER David Meltzer (2006c) who wrote [my inserts -NOT Meltzer’s! - at “. . . . [Hake’s insert]. . .”]:

A key part of the article by Kirschner, Sweller, and Clark reads as follows (p. 82):
“The work of Klahr and Nigam. . . [2004]. . . unambiguously demonstrated the advantages of direct instruction in science. There is a wealth of such evidence. A series of reviews by the U.S. National Academy of Sciences has recently described the results of experiments that provide evidence for the negative consequences of unguided science instruction. . . . [McCray et al. (2003)]. . . reviewed studies and practical experience in the education of college undergraduates in engineering, technology, science, and mathematics. . . (this and other publications) amply document the lack of evidence for unguided approaches and the benefits of more strongly guided instruction. . . .”

Chapter 3 of . . . [McCray et al. (2003)]. . . is entitled “Evaluating Effective Instruction.” The full list of physics instructional methods cited as “exemplary” in this chapter. . . [consistent with the tunnel vision of physicists contributing to McCray et al. (2003), “Socratic Dialogue Inducing” Labs (Hake, 1987, 1992, 2002j; Tobias & Hake (1988) were omitted] . . . is as follows:

- Context-rich cooperative problem solving
- Explorations in Physics
- Interactive Physics
- Just in Time Teaching
- [Interactive] lecture demonstrations
- Models in Physics Instruction
- Peer Instruction
- Physics by Inquiry
- Powerful Ideas in Physical Science
- Real Time Physics
- Studio Physics
- Tutorials in Introductory Physics
- Workshop Physics

(Several of these methods are discussed in some detail.)

The Summary of this chapter states:

“Accumulating research shows that the traditional didactic lecture format . . . [this is what “direct instruction” means to most education researchers in the disciplines, but not to Klahr & Nigam (2004) or to Kirchner et al. (2006)]. . . can support memorization of factual information but may be less effective than other instructional strategies in promoting understanding of complex concepts or the ability to apply such concepts in new situations. Instructional programs known to the workshop participants to be effective in eliciting such learning. . . . recognize that students have diverse learning styles; provide varied experiences for students to develop functional understanding of a subject; promote students’ ability to work cooperatively and to communicate orally and in writing; invest in training and mentoring of instructors; and promote research on teaching and learning....

An important element of effective instruction involves engaging students’ preconceptions and prior beliefs in ways that help them achieve a more mature understanding. Effective instructional strategies for correcting misconceptions and producing conceptual understanding for most students require situations that demand active intellectual engagement, such as tutorials, small group learning, hands-on activities, case studies, and problem-solving exercises with appropriate scaffolding. Scaffolding (i.e., support and guidance in learning specific concepts or tasks) can be provided by an expert (instructor, teaching assistant, or peer learning coach) or by a computer program. When instructors employ effective instructional strategies of the types described, they model in their teaching the ways in which their own students should teach if those students go on to become graduate teaching assistants, K–12 teachers, or science faculty.”

Now.... does this mean that Kirschner et al. are clearly and consciously presenting all of those standard PER-based instructional approaches as shining examples of what they call “direct instruction,” . . . [Hake’s guess is that they are]. . . or is it rather the case that they failed to read and/or understand the extensive discussion in . . . [McCray et al. (2003)]. . . regarding the principles and methods of research-based physics instruction? I don’t know, but if nothing else this underscores the utter futility of attempting to discuss these issues in terms of meaningless empty phrases such as “direct instruction,” “unguided instruction,” “discovery learning,” etc. *If there is no direct and explicit reference to the actual details of the instructional methods, it is pointless (I think) to get involved in such a debate.* [My italics - well said David!]

The dreadful abyss that appears to separate many psychologists, education specialists, and psychometricians (PEP’s) from education researchers and developers in the disciplines is, I think, well illustrated by the above discussion.

5. Why Are Interactive Engagement (IE) Courses More Effective Than Traditional (T) Passive-Student Courses?

The Brain . . . Use It or Lose It. . . no matter what form enrichment takes, it is the challenge to the nerve cells that is important. Data indicate that passive observation is not enough; one must interact with the environment. Marian Diamond (1996)

The superiority of IE methods in promoting conceptual understanding and higher-order learning is probably related to the “enhanced synapse addition and modification” induced by those methods. Cognitive scientists Bransford et al. (1999, 2000) stated:

. . . 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.

Consistent with the above, biologist Robert Leamson (1999, 2000) has stressed the relationship of biological brain change to student learning. In his first chapter “Thinking About Thinking and Thinking About Teaching,” Leamson (1999) defines teaching and learning thusly [my italics]:

. . . teaching means any activity that has the conscious intention of, and potential for, *facilitation of learning* in another. . . . *Learning is defined as stabilizing, through repeated use, certain appropriate and desirable synapses in the brain.* . . .

And biologist James Zull (2003) in “What is *The Art of Changing the Brain?*” wrote [my italics]:

Although the human brain is immensely complicated, we have known for some time that it carries out four basic functions: getting information (sensory cortex,) making meaning of information (back integrative cortex), creating new ideas from these meanings (front integrative cortex,) and acting on those ideas (motor cortex.). . . [for Zull’s schematic of the brain see <<http://www.case.edu/artsci/biol/people/zull.html>>]. . . . From this I propose that there are four pillars of human learning: gathering, analyzing, creating, and acting. This isn’t new, but its match with the structure of the brain seems not to have been noticed in the past. *So I suggest that if we ask our students to do these four things, they will have a chance to use their whole brain.*

For pro and con articles on the relevance of neuroscience to present-day classroom instruction see e.g., PRO: Lawson (2006) and Willis (2006); CON: Marchese (2002) and Bruer (1997, 2006). See also the commentary on Willis (2006) and Bruer (2006) by Hake (2006i) and Redish (2006).

Despite controversy regarding the applicability of neuroscience research to *current* classroom practice, few would doubt its *potential* value. Neuroscientists Squire & Kandel (2000) write:

From a cognitive point of view, cellular and molecular approaches have made headway toward answering some of the key unsolved problems in the psychology of memory. What is the molecular relationship between nondeclarative [“knowing how,” sometimes called “operative” or “procedural” (Anderson, 2004)]. . . . and declarative . . . [“knowing that”]. . . .memory storage. How do short-term forms of memory relate to long-term forms? Most important, molecular approaches have provided an initial bridge. . . [cf., “A Bridge Too Far,” Bruer (1997)]. . . .connecting behavior in intact animals to molecular mechanisms in single cells. Thus, what were formally only psychological constructs like association, learning, storing, remembering, and forgetting can now be approached in terms of cell and molecular mechanisms and in terms of brain circuits and brain systems. In this way, it has become possible to achieve deep insights into the fundamental questions about learning and memory.

6. Is There an Education Community Map?

It is not enough to observe, experiment, theorize, calculate and communicate; we must also argue, criticize, debate, expound, summarize, and otherwise transform the information that we have obtained individually into reliable, well established, public knowledge.

John Ziman (1969).

The approximately two-sigma superiority of IE over T courses in introductory mechanics has been independently corroborated in hundreds of courses with widely varying types of instructors, institutions, and student populations [see e.g., the references in Hake (2002a,b)], thus satisfying Shavelson & Towne’s (2002) fifth principle of good scientific practice [my *italics*]:

Replicate and Generalize Across Studies: By one replication we mean, at an elementary level, that if one investigator makes a set of observations, another investigator can make a similar set of observations under the same conditions At a somewhat more complex level, *replication means the ability to repeat an investigation in more than one setting (from one laboratory to another or from one field site to a similar field site) and reach similar conclusions.*

Associated with the need for replication, good science requires “continual interaction, exchange, evaluation, and criticism so as to build a . . . *community map*” [Redish (1999)]. The latter crucial feature of the scientific method has also been emphasized by, e.g., Gottfried & Wilson (1997), Cromer (1997), Gere (1997), Newton (1997), Ziman (2000), and Hake (2000a), *but does not generally characterize education research*, as a visit to most (but not all) American Educational Research Association (AERA) discussion lists will demonstrate [Hake (2005b)]. In fact, whether or not an “education community” even exists is problematic. Lagemann (2000, p. 239) writes [my *italics*]:

. . . . there are very few filters of quality in education. There is neither a Better Business Bureau nor the equivalent of the Federal Food and Drug Administration. Caveat emptor is the policy of this field. This is because *education research has never developed a close-knit professional community*, which is the prerequisite for the creation of regulatory structures that can protect both the welfare and safety of the public at large and the integrity of the profession. Such communities exist in some disciplines, for example, physics, and, to a lesser extent, psychology; they also exist in some professions, notably medicine and law. *But such a community has never developed in education.*

7. Resistance To Change: Why Is Higher Education So Slow to Recognize the Value of Interactive Engagement Methods in Promoting Higher-Level Learning?

It's harder to move the faculty than a graveyard.

Richard Cyert, former president of Carnegie Mellon University

The lack of a reliable community map for education is doubtless one of the reasons for the failure of higher education to recognize and benefit from research advances. Among the heavy hitters who have come to the plate for the last-place reform team are [my *italics* and inserts at “. . . [insert]. . .”]:

Derek Bok (2005b), leading off as former president of Harvard University, wrote in “Are colleges failing? Higher ed needs new lesson plans”:

. . . 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*. Other research has documented the widespread use of other practices that impede effective learning, such as the lack of prompt and adequate feedback on student work, the prevalence of tests that call for memory rather than critical thinking, and the reliance on teaching methods that allow students to do well in science courses by banking on memory rather than truly understanding the basic underlying concepts.

James Duderstadt, former president of the University of Michigan, in *A University for the 21st Century*, made much the same point, writing:

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.*

Richard Cyert, former president of Carnegie Mellon University, wrote in Tuma & Reif (1980):

The academic area is one of the most difficult areas to change in our society. We continue to use the same methods of instruction, particularly lectures, that have been used for hundreds of years. Little scientific research is done to test new approaches, and little systematic attention is given to the development of new methods. *Universities that study many aspects of the world ignore the educational function in which they are engaging and from which a large part of their revenues are earned.*

*Yes, I know, leeches are finding some *isolated* use in modern medicine

<<http://www.pbs.org/wgbh/nova/sciencenow/dispatches/050915.html>>, but that doesn't invalidate Duderstadt's point. Nor does the observation that passive student lectures (PSL's) can be used to advantage for material that can be rote memorized invalidate the point that for higher-level learning PSL's are relatively ineffective.]

Fred Reif (1974), batting clean up as a physics-education-research pioneer, in commentary as relevant today as it was 32 years ago, wrote:

The university does not systematically encourage faculty members to turn their talents to educational endeavors; in fact such endeavors are usually regarded as being of dubious legitimacy compared to more prestigious activities. . . . educational innovations are few in number and often marginal in their impact. Nor is this situation surprising, since the university, unlike any progressive industry, is not in the habit of improving its own performance by systematic investment in innovative research and development. Indeed the resources allocated by the university to educational innovation are usually miniscule or nonexistent. . . . The present educational role of the University is neither desirable nor necessarily immutable. Is it too far fetched to suggest that the university should take education at least as seriously as the Bell Telephone Company takes communication? The university would then be expected to further progress in education by engaging deliberately in suitable research, development, and deployment. In short, it would have to view its mission, in education as well as in other functions, as one of carrying out excellent work, developing new ideas and methods, and fostering the diffusion of innovations throughout the rest of society.

And Ted Marchese (2006), former vice president of the American Association for Higher Education (AAHE) and former editor of *Change* magazine, lamenting the apparent absence of reform in undergraduate education since the 1990's, wrote:

What's at stake? Does this matter? Does it matter that university completion rates are 44 percent and slipping? That just 10 percent from the lowest economic quartile attain a degree? *That figures released this past winter show huge chunks of our graduates who cannot comprehend a New York Times editorial or their own checkbook?* That frustrated public officials edge closer and closer to imposing a standardized test of college outcomes? . . . [see, e.g., USDE (2006)]. . . Does it matter that we look to our publics like an enterprise more eager for status and funding than self-inquiry and improvement? *When our absolute core function-undergraduate teaching and learning-runs on yesterday's ideas, it runs on empty.* Good as yesterday's ideas may be, I fear we are not asking hard, new questions about that function, producing new intellectual capital, and hatching new idea champions.

With regard to science education, Robert DeHaan (2005), the Charles Candler Professor of Cell Biology Emeritus, at Emory University, in his abstract to “The Impending Revolution in Undergraduate Science Education” wrote:

There is substantial evidence that scientific teaching in the sciences, i.e., teaching that employs instructional strategies that encourage undergraduates to become actively engaged in their own learning, can produce levels of understanding, retention and transfer of knowledge that are greater than those resulting from traditional lecture/lab classes. *But widespread acceptance by university faculty of new pedagogies and curricular materials still lies in the future.*

And biologists William Wood & and James Gentile (2003) wrote [references have been changed to match those in the reference list below]:

Unknown to many university faculty in the natural sciences, particularly at large research institutions, is a large body of recent research from educators and cognitive scientists on how people learn [Bransford et al. (2000)]. The results show that *many standard instructional practices in undergraduate teaching, including traditional lecture, laboratory, and recitation courses, are relatively ineffective at helping students master and retain the important concepts of their disciplines over the long term.* Moreover, these practices do not adequately develop creative thinking, investigative, and collaborative problem-solving skills that employers often seek. *Physics educators have led the way in developing and using objective tests* [Hestenes et al. (1992), Hake (1998a), NCSU (2006)] to compare student learning gains in different types of courses, and chemists, biologists, and others are now developing similar instruments [Mulford & Robinson (2002), Klymkowsky et al. (2003), Klymkowsky (2006)]. *These tests provide convincing evidence that students assimilate new knowledge more effectively in courses including active, inquiry-based, and collaborative learning, assisted by information technology, than in traditional courses* [Hake (1998a), NCSU (2006)].

In a masterful review “Where's the evidence that active learning works?” medical education researcher/developer Joel Michael (2006) stated:

There is a long history of educational research being done by the physics education community, and a long history of attempts to reform the teaching of physics. Some of the first, and still most striking, demonstrations of misconceptions were uncovered in the domain of Newtonian motion. (McCloskey (1983) . . . [a psychologist!]. . . discusses this early work). This work on physics misconceptions led in a fairly direct way to the development of the force concept inventory (FCI) by Hestenes et al. (1992), *an assessment tool that focuses on students' understanding of the concepts of physics rather than the ability to analyze problems and solve them using equations.* There are now a number of similar inventories in other areas of physics. The availability of an assessment tool like the FCI has provided a valuable tool for the classroom teacher [Savinainen & Scott (2002a,b)] but it has also provided a valuable tool for the educational research community. Although there are controversies about the FCI. . . [see e.g., Hestenes & Halloun (1995)]. . . it has provided a measure of student conceptual learning that can be used in asking questions about the efficacy of different approaches to teaching physics. One of the most striking findings [Hake (1998a,b)] came from comparison of the learning outcomes (as measured by the FCI and a related inventory on mechanics) from 14 traditional courses (2,084 students) and 48 courses using "interactive-engagement" (active learning) techniques (4,458 students). *The results on the FCI assessment showed that students in the interactive-engagement courses outperformed students in the traditional courses by 2 SDs.* Similarly, students in the interactive-engagement courses outperformed students in the traditional courses on the. [Mechanics Baseline Test (Hestenes & Wells, 1992)]. . . , a measure of problem-solving ability. *This certainly looks like evidence that active learning works! Research in physics education is having a profound effect on the development of instructional materials.* Research to determine what students know and can do and what they don't know or can't do in the domain of electricity by [McDermott & Schaffer (1992)]. . . has thus led to the development of instructional materials by this same group that do a better job of helping students master this subject matter. A similar approach has been pursued in the domain of optics [Wosilait et al. (1998)]. The interactive-engagement physics courses studied by Hake (1998a,b) used instructional approaches based on research findings from within the physics education community.

Education leaders Bok, Duderstat, Cyert, Reif, Marchese, DeHaan, and Wood & Gentile all deplore the general failure of higher education to benefit from research and development of the type reviewed by Michael. Such resistance to change is consistent with Lesson #13 of the physics education reform effort [Hake (2002a)]:

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

Inertia is an oft-lamented condition in many systems - see e.g., “Eleven Quotes in Honor of Inertia” [Hake (2006b)] and “Re: Innovation and change vs. the status quo” [Hake (2006c)]. But few systems can match the monumentally ponderous inertia of the U.S. educational system. Eight years ago a special issue of *Daedalus* (1998) contained essays by researchers in education and by historians of more rapidly developing institutions such as power systems, health care, communications, and agriculture. The issue was intended to help answer a challenge posed by physics Nobelist Kenneth Wilson:

If other major American “systems” have so effectively demonstrated the ability to change, why has the education “system” been so singularly resistant to change? What might the lessons learned from other systems’ efforts to adapt and evolve have to teach us about bringing about change—successful change—in America’s schools?

8. The Failure of Higher Education to Improve the Public Schools

Although we in higher education are very skillful at ignoring the obvious, it is gradually dawning on some of us that we bear a substantial part of the responsibility for this sad situation [in K-12 education].

Don Langenberg [BHEF (2001), p. 23], physicist and (at the time) Chancellor of the University of Maryland System

Judging from the contents of the special education/learning issues of *Daedalus* (1998, 2002, & 2004), the American Academy of Arts and Sciences <<http://www.amacad.org/>> (former publishers of *Daedalus*) has failed to find contributors who can satisfactorily answer the question posed above by Kenneth Wilson. Nor, as far as I know, has any other journal or publisher. But could it be that the U.S. education system is so singularly resistant to change because, in part, higher education has failed to properly educate prospective K-12 teachers and administrators? The NSF’s (1996) report *Shaping the Future* hit the nail on the head [my insert at “. . . [insert]. . .”]:

Many faculty in SME&T. . . [Science, Mathematics, Engineering, & Technology]. . . at the post-secondary level continue to blame the schools for sending underprepared students to them. But, increasingly. . . [but not conspicuously]. . . the higher education community has come to recognize the fact that *teachers and principals in the K-12 system are all people who have been educated at the undergraduate level, mostly in situations in which SME&T programs have not taken seriously enough their vital part of the responsibility for the quality of America’s teachers.*

The failure of higher education to play a substantive role in the improvement of K-12 education has been lamented by, e.g., [my *italics* and inserts at “. . .[insert]. . .”]:

a. Sherman Stein (1997), writing of mathematics education, but his comments apply as well to other branches of education:

The first stage in the reform movement should have been to improve the mathematical knowledge of present and prospective elementary teachers. Unfortunately, the cart of curriculum reform has been put before the horse of well-prepared teachers. . . .If all teachers were mathematically well prepared, I for one would stop worrying about the age-old battle still raging between “back to basics” and “understanding.” *On the other hand, if mathematics departments do nothing to improve school mathematics, they should stop complaining that incoming freshmen lack mathematical skills.*

b. Herbert Clemens (1989), again concerned with math education, but he could have been talking about almost any discipline:

Why don't mathematicians from universities and industry belong in math education? The first reason is that it is self-destructive. *The quickest way to be relegated to the intellectual dustbin in the mathematics departments of most research universities today is to demonstrate a continuing interest in secondary. . . .[or even worse, primary]. . . mathematics education.* Colleagues smile tolerantly to one another in the same way family members do when grandpa dribbles his soup down his shirt. Math education is certainly an acceptable form of retiring as a mathematician, like university administration (unacceptable forms being the stock market, EST. . . .[Erhard Seminar Training?]). . . , or a mid-life love affair). *But you don't do good research and think seriously about education.*

c. Kati Haycock (1999), Director of the Education Trust <<http://www2.edtrust.org/edtrust>>, who wrote:

Higher education . . . (unlike Governors and CEO's) . . . has been left out of the loop and off the hook (in the effort to improve America's public schools since the release of *A Nation at Risk* in 1983). . . . Present neither at the policy tables where improvement strategies are formulated nor on the ground where they are being put into place, most college and university leaders remain blithely ignorant of the roles their institutions play in helping K-12 schools get better—and the roles they currently play in maintaining the status quo *How are we going to get our students to meet high standards if higher education continues to produce teachers who don't even meet those same standards?* How are we going to get our high school students to work hard to meet new, higher standards if most colleges and universities will continue to admit them regardless of whether or not they even crack a book in high school?

Unfortunately the passive-student lecture method generally employed in higher education generally fails to provide average prospective K-12 teachers with the deep content knowledge and *pedagogical*-content knowledge [Shulman (1986, 1987)] that's required to promote student learning. According to Arthur Levine's (2006) report *Educating School Teachers* [my *italics*]:

The nation's teacher education programs are inadequately preparing their graduates to meet the realities of today's standards-based, accountability-driven classrooms, in which the primary measure of success is student achievement. A new study conducted by Arthur Levine, who recently left the presidency of Teachers College, Columbia University to become president of the Woodrow Wilson National Fellowship Foundation. . . .[<<http://www.woodrow.org/>>]. . . , concludes that a majority of teacher education graduates are prepared in university-based programs that suffer from low

admission and graduation standards. Their faculties, curriculums and research are disconnected from school practice and practitioners. There are wide variations in program quality, with the majority of teachers prepared in lower quality programs. Both state and accreditation standards for maintaining quality are ineffective. . . . *Many students seem to be graduating from teacher education programs without the skills and knowledge they need to be effective teachers.* More than three out of five teacher education alumni surveyed (62 percent) report that schools of education do not prepare their graduates to cope with the realities of today's classrooms. . . . *Universities use their teacher education programs as "cash cows," requiring them to generate revenue to fund more prestigious departments.* This forces them to increase their enrollments and lower their admissions standards. Schools with low admissions standards also tend to have low graduation requirements. While aspiring secondary school teachers do well compared to the national average on SAT and GRE exams, the scores of future elementary school teachers fall near the bottom of test takers. Their GRE scores are 100 points below the national average.

The crucial importance of effective K-12 teachers has been emphasized by, e.g., [my *italics*]:

a. John Goodlad (1990):

Few matters are more important than the quality of the teachers in our nation's schools. Few matters are as neglected.... A central thesis of this book is that there is a natural connection between good teachers and good schools and that this connection has been largely ignored. . . . *It is folly to assume that schools can be exemplary when their stewards are ill-prepared.*

b. Arnold Arons (2000) [commenting on the ground-breaking work of the forgotten pioneer Louis Paul Benezet (1935/36) [see also Mahajan & Hake (2000)]:

. . . *I have looked at the Benezet papers, and I find the story congenial.* The importance of cultivating the use of English cannot be exaggerated. I have been pointing to this myself since the '50's, and am delighted to find such explicit agreement. I can only point out that my own advocacy has had no more lasting effect than Benezet's. [You will find some of my views of this aspect in (Arons 1959)] . . . Benezet taught excellent arithmetic from the very beginning just as it should be taught. What he removed was useless drill on memorized algorithms that had no connection to experience and verbal interpretation. . . . This, of course, brings us back to the same old problem: *Whence do we get the teachers with the background, understanding, and security to implement such instruction? They will certainly not emerge from the present production mills.*

c. TheodoreSizer (2002):

Finally, in reality *it all comes down to the teachers.* However brilliant the "curriculum frameworks" and however scholarly the textbooks, what the teachers do with them is most of the game. Ravitch (2002) surely would agree.

d. Larry Cuban (2003):

. . . I know from both experience and research that the teacher is at the heart of student learning and school improvement by virtue of being the classroom authority and gatekeeper for change. *Thus the preparation, induction, and career development of teachers remain the Archimedean lever for both short- and long-term improvement of public schools.*

e. Linda Darling-Hammond (2007):

Research indicates that expert teachers are the most important—and the most inequitably distributed—school resource. In the United States, however, schools serving more than 1 million of our highest-need students are staffed by a parade of underprepared and inexperienced teachers who know little about effective instruction, and even less about teaching English-language learners and students with disabilities. Many of these teachers enter the classroom with little training and leave soon after, creating greater instability in their wake. Meanwhile, affluent students receive teachers who are typically better prepared than their predecessors, further widening the achievement gap.

If we are serious about leaving no child behind, we need to go beyond mandates to ensure that *all* (emphasis in original) students have well-qualified teachers. Effective action can be modeled after practices in medicine. Since 1944, the federal government has subsidized medical training to fill shortages and build teaching hospitals and training programs in high-need areas – a commitment that has contributed significantly to America’s world-renowned system of medical training and care.

Intelligent, targeted incentives can ensure that all students have access to teachers who are indeed highly qualified. *An aggressive national policy on teacher quality and supply, on the order of the post-World War II Marshall Plan, could be accomplished for less than 1 percent of the more than \$300 billion spent thus far in Iraq, and, in a matter of only a few years, would establish a world-class teaching force in all communities.*

In “Physics First: Opening Battle in the War on Science/Math Illiteracy?” [Hake (2002e)], I list seven steps that might be taken to increase the pool of effective science/math teachers, the most important being to motivate universities to discharge their obligation to adequately educate prospective K-12 teachers.

And, considering the dire shortage of effective teachers generally, Don Langenberg (2000) has written [my *italics*]:

The recognition that inadequate teacher performance is a major cause of low student performance has focused school reform initiatives across the nation on teacher preparation and certification, and has drawn into those initiatives the colleges and universities that educate and train teachers, often under the aegis of “K- 16” partnerships. . . . Laudable though such efforts may be, however, it is becoming increasingly clear that they are grossly insufficient. There is a Sword of Damocles hanging over the teacher improvement movement. *It is the demand in this decade for about twice as many new teachers as our schools of education seem likely to graduate, conducting business as usual.* We must have better teachers, and we must also have many more teachers. Just how we are going to accomplish both, and soon, seems to be a mystery to just about everybody, including the current presidential candidates. Where is Superman when we need him? It seems to me that, while we’re awaiting a visitor from Krypton, we ought to consider—and take—actions appropriate to the nature and the scale of the problem before us. *Its nature and scale are such that many of the necessary actions will have to be of the unthinkable and impossible variety.* So be it, for we have a very real national crisis on our hands, and we must do what Americans always do in such circumstances, i.e., rise to the challenge! Here’s what I think we need to do:

1. *Focus evaluation and compensation of teachers primarily on their impacts on the performance of their students.* A teacher's academic background or longevity is far less important than his or her effect on the progress of students for whom the teacher is responsible. . . .
2. *Pay good teachers what they are worth.* It's time we recognized that those responsible for developing our children's brains are worth at least as much as those who develop our electronic brains. I'd guesstimate that, on average, teacher's salaries ought to be about 50% higher than they are now. . . .
3. *Cease employing inadequate teachers.* Teachers who are unqualified, underqualified, or simply incompetent have no place in our classrooms and we ought to stop employing them. . . .
4. *Treat teachers – and expect them to behave – like members of a learned profession.* This means profound changes in the behavior of our schools, who employ teachers, *of our universities, who prepare teachers*, and of teachers themselves.

If one is willing to entertain the provocative premise that the ponderous inertia of the *entire* U.S. educational system is due in part to the inertia of the higher-education component in failing to improve K-12 education, then it may be worthwhile to consider the eleven reasons enumerated below for higher education's resistance to change.

9. Eleven Barriers to Change in Higher Education

The prima facie affront: Whereas I have spent a significant fraction of my professional life perfecting my lectures and otherwise investing conscientiously in the status quo, therefore to suggest an alternative is, by definition, to attack me.

Robert Halfman, Margaret MacVicar, W.T. Martin,
Edwin Taylor, and Jerrold Zacharias (1977)

Shiela Tobias (2000) in her "Guest Comment: From Innovation to Change: Forging a Physics Education Agenda for the 21st Century," listed the following four barriers to educational change:

1. In-class and standardized tests (MCAT, SAT, GRE) drive the curriculum in a traditional direction.
2. Effectiveness of teaching has little effect on promotion/tenure decisions or on national departmental rankings.
3. Clients for the sciences are not cultivated among those who do not wish to obtain PhD's.
4. Class sizes are too large.

In “Lessons from the Physics Education Reform Effort” [Hake (2002a)], I indicated two more barriers in Lesson #12: *Various institutional and political factors, including the culture of research universities, slow educational reform:*

5. Universities fail to think of education in terms of student learning rather than the delivery of instruction [Barr and Tagg (1995)] - for a discussion see Hake (2005a).
6. Universities fail to effectively consider crucial multidisciplinary societal problems such as education. In the words of Karl Pister (1996), former Chancellor of the University of California - Santa Cruz:
. . . . we need to encourage innovative ways of looking at problems, moving away from the increasing specialization of academia to develop new interdisciplinary fields that can address complex real-world problems from new perspectives.

Still other impediments to the reform of undergraduate education are:

7. The gross misuse of Student Evaluations of Teaching (SET's) *as gauges of student higher-order learning*, as discussed in “Re: Problems with Student Evaluations: Is Assessment the Remedy?” [Hake (2002d)]; “SET's Are Not Valid Gauges of Teaching Performance” [Hake (2006j)]; “Re: Adjunct Faculty: Improving Results” [Hake (2006k)]; “Re: Research into student evaluations” [Hake (2006l)]; and “The High Risks of Improving Teaching” [Rhem (2006)].
8. The provincialism of educational researchers in various disciplines, as discussed in Section II, “The Insularity of Educational Research” of an article “Design-Based Research: A Primer for Physics Education Researchers” [Hake (2004c)].
9. The failure of professors to “shut up and listen to students.” The late Arnold Arons (1974) wrote [pertaining to physics education, but doubtless relevant to most other subjects; my *italics*]:
If a teacher disciplines himself to conduct such Socratic Dialogues he begins to see the difficulties, misconceptions, verbal pitfalls, and conceptual problems encountered by the learner. . . . In my own case, *I have acquired in this empirical fashion virtually everything I know about the learning difficulties encountered by students.* I have never been bright enough or clairvoyant enough to see the more subtle and significant learning hurdles *a priori*. . . . I am deeply convinced that a statistically significant improvement would occur if more of us learned to listen to our students . . . *By listening to what they say in answer to carefully phrased, leading questions, we can begin to understand what does and does not happen in their minds, anticipate the hurdles they encounter, and provide the kind of help needed to master a concept or line of reasoning without simply ‘telling them the answer.’* . . . Nothing is more ineffectually arrogant than the widely found teacher attitude that “all you have to do is say it my way, and no one within hearing can fail to understand it.” . . . *Were more of us willing to relearn our physics by the dialogue and listening process I have described, we would see a discontinuous upward shift in the quality of physics teaching.* I am satisfied that this is fully within the competence of our colleagues; the question is one of humility and desire.

For a discussion of Arons's insights on education and Socratic pedagogy see “The Arons Advocated Method” [Hake (2004d)].

10. The schizophrenia of academic scientists (counterpart to the pre/post paranoia of PEP's) has been portrayed by physicist Robert Hilborn (2005) in "Academic Schizophrenia, the Scholarship of Teaching, and STEM Leadership" [my *italics*]:

Scientists in academe are schizophrenic. That diagnosis is not a formal psychiatric assessment of their mental states, but a factual statement about academic scientists' behavior. *When they deal with scientific research and when they work on educational projects, faculty members display totally different personalities.* . . . This schizophrenia is not entirely irrational. *Our scholarly communities have acculturated us to behave differently when doing research and when teaching.* For example, I know from my experiences in scientific research that if I wrote a research proposal to the National Science Foundation without references to what had already been done in the field and if I indicated that I would not bother to evaluate the validity of my research results nor publish them in journals, I would be dismissed as a lunatic or a fraud (or perhaps both). Not wishing to be thrust into either category, I do my homework and acknowledge previous research work and write explicitly about what my work will add to the body of scientific knowledge and how I will communicate my results to the scientific community. But our community standards and expectations for teaching have evolved differently: *my teaching might be viewed as perfectly fine, and indeed exemplary (as long as the students are happy), without any of the scholarly attributes that are viewed as a necessary part of research.* . . . I argue that academic freedom should not include the freedom to "screw up" by using teaching practices that have been shown to be ineffective. For example, *physics education research has shown with studies of thousands of students in many different kinds of institutions [Hake (1998a)], that straight lecturing does little to build conceptual understanding of basic physics ideas.*

The last listed barrier to reform is a theme of this essay:

11. The pre/post paranoic aversion to the *formative* assessment of change in students' conceptual understanding over the period of a course.

Research on innovation diffusion (or lack thereof) of the type pioneered by Everett Rogers (2003) [see e.g., <http://en.wikipedia.org/wiki/Diffusion_of_innovations>], and Clayton Christensen (2006)) [see e.g., <http://en.wikipedia.org/wiki/Clayton_Christensen>] might yield more information on the factors responsible for the pathologically slow diffusion of innovation in higher education. Some physics education researchers are now pursuing such studies. For example, Henderson & Dancy (2006a,b) are interviewing faculty to determine their attitudes toward innovations in physics education and the extent to which faculty have incorporated such innovations into their own classroom practice.

10. Formative vs Summative Assessment

... standards can be raised only by changes that are put into direct effect by teachers and pupils in classrooms. There is a body of firm evidence that formative assessment is an essential component of classroom work and that its development can raise standards of achievement. We know of no other way of raising standards for which such a strong prima facie case can be made. Our plea is that national and state policy makers will grasp this opportunity and take the lead in this direction.

Paul Black and Dylan Wiliam (1998)

It should be emphasized that such low-stakes *formative* pre/post testing as discussed above is the polar opposite of the high-stakes *summative* testing mandated by the U.S. Department of Education's "No Child Left Behind Act" for K-12 [USDE (2005)], now contemplated for higher education [USDE (2006)]. As the NCLB experience shows, such testing often falls victim to *Campbell's Law* [Campbell (1976), Nichols & Berliner (2005, 2007)]:

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.

Why is the pre/post testing discussed above regarded as *formative*? Because both teachers' *action research* and education researchers' scientific research is carried out to improve classroom teaching and learning, NOT to rate instructors or students. Thus it's "formative" as defined by JCSEE (1994): "Formative evaluation is evaluation designed and used to improve an object, especially when it is still being developed." For a four-quadrant delineation of the formative/summative and public/private dimensions of assessment/evaluation see Hake (2003a).

11. Why Worry About Student Learning in Higher Education?

Our education in America at the undergraduate level must be focused on preparing students for roles in a competitive environment in which qualities that cannot be off-shored are developed in our students. . . . A liberal education, combined with deep knowledge and understanding of some particular field, is the best preparation for the world-scale competition we all face in the 21st century, and the best way to educate our citizens to play active roles in a democratic society.

David Oxtoby - President, Pomona College; from the Keynote Address at the 2006 *Assessing Student Achievement Conference* as reported in PKAL (2006b).

Although international competitiveness is often cited by educational leaders, politicians, and business executives, more crucial in my view is the need to overcome the monumental problems now threatening life on planet Earth. In "The General Population's Ignorance of Science Related Societal Issues: A Challenge for the University" [Hake (2000b)] I list a few (14) such problems and cite the imperative to (a) educate more effective science majors and science-trained professionals, and (b) raise the appallingly low level of science literacy among the general population by properly educating prospective K-12 teachers.

Human history becomes more and more a race between education and catastrophe.

H. G. Wells (1920)

12. Summary and Conclusions

The success of formative pre/post testing in motivating and guiding the enhancement of student learning in some sectors of higher education appears to contradict the advice of Cronbach & Furby (1970) that “gain scores are rarely useful, no matter how they may be adjusted or refined.” On the contrary, that “Interactive Engagement” methods can markedly increase higher-order learning in conceptually difficult areas now seems clear, thanks to 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. Furthermore, 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 required to develop multiple-choice tests of higher-level learning of those concepts, so as to gauge the need for and effects of non-traditional pedagogy, and (c) develop “Interactive Engagement” methods suitable to their disciplines.

13. Acknowledgments

I have benefited from recent 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, Larry Woolf, 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 Hake (2005f)]. Most importantly, I thank Michael Scriven for suggesting that I write this book chapter and for his many valuable suggestions.

Reference & Footnotes [Tiny URL's courtesy <<http://tinyurl.com/create.php>>; all URL's were accessed on 9-12 January 2007.

Albacete, P.L., & K. VanLehn. 2000. "Evaluation the effectiveness of a cognitive tutor for fundamental physics concepts," in L.R. Gleitman & A.K. Joshi, eds., *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society.*, pp. 25-30). Erlbaum; online at <<http://www.pitt.edu/~vanlehn/distrib/Papers/last-cogsci-albacete.pdf>> (48 kB).

Anderson, J.R. 2004. *Cognitive Psychology and Its Implications*. Worth Publishers, 6th edition. For a physicist's view of cognitive science see Redish (1994).

Arons, A.B. 1959. "Structure, methods, and objectives of the required freshman calculus-physics course at Amherst College. *Am. J. Phys.* **27**(9): 658–666; online to subscribers at <<http://scitation.aip.org/dbt/dbt.jsp?KEY=AJPIAS&Volume=27&Issue=9>>.

Arons, A.B. 1973. "Toward wider public understanding of science," *Am. J. Phys.* **41**(6): 769-782; an abstract is online at <<http://tinyurl.com/yxxcud>>.

Arons, A.B. 1974. "Toward wider public understanding of science: Addendum," *Am. J. Phys.* **42**(2): 157-158. An addendum to Arons (1973). Online to subscribers at <<http://scitation.aip.org/dbt/dbt.jsp?KEY=AJPIAS&Volume=42&Issue=2>>.

Arons, A.B. 1990. *A Guide to Introductory Physics Teaching*. Wiley; reprinted with minor updates in Arons (1997). For an excellent review see Knight (1995).

Arons, A.B. 1993. "Guiding Insight and Inquiry in the Introductory Physics Laboratory," *Phys. Teach.* **31**(5), 278-282 (1993); online to subscribers at <<http://scitation.aip.org/dbt/dbt.jsp?KEY=PHTEAH&Volume=31&Issue=5>>.

Arons, A.B. 1994. *Homework and Test Questions for Introductory Physics Teaching*. Wiley.

Arons, A.B. 1997. *Teaching Introductory Physics*. Wiley. Contains a slightly updated version of Arons (1990), plus *Homework and Test Questions for Introductory Physics Teaching* [Arons (1994)], plus a new monograph *Introduction to Classical Conservation Laws*.

Arons, A.B. 2000. Private communication to R.R. Hake of 30 June concerning the Louis Paul Benezet's (1935/36) pioneering replacement of rote memorization with interactive engagement in a sector of the Manchester, New Hampshire public school system.

Bao, L. 2006. "Theoretical comparisons of average normalized gain calculations," *Am. J. Phys.* **74**(10): 917-922; abstract online at <<http://tinyurl.com/yx3h9q>>.

Baker, F. 2001. *The Basics of Item Response Theory*, online at <<http://edres.org/irt/baker/>>, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD; edited and updated by Carol Boston of ERIC. Baker's book was first published in 1985. For many other online IRT resources see Rudner (2001).

Bardar, E.M.W., E.E. Prather, K. Brecher, & T.F. Slater. 2006. "The Need for a Light and Spectroscopy Concept Inventory for Assessing Innovations in Introductory Astronomy Survey Courses," *Astronomy Education Review* 4(2): 20-27; online at <<http://aer.noao.edu/cgi-bin/article.pl?id=160>>. They write [my *italics* and inserts at "...[insert]. ..."]:

Cited more than any other work in physics education is the far-reaching work by ... [Hestenes et al. (1992)]. ... which focuses on the assessment of student conceptions of Newton's laws of motion. The Force Concept Inventory (FCI) was developed to serve as a diagnostic tool capable of assessing change in students' understanding of concepts related to force and motion, as a result of introductory university physics instruction. The FCI is a 29-item ... [now 30]. ... multiple-choice survey typically administered at the beginning and end of a course. The FCI requires no calculators and is written in the natural language of students, such that it is reliable and valid as a pretest when used prior to instruction. What is unique about this survey is that the items initially appear trivially easy to most faculty. However, research on how students actually respond to FCI questions reveals that the multiple-choice distracters are so enticing and attractive that students consistently and confidently select answers that are wholly incompatible with accurate scientific thinking. Even more revealing is the finding that traditional lecture-based instruction seems to have little, if any, impact on precourse/postcourse student gain scores. On the other hand, FCI results from classes that implemented innovative learner-centered instructional strategies informed by physics education research on student learning showed dramatic precourse/postcourse gains. *In this way, the FCI has been able to successfully document that student achievement in traditional lecture-based physics courses consistently falls below the achievement of students who are taught using research-based instructional methods* (Hake 1998).

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>>.

Begley, S. 2004. "To Improve Education, We Need Clinical Trials To Show What Works," *Wall Street Journal*, 17 December, page B1; online for educators at <<http://tinyurl.com/ycgq2f>> (scroll to the Appendix).

Benezet, L.P. 1935/36. "The teaching of arithmetic I, II, III: The story of an experiment," *Journal of the National Education Association* 24(8), 241-244 (1935); 24(9), 301-303 (1935); 25(1), 7-8 (1936). The articles were: (a) reprinted in the *Humanistic Mathematics Newsletter* #6: 2-14 (May 1991); (b) placed on the web along with other Benezetia at the Benezet Centre; online at <<http://www.inference.phy.cam.ac.uk/sanjoy/benezet/>>. See also Mahajan & Hake (2000).

Black, P. & D. Wiliam. 1998. "Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan* **80**(2): 139-144, 146-148; online at <http://www.pdkintl.org/kappan/kbla9810.htm>.

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; online at http://www.ncsu.edu/per/Articles/Varenna_SCALEUP_Paper.pdf (1MB).

BHEF. 2001. Business - Higher Education Forum (a partnership of the American Council on Education and the National Alliance of Business), Winter, "Sharing Responsibility: How Leaders in Business and Higher Education Can Improve America's Schools" online at http://www.bhef.com/includes/pdf/sharing_responsibility.pdf (244 kB), see page 23.

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; online "soon" (but don't hold your breath) at <http://www.aera.net/publications/?id=331>. An ERIC abstract is online at <http://tinyurl.com/yxmyze>. 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 - information including the preface and Chapter 1 is online at <http://press.princeton.edu/titles/8125.html>.

Bok, D. 2005b. "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).

Bond, L. 2005. "Carnegie Perspectives: A different way to think about teaching and learning - Who has the lowest prices," online at <http://www.carnegiefoundation.org/perspectives/sub.asp?key=245&subkey=569>; see also the commentary at <http://www.carnegiefoundation.org/perspectives/sub.asp?key=244&subkey=35>, including those by Hake at <http://www.carnegiefoundation.org/perspectives/sub.asp?key=244&subkey=1814>, and <http://www.carnegiefoundation.org/perspectives/sub.asp?key=244&subkey=1810>.

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>>. The quote is from page 106 of the earlier 1999 edition.

Breslow, L. 1999. "New Research Points to the Importance of Using Active Learning in the Classroom," MIT Teaching & Learning Laboratory, Vol. XIII, No. 1, September/October; online at <<http://web.mit.edu/tll/tll-library/teach-talk/new-research.html>>.

Breslow, L. 2000. "Active Learning, Part II, Suggestions for Using Active Learning Techniques in the Classroom," Vol. XII, No. 3, January/February 2000 <<http://web.mit.edu/tll/tll-library/teach-talk/active-learning-2.html>>.

Bridgman, P.W. 1927. *The Logic of Modern Physics*. Now available in a 1960 edition from Macmillan.

Brookes, D. 2006. "Re: Best Practices in Science Education," PhysLrnR post of 29 Sep 2006 06:07:24-0700, online at <<http://tinyurl.com/ygju5d>>.

Bruer, J.T. 1997. "Education and the Brain: A Bridge Too Far," *Educational Researcher* **26**(8): 4-16, online "soon" (but don't hold your breath) at <<http://www.aera.net/publications/?id=331>>.

Bruer, J.T. 2006. "Points of View: On the Implications of Neuroscience Research for Science Teaching and Learning: Are There Any? A Skeptical Theme and Variations: The Primacy of Psychology in the Science of Learning," *CBE-Life Sciences Education* **5**: 104-110, Summer, online at <<http://www.lifescied.org/cgi/reprint/5/2/104>>. Bruer writes [my insert at "...[insert]. ..."]:

I am notorious for my skepticism about what neuroscience can currently offer to education. My skepticism derives from several concerns, but a common theme runs through all of them: attempts to link neuroscience with education pay insufficient attention to psychology. . . [but, in my opinion (Hake, 2005d), psychology pays insufficient attention to classroom teaching]. . . .In what follows, I will present four variations on this theme. First, for those who are committed to developing a science-based pedagogy and solving existing instructional problems, cognitive psychology offers a mother-lode of still largely untapped knowledge. Second, attempts to link developmental neurobiology to brain development and education ignore, or are inconsistent with, what cognitive psychology tells us about teaching and learning. Third, cognitive neuroscience is the brain-based discipline that is most likely to generate educationally relevant insights, but cognitive neuroscience presupposes cognitive psychology and, to date, rarely constrains existing cognitive models. And fourth, the methods of cellular and molecular neuroscience are powerful, but it is not always clear that the concepts of learning and memory used by neuroscientists are the same as those used by psychologists, let alone by classroom teachers.

Buck, J.R. & K.E. Wage. 2005. "Active and Cooperative Learning in Signal Processing Courses," *IEEE Signal Processing Magazine* **22**(2): 76-81, March; online as a 573 kB pdf at <<http://tinyurl.com/ymk5x8>> / "Publications," where "/" means "click on."

Campbell, D. T. 1976. "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). For a memorial to the late Donald Campbell see <http://pespmc1.vub.ac.be/CAMPBEL.html>: "He was the founder of the domain of 'evolutionary epistemology' <http://pespmc1.vub.ac.be/EVOLEPIST.html> (a label he created), in which he generalized Popper's falsificationist philosophy of science to knowledge processes at all biological, psychological and social levels."

Christensen, M. 2006. *The Innovator's Dilemma*. Harper Collins - information at <http://tinyurl.com/yx9k9h>. First published in 1997 by the Harvard Business School Press.

Clemens, H. 1989. "Is There a Role for Mathematicians in Math Education?" *Notices of the American Mathematical Society* **36**(5): 542-544.

Cohen, P., J. Cohen, L.S. Aiken, & S.G. West. 1999. "The Problem of Units and the Circumstance for POMP," *Multivariate Behavioral Research* **34**(3): 315-346; online as a 120kB pdf at <http://tinyurl.com/pcxt3>. The abstract reads [my *italics*]:

Many areas of the behavioral sciences have few measures that are accepted as the standard for the operationalization of a construct. One consequence is that there is hardly ever an articulated and understood framework for the units of the measures that are employed. Without meaningful measurement units, theoretical formulations are limited to statements of the direction of an effect or association, or to effects expressed in standardized units. Thus the long term scientific goal of generation of laws expressing the relationships among variables in scale units is greatly hindered. This article reviews alternative methods of scoring a scale. Two recent journal volumes are surveyed with regard to current scoring practices. Alternative methods of scoring are evaluated against seven articulated criteria representing the information conveyed by each in an illustrative example. *Converting scores to the percent of maximum possible score (POMP) is shown to provide useful additional information in many cases.*

Collins, L.M. & J.L. Horn, eds. 1991. *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*, report of a 1989 conference. American Psychological Association. See also Collins & Sayer (2001).

Collins, L.M. & A.G. Sayer, eds. 2001. *New Methods for the Analysis of Change*. American Psychological Association. APA information at <http://www.apa.org/books/4318991.html>, including the Table of Contents <http://www.apa.org/books/4318991t.html>.

Cook, T.D. 2006. "Describing what is Special about the Role of Experiments in Contemporary Educational Research?: Putting the 'Gold Standard' Rhetoric into Perspective," *Journal of MultiDisciplinary Evaluation*, Number 6, November, online at http://evaluaton.wmich.edu/jmde/JMDE_Num006.html. Note that "experiments" is standard psychometric jargon for "randomized control trials" (RCT's), but Cook uses the term "randomized experiment," presumably to distinguish "randomized experiment" from the less venerated "quasi-experiment." See the response by Scriven (2006).

Cromer, A. 1997. *Connected Knowledge*. Oxford University Press, information at <<http://www.oup.com/uk/catalogue/?ci=9780195096361>>.

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

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/sbys4>>.

CSTA. 2006a. California Science Teachers Association <<http://www.cascience.org/>>, "Current News: California state board backs hands-on science," online at <<http://www.cascience.org/currentnews.html-handson>>.

CSTA. 2006b. California Science Teachers Association, "Instructional Materials: K-8 Instructional Materials Adoption: State board adopts revised materials criteria, declines to limit hands-on materials," online at <<http://www.cascience.org/IMCriteria.html>>.

Cuban, L. 2003. *Why Is It So Hard To Get Good Schools?* Teachers College Press - information at <<http://store.tcpres.com/0807742945.shtml>>.

Daedalus. 1998. Special issue on "Education yesterday, education tomorrow," volume **127**(4). Focuses on K-12 education. The online description, formerly at <<http://daedalus.amacad.org/inprint.html>> has rotted. The American Academy of Arts and Sciences <<http://www.amacad.org/>>, former publishers of *Daedalus*, has seen fit to give contents of issues at <http://www.amacad.org/publications/back_issues.aspx> only back to Fall 2001. The current publishers, MIT press, display an even more limited list of issues back to Fall 2003 at <<http://www.mitpressjournals.org/loi/daed>>. See also *Daedalus* (2002, 2004)).

Daedalus. 2002. Special issue on Education, Summer, cover online at <<http://www.amacad.org/publications/summer2002/Su2002coverweb.pdf>> (828 kB). As the previous 1998 education issue, this issue focuses on K-12. Of the 15 articles, 5 are available online:

- a. Ravitch, D. "Education after the culture wars",
<<http://www.amacad.org/publications/summer2002/ravitch.pdf>> (704 kB);
- b.Sizer, T.R. "A better way,"
<<http://www.amacad.org/publications/summer2002/sizer.pdf>> (444 kB);
- c. Hirsch, E. D. "Not to worry?"
<<http://www.amacad.org/publications/summer2002/hirsch.pdf>> (448 kB);;
- d. Delbanco, A. "It all comes down to the teachers,"
<<http://www.amacad.org/publications/summer2002/delbanco.pdf>> (424 kB);
- e. Bloom, D.E. & J.E. Cohen, "Education for all: an unfinished revolution,"
<<http://www.amacad.org/publications/summer2002/bloomcohen.pdf>> (648 kB).

Daedalus. 2004. Special issue on Learning, Winter issue, cover is online at <http://www.amacad.org/publications/winter2004/cover_winter_2004.pdf> (53 kB) Of the 8 articles, 3 are available online:

- a. Gopnik, A. "Finding our inner scientist," <<http://www.amacad.org/publications/winter2004/gopnik.pdf>> (372 kB);
- b. Povinelli, D.J. "Behind the ape's appearance: escaping anthropocentrism in the study of other minds," <<http://www.amacad.org/publications/winter2004/povinelli.pdf>> (424 kB);
- c. Churchland, P.S. "How do neurons know?" <<http://www.amacad.org/publications/winter2004/churchland.pdf>> (384 kB).

Dancy, M.H. & R.J. Beichner. 2002. "But Are They Leaning? Getting Started in Classroom Evaluation," *Cell Biology Education* 1(3): 87-94; online at <<http://www.lifescied.org/cgi/content/full/1/3/87>> .

Darling-Hammond, L. 2007. "A Marshall Plan for Teaching: What It Will Really Take to Leave No Child Behind" *Education Week* 26 (issue 18): pp. 28,48; online at <<http://www.edweek.org/ew/toc/2007/01/10/index.html>>.

DeHaan, R.L. 2005. "The Impending Revolution in Undergraduate Science Education," *Journal of Science Education and Technology* 14(2): 253-269; abstract online at <<http://tinyurl.com/ymwwe3>>.

Diamond M. C. 1996. "The Brain. Use It or Lose It." *Mind Shift Connection* 1: 1; online at <http://www.newhorizons.org/neuro/diamond_use.htm>.

Dori, Y.J. & J. Belcher. 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>>. See also Duderstadt (2001).

Duderstadt, J. J. 2001. "Science policy for the next 50 years: from guns to pills to brains," in *Proceedings of the AAAS Annual Meeting*, San Francisco, February, 2001. Available online at: <http://milproj.umm.umich.edu/publications/aaas_text_2>.

Elliott, C. 2003. "Using a Personal Response System in Economics Teaching," *International Review of Economics Education* 1(1): 80-86, online at <<http://www.economics.ltsn.ac.uk/iree/i1/elliott.htm> - fn0> (36 kB).

Evans, D.L., G.L. Gray, S. Krause, J. Martin, C. Midkiff, B.M. Notaros, M. Pavelich, D. Rancour, T. Reed-Rhoads, P. Steif, R. Streveler, & K. Wage. 2003. "Progress on Concept Inventory Assessment Tools," *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference*, pp. T4G1-T4G8; online at http://ece.gmu.edu/~kwage/research/ssci/papers/evans_etal_fie_nov03.pdf (60 kB).

Epstein, J. 2006. "The Calculus Concept Inventory," abstract online at *2006 Conference on Research in Undergraduate Mathematics Education*, online at <http://mathed.asu.edu/CRUME2006/Abstracts.html>, scroll down about one third of the way to the bottom.

Feldman, A. & J. Minstrell. 2000. "Action research as a research methodology for the study of the teaching and learning of science," in E. Kelly & R. Lesh, eds., *Handbook of research design in mathematics and science education*. Lawrence Erlbaum; online at <http://www-unix.oit.umass.edu/~afeldman/ActionResearchPapers/FeldmanMinstrell2000.PDF> (20 kB).

FLAG. 2006. "Field-tested Learning Assessment Guide," online at <http://www.flaguide.org/>: "... offers broadly applicable, self-contained modular classroom assessment techniques (CAT's) and discipline-specific tools for STEM [Science, Technology, Engineering, and Mathematics] instructors interested in new approaches to evaluating student learning, attitudes, and performance. Each has been developed, tested, and refined in real colleges and university classrooms." Assessment tools for physics and astronomy (and other disciplines) are at <http://www.flaguide.org/tools/tools.php>.

Froyd, J., J. Layne, & K. Watson. 2006. "Issues Regarding Change in Engineering Education," *2006 Frontiers in Education Conference (FIE 2006)*, online at <http://fie.engrng.pitt.edu/fie2006/papers/1645.pdf> (208 kB).

Galley, M. 2004a. "Calif. Mulls Limiting Hands-On Science Lessons," *Education Week* **23**(issue 24): 5, 25 February, online at <http://www.edweek.org/ew/articles/2004/02/25/24science.h23.html?print=1>.

Galley, M. 2004b. "News in Brief: California State Board Backs Hands-On Science," *Education Week* **23**(issue 28): 22, 24 March, online at <http://www.edweek.org/ew/articles/2004/03/24/28caps.h23.html>.

Gery, F.W. 1972. "Does mathematics matter?" in A. Welch, ed., *Research papers in economic education*, Joint Council on Economic Education. pp. 142-157.

Giere, R.N. 1997. *Understanding Scientific Reasoning*. Holt, Rinehart, and Winston.

Goodlad, J.I. 1990. *Teachers For Our Nation's Schools* (Jossey-Bass).

Gottfried, K. & K.G. Wilson. 1997. "Science as a cultural construct," *Nature* **386**: 545. They attack the "strong program" of the Edinburg school of social constructivists. An abstract with references is online at <<http://www.nature.com/nature/journal/v386/n6625/abs/386545a0.html>>. The abstract is: "Scientific knowledge is a communal belief system with a dubious grip on reality, according to a widely quoted school of sociologists. But they ignore crucial evidence that contradicts this allegation."

Hake, R.R. 1987. "Promoting Student Crossover to the Newtonian World," *Am. J. Physics* **55**(10): 878-884 (1987); online at <<http://www.physics.indiana.edu/~hake/PromotingCrossover.pdf>> (788 kB).

Hake, R.R. 1992. "Socratic pedagogy in the introductory physics lab." *Phys. Teach.* **30**: 546-552; updated version (4/27/98) 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**(1): 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. 2000a. "Towards Paradigm Peace in Physics Education Research," presented at the annual meeting of the American Educational Research Association, New Orleans, 24-28 April; online at <http://www.physics.indiana.edu/~sdi/AERA-Hake_11.pdf> (168 KB). Also at that location is a pdf version <<http://www.physics.indiana.edu/~hake/ParadigmSlides.pdf>> (244 kB) of the PowerPoint slides shown at the meeting.

Hake, R.R. 2000b. "The General Population's Ignorance of Science Related Societal Issues: A Challenge for the University," *AAPT Announcer* **30**(2): 105; online at <<http://www.physics.indiana.edu/~hake/GuelphSocietyG.pdf>> (2.1MB). Based on an earlier libretto with the leitmotiv: "The road to U.S. science literacy begins with effective university science courses for pre-college teachers." The opera dramatizes the fact that the failure of universities *throughout the universe* to properly educate pre-college teachers is responsible for our failure to observe any signs of either terrestrial or extraterrestrial intelligence.

Hake, R.R. 2002a. "Lessons from the Physics Education Reform Effort," *Ecology and Society* **2**: 28; online at <<http://www.ecologyandsociety.org/vol5/iss2/art28/>>. *Ecology and Society* is only online so the URL *must* be specified.

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

Hake, R.R. 2002c. "Assessment of Student Learning in Introductory Science Courses," 2002 PKAL Roundtable on the Future: Assessment in the Service of Student Learning, online at <<http://www.physics.indiana.edu/~hake/ASLIS.Hake.060102f.pdf>> (192 kB).

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

Hake, R.R. 2002e. "Physics First: Opening Battle in the War on Science/Math Illiteracy?" Submitted to the *American Journal of Physics* on 27 June 2002; online at <<http://www.physics.indiana.edu/~hake/PhysFirst-AJP-6.pdf>> (220 kB).

Hake, R.R. 2002f. "Re: Socratic Method," online at <<http://lists.nau.edu/cgi-bin/wa?A2=ind0211&L=phys-l&P=R9157>>. Post of 14 Nov 2002 14:32:54-0800 to Phys-L, PhysLrnR, and AP-Physics.

Hake, R.R. 2002g. "Re: Socratic Method," online at <<http://lists.nau.edu/cgi-bin/wa?A2=ind0212&L=phys-l&P=R15999>>. Post of 13 Dec 2002 13:33:26-0800 to Phys-L, PhysLrnR, and AP-Physics.

Hake, R.R. 2002h. "Re: Socratic Method," online at <<http://lists.nau.edu/cgi-bin/wa?A2=ind0212&L=phys-l&P=R19258>>. Post of 18 Dec 2002 to Phys-L, PhysLrnR, and AP-Physics.

Hake, R.R. 2002i. "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, Idaho; August 2002; online at <<http://www.physics.indiana.edu/~hake/PERC2002h-Hake.pdf>> (220 KB).

Hake, R.R. 2002j. "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; also online at <<http://www.physics.indiana.edu/~hake/Hake-SriLanka-SDIb.pdf>> (44 KB).

Hake, R.R. 2003a. "Re: A taxonomy," POD posts of 9 & 12 Jul 2003, online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0307&L=pod&P=R4226&I=-3>> (a diagram is shown), and <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0307&L=pod&P=R5361&I=-3>>. In this four-quadrant taxonomy:

- (a) no distinction is made between *assessment* and *evaluation*;
- (b) the X-axis represents a continuum from pure *formative* to pure *summative* assessment of either teaching or learning, while the Y-axis represents a continuum from complete *privacy* to complete *public* disclosure of results;
- (c) "Scientific Research" [see e.g. Shavelson & Towne (2003)] is in the first and second quadrants, on the *public* side of the Y-axis and ranging from pure *formative* to pure *summative* on the X-axis;
- (d) "Action Research" [see e.g. Feldman & Minstrell (2000) and Bransford et al. (2000)] is in the third quadrant, on the *private* side of the Y-axis and the *formative* side of the X axis;
- (e) "Institutional/Administrative Research" is usually in the fourth quadrant, on the *private* side of the Y-axis, and the *summative* side of the X-axis (although it could approach the formative for those who study and attempt to improve institutional/administrative practice.

Hake, R.R. 2003b. "Spare Me That Passive-Student Lecture," POD post of 1 Oct 2003 17:07:48 -0700 to; online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0310&L=pod&O=D&P=947>>. I list seven deficiencies in Powell's otherwise interesting and generally accurate review and respond to three provocative statements by Powell interviewees.

Hake, R.R. 2003c. "NRC's CUSE: Oblivious of the Advantage of Pre/Post Testing With High Quality Standardized Tests?" POD post of 25 Jul 2003 13:07:23-0700; online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0307&L=pod&O=D&P=17145>>.

Hake, R.R. 2003d. "NRC's CUSE: Stranded on Assessless Island?" POD post of 3 Aug 2003; online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0308&L=pod&F=&S=&P=391>>.

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>>. POD post of 19 Aug 2004 13:56:07-0700.

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. 2004c. "Design-Based Research: A Primer for Physics Education Researchers," submitted to the *Am. J. Phys.* on 10 June 2004; online at <<http://www.physics.indiana.edu/~hake/DBR-AJP-6.pdf>> (310kB); see Section II "The Insularity of Educational Research." See also Hake (2007c).

Hake, R.R. 2004d. "The Arons Advocated Method," submitted to the *Am. J. Phys.* on 24 April 2004; online at <<http://www.physics.indiana.edu/~hake/AronsAdvMeth-8.pdf>> (144 kB).

Hake, R.R. 2004e. "Direct Science Instruction Suffers a Setback in California - Or Does It?" *AAPT Announcer* **34**(2): 177; online at <<http://www.physics.indiana.edu/~hake/DirInstSetback-041104f.pdf>> (420 KB) [about 160 references and 180 hot-linked URL's]. A pdf version of the slides shown at the meeting is also available at <<http://www.physics.indiana.edu/~hake/AAPT-Slides.pdf>> (132 kB).

Hake, R.R. 2004f. "Re: Back to Basics vs. Hands-On Instruction," online at POD post of 24 Feb 2004 17:45:19-0800, online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0402&L=pod&O=D&P=18509>>.

Hake, R.R. 2004g. "Re: Back to Basics vs. Hands-On Instruction," POD post of 29 Feb 2004 17:57:25-0800, online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0402&L=pod&P=R17377>>.

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). This is a slightly edited version of an article that was (a) published in the *National Teaching and Learning Forum* **15**(1), December, online to subscribers at <<http://www.ntlf.com/FTPSite/issues/v15n1/physics.htm>>, and (b) disseminated by the *Tomorrow's Professor* list <<http://ctl.stanford.edu/Tomprof/postings.html>> as Msg. 698 on 14 Feb 2006. For an executive summary see Hake (2006a).

Hake, R.R. 2005b. "Why Aren't AERA Discussion Lists More Active?" POD post of 11 Jun 2005 11:44:58-0700, online at <<http://lists.asu.edu/cgi-bin/wa?A2=ind0506&L=aera-l&T=0&O=D&P=1246>>.

Hake, R.R. 2005c. "In Defense of Cross Posting," POD post of 24 Jul 2005 21:54:34-0700, online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0507&L=pod&P=R13219&I=-3>>.

Hake, R.R. 2005d. "Do Psychologists Research the Effectiveness of Their Courses? Hake Responds to Sternberg," POD post of 21 Jul 2005 22:55:31-0700, online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0507&L=pod&P=R11939&I=-3>>.

Hake, R.R. 2005e. "Will the No Child Left Behind Act Promote Direct Instruction of Science?" *Am. Phys. Soc.* **50**: 851 (2005); APS March Meeting, Los Angeles, CA. 21-25 March; online at <<http://www.physics.indiana.edu/~hake/WillNCLBPromoteDSI-3.pdf>> (256 kB).

Hake, R.R. 2005f. "The Insularity of Educational Research," online at <<http://tinyurl.com/aoe4p>>. Post of 21 May 2005 12:11:27-0700 to AERA-L and PhysLrnR. Contains a list of discussion-list archive URL's.

Hake, R.R. 2006a. "A Possible Model For Higher Education: The Physics Reform Effort (Author's Executive Summary)," *Spark* (American Astronomical Society Newsletter), June, online at <<http://www.aas.org/education/spark/SparkJune06.pdf>> (1.9MB). Scroll down about 4/5 of the way to the end of the newsletter.

Hake, R.R. 2006b. "Eleven Quotes in Honor of Inertia," POD post of 13 Jun 2006 15:01:14-0700, online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0606&L=pod&P=R7910&I=-3>>.

Hake, R.R. 2006c. "Re: Innovation and change vs. the status quo," POD post of 13 Jun 2006 21:19:23-0700, online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0606&L=pod&P=R8407&I=-3>>.

Hake, R.R. 2006d, e, f, g. "Ceiling Effects: Performance and Instrumental #1, #1 (addendum), #2, and #3," POD posts of 19, 20, and 21 July, online at, respectively, <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0607&L=pod&O=D&P=4617>>, <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0607&L=pod&P=R3602&I=-3>>, <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0607&L=pod&P=R3749&I=-3>>, <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0607&L=pod&P=R5240&I=-3>>.

Hake, R.R. 2006h. "Forward from Lee Sechrest: Re: Ceiling Effects: Performance and Instrumental," POD post of 21 Jul 2006 08:35:25-0700; online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0607&L=pod&P=R4075&I=-3>>.

Hake, R.R. 2006i. "Can Neuroscience Benefit Classroom Instruction?" POD post of 12 Oct 2006 16:26:32-0700; online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0610&L=pod&P=R6888&I=-3>>.

Hake, R.R. 2006j. "SET's Are Not Valid Gauges of Teaching Performance," AERA-L post of 20 Jun 2006 09:40:52-0700, online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0606&L=pod&O=A&P=12575>>.

Abstract: It is argued that *if universities value teaching that leads to student higher-level learning, then student evaluations of teaching (SET's) do NOT afford valid evidence of teaching performance*. Instead, institutions should consider 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.

Hake, R.R. 2006k. "Re: Adjunct Faculty: Improving Results." ASSESS post of 3 Jul 2006 17:00:06-0700; online at <<http://lsv.uky.edu/scripts/wa.exe?A2=ind0607&L=ASSESS&P=R2&I=-3>>.

Abstract: I respond in order to 20 points made by Dan Tompkins in ASSESS posts of 7-29 June 2006, titled "Re: Adjunct Faculty: Improving Results," "Re: SET's Are Not Valid Gauges of Teaching Performance," and "Re: What if students learn better in a course they don't like?" A subtitle might be "Is It Possible to Construct a 'Philosophy Concept Test' of students' higher-level learning?"

Hake, R.R. 2006l. "Re: Research into student evaluations," POD post of 7 Nov 2006 21:03:23-0800; online at <<http://listserv.nd.edu/cgi-bin/wa?A2=ind0611&L=pod&P=R3843&I=-3>>.

Hake, R.R. 2006m. "Possible Palliatives for the Paralyzing Pre/Post Paranoia that Plagues Some PEP's," *Journal of MultiDisciplinary Evaluation*, Number 6, November, online at <http://evaluation.wmich.edu/jmde/JMDE_Num006.html>. This even despite the admirable anti-alliteration advice at psychologist Donald Zimmerman's site <<http://mypage.direct.ca/z/zimmerma/>> to "Always assiduously and attentively avoid awful, awkward, atrocious, appalling, artificial, affected alliteration."

Hake, R.R. 2006n. "Best Practices in Science Education #2," PhysLrnR post of 30 Sep 2006 11:07:55-0700, online at <<http://tinyurl.com/y5akyf>>.

Hake, R.R. 2007a. "Should We Measure Change? Yes!" The present article. The latest version is online as ref. 43 at <<http://www.physics.indiana.edu/~hake>>. To appear as a chapter in Hake (2007b). I welcome comments and suggestions directed to <rrhake@earthlink.net>.

Hake, R.R. 2007b. *Evaluation of Teaching and Student Learning in Higher Education*, Monograph, American Evaluation Association <<http://www.eval.org/>>, in preparation.

Hake, R.R. 2007c. "Design-Based Research in Physics: A Review," to appear in Kelly & Lesh (2007).

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>>. The print version contains the *Mechanics Diagnostic* test, precursor to the *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>>.

Halfman, R., M.L.A. MacVicar, W.T. Martin, E.F. Taylor, & J.R. Zacharias. 1977. "Tactics for Change." MIT Occasional Paper No. 11. online at <<http://web.mit.edu/jbelcher/www/TacticsForChange/>>. Thanks to John Belcher for placing this gem on the web.

Handelsman, J., D. Ebert-May, R. Beichner, P. Bruns, A. Chang, R. DeHaan, J. Gentile, S. Lauffer, J. Stewart, S.M. Tilghman, & W.B. Wood. 2004. "Scientific Teaching," *Science* **304** (23): 521-522, April; online at <<http://www.plantpath.wisc.edu/fac/joh/scientificteaching.pdf>> (100 kB). See also the supporting material at <<http://scientificteaching.wisc.edu/resources.htm>> [URL's are specified for some, but (unfortunately) not all, online materials].

Haycock, K. 1999. "The role of higher education in the standards movement," in *1999 National Education Summit Briefing Book*, formerly online at the Alliance for Excellent Education (AEE) <<http://www.all4ed.org/index.html>>. I have requested AEE administrators to return this valuable resource to their website.

Hegedus, S.J. & J.J. Kaput. 2004. "An Introduction To The Profound Potential Of Connected Algebra Activities: Issues of Representation, Engagement and Pedagogy." *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 3: 129–136; online at <http://www.emis.de/proceedings/PME28/RR/RR261_Kaput.pdf> (408 kB).

Heller, K. J. 1999. "Introductory physics reform in the traditional format: an intellectual framework," *AIP Forum on Education Newsletter* (Summer): pages 7-9. The entire newsletter is available online at <<http://units.aps.org/units/fed/newsletters/summer99.pdf>> (8.1 MB); an illustrated form of Heller's AAPT talk is at <<http://groups.physics.umn.edu/physed/Talks/Heller1.pdf>> (844 kB).

Henderson, C. and M. Dancy. 2006a. "Physics Faculty and Educational Researchers: Divergent Expectations as Barriers to the Diffusion of Innovations," submitted in April 2006 to *Am. J. Phys.* (*Physics Education Research Section*); online at <<http://homepages.wmich.edu/~chenders/Publications/DivergentExpectationsSubmitted.pdf>> (224 KB).

Henderson, C. & M. Dancy. 2006b. "Barriers to the Use of Research-Based Instructional Strategies: The Dual Role of Individual and Situational Characteristics," submitted in October 2006 to *Physical Review Special Topics: Physics Education Research*; online at <<http://homepages.wmich.edu/~chenders/Publications/SituationalPaperSubmitted.pdf>> (184 KB).

Heron, P.R.L. & D.E. 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).

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.

Hestenes D. & M. Wells. 1992. "A Mechanics Baseline Test," *Phys. Teach.* **30**: 159-166; online (except for the test itself) at <<http://modeling.asu.edu/R&E/Research.html>>. The test (password protected) is available in English, Spanish, German, Malaysian, Finnish, and Turkish.

Hestenes, D & I. Halloun. 1995. "Interpreting the force concept inventory: a response to Huffman and Heller," *Physics Teacher* **33**: 502–506, online at <http://modeling.asu.edu/R&E/Research.html> under "Articles Describing the FCI."

Hilborn, R.C. 2005. "Academic Schizophrenia, The Scholarship of Teaching, and STEM Leadership," Project Kaleidoscope: What works, what matters, what lasts; online at http://www.pkal.org/documents/Hilborn_Schizophrenia.pdf (192 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; abstract online at <http://tinyurl.com/br88n>.

Holton, G. & S.G. Brush. 2001. *Physics the Human Adventure: From Copernicus to Einstein and Beyond*. 3rd edition, Rutgers University Press, pp. 161-162; see Rutgers University Press information at http://rutgerspress.rutgers.edu/acatalog/_Physics_the_Human_Adventure_72.html - 26>.

Hovland, C. I., A. A. Lumsdaine, and F. D. Sheffield. 1949. "A baseline for measurement of percentage change," in C. I. Hovland, A.A. Lumsdaine, and F.D. Sheffield, eds. 1965, "Experiments on mass communication." Wiley (first published in 1949). Reprinted as pages 77-82 in P. F. Lazarsfeld and M. Rosenberg, eds. 1955, *The language of social research: a reader in the methodology of social Research*. Free Press.

IES. 2006. Institute for Education Sciences (U.S. Department of Education) <http://www.ed.gov/about/offices/list/ies/index.html?src=oc>.

JCSEE. 1994. Joint Committee on Standards for Educational Evaluation, *The Program Evaluation Standards*, 2nd ed., Sage. A glossary of evaluation terms from this publication is online at <http://ec.wmich.edu/glossary/prog-glossary.htf>.

Kandel. E.R. 2006. *In Search of Memory: The Emergence of a New Science of Mind*. W.W. Norton. Norton information at <http://www2.wwnorton.com/catalog/fall06/032937.htm>. See also Squire & Kandel (2000).

Kelly A.E. & R.A. Lesh, eds., 2007. *Handbook: Design Research in Education*, in preparation. Erlbaum.

Khodor, J. , D.G. Halme, and G.C. Walker. 2004. "A Hierarchical Biology Concept Framework: A Tool for Course Design," *Cell Biology Education* **3**: 111–121, Summer; online at <http://www.lifescied.org/cgi/reprint/3/2/111>>. They write [see that article references other than Hestenes & Wells (1992), Hestenes et al. (1992), Hake (1998a,b), and Klymkowsky et al. (2003); my insert at ". . . [insert]. . ."]:

The idea of a Concept Inventory as an assessment tool dates back to 1992, when the Force Concept Inventory (FCI) was developed to measure students' conceptual understanding of motion and force (Hestenes and Wells, 1992; Hestenes *et al.*, 1992). . . [more accurately, in my opinion, to the pioneering work of Halloun & Hestenes (1985a,b) in developing the Mechanics Diagnostic test, precursor to the FCI]. . . . A major accomplishment of this work was to create a multiple-choice test in which the erroneous answers diagnose the misconceptions held by students about particular concepts. The FCI has been used over the past decade by physicists at several institutions of higher learning to assess the effectiveness of different teaching methods (Hake, 1998). Similar multiple-choice exams have been developed for astronomy (Astronomy Diagnostic Test [Zeilik *et al.*, 1997; Deming, 2002; Hufnagel, 2002; Zeilik, 2003]) and chemistry (ConceptTests: <http://chem.wisc.edu/concept/>). . . [that link has rotted - try <http://tinyurl.com/sgg4o>]. . . Efforts to create similar standardized tests in biology are now under way. A group headed by Michael Klymkowsky at the University of Colorado at Boulder has been creating concept tests to cover "introductory, genetics, molecular, cellular, and developmental biology" (Klymkowsky *et al.*, 2003). Dianne Anderson and colleagues (2002) have published a concept inventory of natural selection.

Kirschner, P. A., J. Sweller, & R.E. Clark. 2006. "Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching." *Educational Psychologist* **41**(2): 75-86; online at http://www.cogtech.usc.edu/publications/kirschner_Sweller_Clark.pdf> (176 kB).

Klahr, D. & M. Nigam. 2004. "The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning," *Psychological Science* **15**(10): 661-667; online at <http://www.psy.cmu.edu/faculty/klahr/papers.html>>. For a discussion of widespread misinterpretation of this paper see, e.g., "Will the No Child Left Behind Act Promote Direct Instruction of Science?" [Hake (2005e)].

Kluck, A.S. 2005. "Integrating Multiculturalism into the Teaching of Psychology: Why and How?" E-xcellence in Teaching e-column, online at <http://teachpsych.lemoyne.edu/teachpsych/eit/eit2005/eit05-13.pdf>> (48 KB)

Klymkowsky, M.W., K. Garvin-Doxas, & M. Zeilik. 2003. "Bioliteracy and Teaching Efficiency: What Biologists Can Learn from Physicists," *Cell Biology Education* 2: 155-161; online at <<http://www.lifescied.org/cgi/reprint/2/3/155>>. The abstract reads:

The introduction of the Force Concept Inventory (FCI) by David Hestenes and colleagues in 1992 produced a remarkable impact within the community of physics teachers. An instrument to measure student comprehension of the Newtonian concept of force, the FCI demonstrates that active learning leads to far superior student conceptual learning than didactic lectures. Compared to a working knowledge of physics, biological literacy and illiteracy have an even more direct, dramatic, and personal impact. They shape public research and reproductive health policies, the acceptance or rejection of technological advances, such as vaccinations, genetically modified foods and gene therapies, and, on the personal front, the reasoned evaluation of product claims and lifestyle choices. While many students take biology courses at both the secondary and the college levels, there is little in the way of reliable and valid assessment of the effectiveness of biological education. This lack has important consequences in terms of general bioliteracy and, in turn, for our society. Here we describe the beginning of a community effort to define what a bioliterate person needs to know and to develop, validate, and disseminate a tiered series of instruments collectively known as the Biology Concept Inventory (BCI), which accurately measures student comprehension of concepts in introductory, genetic, molecular, cell, and developmental biology. The BCI should serve as a lever for moving our current educational system in a direction that delivers a deeper conceptual understanding of the fundamental ideas upon which biology and biomedical sciences are based.

Klymkowsky, M.W. 2006. "Bioliteracy.net," online at <<http://bioliteracy.net/>>:

Our goal is to generate, test and distribute the tools to determine whether students are learning what teachers think they are teaching. We assume that accurate and timely assessment of student knowledge will pressure the educational world toward more effective teaching. WHY? (a) Because basic understanding of the biological sciences impacts our lives in more and more dramatic ways every year. (b) A wide range of important personal, social, economic and political decisions depend upon an accurate understanding of basic biology and the means by which science generates, tests and extends our knowledge.

Knight, R. 1995. Review of *A Guide to Introductory Physics Teaching* [Arons (1990)], *APS Forum on Education Newsletter*, August 1995 ; online at <<http://www.aps.org/units/fed/newsletters/aug95/knight.cfm>>. Knight ends his insightful review with:

Arons' perspective on the teaching, and learning, of physics is summed up in two questions he insists we pose, over and over, to students: "How do we know ...? Why do we believe ...?" This is, after all, the essence of understanding physics as a science, a way of knowing, rather than as a collection of loosely related formulas. Perhaps, just perhaps, there really is a hope that science education, and physics education in particular, can be improved if Arnold Arons just keeps prodding us along.

Lagemann, E.C. 2000. *An Elusive Science: The Troubling History of Educational Research*. University of Chicago Press - information at <<http://www.press.uchicago.edu/cgi-bin/hfs.cgi/00/13993.cml>>.

Langenberg, D.N. 2000. "Rising to the Challenge," in *Thinking K-14*(1): 19; online as "Honor in the Boxcar" at <<http://tinyurl.com/ulv6w>>.

Lawson, A.E. 2006. "Points of View: On the Implications of Neuroscience Research for Science Teaching and Learning: Are There Any?" *CBE-Life Sciences Education* **5**: 111-117, online at <<http://www.lifescied.org/cgi/content/full/5/2/111>>.

Leammon, R. 1999. "*Thinking About Teaching and Learning: Developing Habits of Learning with First Year College and University Students.*" Stylus. Stylus information at <<http://styluspub.com/Books/BookDetail.aspx?productID=20839>>.

Leammon, R. 2000. "Learning as Biological Brain Change," *Change*, November/December; online at <<http://www.umassd.edu/cas/biology/lbk1.cfm>>.

Levine, A. 2006. *Educating School Teachers: Executive Summary*, Education Schools Project, online at <http://www.edschools.org/pdf/Educating_Teachers_Exec_Summ.pdf> (68 kB). See also the *Inside Higher Ed* report by Powers (2006).

Libarkin, J.C. & S.W. Anderson, 2005. "Assessment of Learning in Entry-Level Geoscience Courses; Results from the Geoscience Concept Inventory," *Journal of Geoscience Education* **53**: 394-401. Online at <http://www.nagt.org/files/nagt/jge/abstracts/Libarkin_v53p394.pdf> (240 kB). Unfortunately, average normalized gains <g> are *not* calculated, so it's difficult to compare this work with similar pre/post testing results in other disciplines. See also Libarkin & Anderson (2006a,b).

Libarkin, J.C. & S.W. Anderson. 2006a. "The Geoscience Concept Inventory: Application of Rasch Analysis to Concept Inventory Development in Higher Education," in Liu & Boone (2006). Online at Libarkin & Anderson (2006b).

Libarkin, J.C. & S.W. Anderson. 2006b. Geoscience Concept Inventory Website at <<http://newton.bhsu.edu/eps/gci.html>>.

Lissitz, R.W., ed. 2005. *Value Added Models in Education: Theory and Applications*. JAM Press. Contents and ordering information are at the Journal of Applied Measurement web site <<http://www.jampress.org>> / "JAM Press Books!" where "/" means "click on." See also <http://home.att.net/~rsmith.arm/LVAM_flyer.pdf> (4 KB). The contributions appear to be focused on K-12.

Liu, X. & W. Boone. 2006. *Applications of Rasch Measurement in Science Education*. JAM Press. Contents and ordering information are at the Journal of Applied Measurement web site <<http://www.jampress.org>> / "JAM Press Books!" where "/" means "click on." See also <http://home.att.net/~rsmith.arm/ARMSE_flyer.pdf>.

Lord, F.M. 1956. "The measure of growth," *Educational and Psychological Measurement* **16**: 42-437.

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

Mahajan, S. & R.R. Hake. 2000. "Is it time for a physics counterpart of the Benezet/Berman math experiment of the 1930's?" *Physics Education Research Conference 2000: Teacher Education*, online at <<http://arxiv.org/abs/physics/0512202>>.

Marchese, T.J. 1997. "The New Conversations About Learning: Insights From Neuroscience and Anthropology, Cognitive Science and Workplace Studies," *Assessing Impact: Evidence and Action*, American Association for Higher Education (now defunct), Washington DC, pp. 79-95, online at <http://www.newhorizons.org/lifelong/higher_ed/marchese.htm>.

Marchese, T.J. 2006. "Carnegie Perspectives: "Whatever Happened To Undergraduate Reform?" online at <<http://www.carnegiefoundation.org/perspectives/sub.asp?key=245&subkey=1736>>; see also the commentary at <<http://www.carnegiefoundation.org/conversations/sub.asp?key=244&subkey=1617>>, including that by Hake at <<http://www.carnegiefoundation.org/perspectives/sub.asp?key=244&subkey=1652>>. See also Marchese (1997).

Marx, J.D. & K. Cummings. 2007. "Normalized change," *Am. J. Phys.* **75**(1): 67-91; an abstract is online at <<http://tinyurl.com/y9kq3p>>.

McCloskey, M. 1983. "Naïve theories of motion," in *Mental Models*, D. Gentner & A.L. Stevens, eds., Erlbaum, pp. 299-324. An ERIC abstract is online at <<http://tinyurl.com/yn2zxc>>.

McConnell, D.A., D.N. Steer, K.D. Owens. 2003. "Assessment and Active Learning Strategies for Introductory Geology Courses," *Journal of Geoscience Education* **51**(2): 205-216, online at <http://www.nagt.org/files/nagt/jge/abstracts/McConnell_Steer_Owens_v51n2p205.pdf> (392 kB).

McCray, R.A., R.L. DeHaan, J.A. Schuck, eds. 2003. *Improving Undergraduate Instruction in Science, Technology, Engineering, and Mathematics: Report of a Workshop*, Committee on Undergraduate STEM Instruction, National Research Council, National Academy Press; online at <<http://www.nap.edu/catalog/10711.html>>. For criticisms of this report's unfortunate neglect of (a) the pioneering research of Halloun & Hestenes (1985a,b), and (b) pre/post testing evidence for the effectiveness of interactive engagement methods, see "NRC's CUSE: Oblivious of the Advantage of Pre/PostTesting with High Quality Standardized Tests?" [Hake (2003c)], and "NRC's CUSE: Stranded on Assessless Island?" [Hake (2003d)].

McDermott, L.C. & P.S. Shaffer. 1992. "Research as a guide for curriculum development: an example from introductory electricity. Part I: investigation of student understanding." *Am. J. Phys.* **60**(11): 994-1003, abstract online at <<http://tinyurl.com/y8hakw>>.

Meltzer, D.E. 2002. "The relationship between mathematics preparation and conceptual learning gains in physics: A possible 'hidden variable' in diagnostic pretest scores," *Am. J. Phys.* **70**(12): 1259-1268; online as article #7 at <<http://www.physicseducation.net/articles/index.html>>.

Meltzer, D.E. 2006a. "Links to United States Physics Education Research Groups," online at <<http://www.physicseducation.net/links/index.html>>.

Meltzer, D.E. 2006b. "Formative Assessment Materials for Large-Enrollment Physics Lecture Classes," invited talk and poster at the National STEM [Science, Technology, Engineering, and Mathematics] Assessment Conference [NSF (2006)]; online at <<http://www.physicseducation.net/talks/index.html>>.

Meltzer, D.E. 2006c. "Re: Article of interest," PhysLrnR post of 5 Dec 2006 21:34:25-0800, online at <<http://tinyurl.com/ye94g8>>.

Michael, J. 2006. "Where's the evidence that active learning works?" *Advances in Physiology Education* **30**: 159-167, online at <<http://advan.physiology.org/cgi/content/full/30/4/159-T1>>.

Mintzes, J.J., J.H. Wandersee, & J.D. Novak. 1999. *Assessing Science Understanding*. Academic Press.

Mislevy, R. 2006: (a) "On approaches to assessing change," and (b) "Clarification"; both online at <<http://www.education.umd.edu/EDMS/mislevy/papers/Gain/>>.

Moore, T. 2005. "Six Ideas That Shaped Physics: Evidence of Success," online at <<http://www.physics.pomona.edu/sixideas/sisuc.html>>.

Morris, G.A., L. Branum-Martin, N. Harshman, S.D. Baker, E. Mazur, S. Dutta, T. Mzoughi, & V. McCauley. 2006. *Am. J. Phys.* **74**(5): 449-453, "Testing the test: Item response curves and test quality," abstract online at <<http://tinyurl.com/yk8qfz>>.

Morote, E, & D. Pritchard. 2002. "What Course Elements Correlate With Improvement On Tests In Introductory Newtonian Mechanics?" online at <<http://relate.mit.edu/effectiveness.pdf>> (104 kB).

Mulford, D. R & W.R. Robinson. 2002. "An inventory for alternate conceptions among first-semester general chemistry students," *J. Chem. Educ.* **79**: 739-744; an abstract is online at <<http://jchemed.chem.wisc.edu/Journal/Issues/2002/Jun/abs739.html>>.

NCSU. 2006. "Assessment Instrument Information Page," Physics Education R & D Group, North Carolina State University; online at <<http://www.ncsu.edu/per/TestInfo.html>>.

Nelson, C. 2000. "Bibliography: How To Find Out More About College Teaching and Its Scholarship: A Not Too Brief, Very Selective Hyperlinked List." (College Pedagogy is A Major Area Of Scholarship!); online at <<http://php.indiana.edu/~nelson1/TCHNGBKS.html>>.

Newton, R.G. 1997. *The Truth of Science: Physical Science and Reality*. Harvard University Press - information at <<http://www.hup.harvard.edu/catalog/NEWWHA.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). See also Nichols & Berliner (2007).

Nichols S.L. & D. Berliner. 2007. *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Harvard Education Press. See <<http://www.hepg.org/hep/Book/62>>, which carries the praise of UCLA emeritus professor W. James Popham:

This savage assault on high-stakes testing in education arrives with a clear concern about those most harmed by high-stakes tests—students and teachers. Nichols and Berliner provide a carefully reasoned analysis laced with frightening accounts drawn from public schools. Not merely another pummeling of *No Child Left Behind*, this is a readable evisceration of the premise that our schools can be evaluated with a single indicator. If you care about public schooling, this is required reading.

NSF. 1996. National Science Foundation Advisory Committee. *Shaping the Future, Volume II: Perspectives on Undergraduate Education in Science, Mathematics, Engineering, and Technology*, Advisory Committee to the National Science Foundation Directorate for Education and Human Resources, chaired by Melvin George, online at: <<http://www.nsf.gov/pubs/1998/nsf98128/nsf98128.pdf>> (1.8 MB). This report is one of the few that emphasizes the crucial role of higher education in determining the quality of K-12 education.

NSF. 2006. *Assessing Student Achievement*, Conference of 19-21 October in Washington, D.C. An announcement is online at <<http://www.drury.edu/multinl/story.cfm?ID=17783&NLID=306>>, and the program is at <<http://www.drury.edu/multinl/story.cfm?ID=17789&NLID=306>>. According to PKAL (2006b), a Proceedings of the conference will be published in 2007.

Nuhfer, E. 2006a. "A Fractal Thinker Looks at Measuring Change: Part 1 - Pre-Post Course Tests and Multiple Working Hypotheses - Educating in Fractal Patterns XVI," *National Teaching and Learning Forum* **15**(4). Online to subscribers at <<http://www.ntlf.com/FTPSite/issues/v15n4/diary.htm>>. If your institution doesn't have a subscription, IMHO it should.

Nuhfer, E. 2006b. "A Fractal Thinker Looks at Measuring Change: Part 2 - Pre-Post Assessments - Are All Interpretations Equally Valid? Educating in Fractal Patterns XVII," *National Teaching and Learning Forum* **15**(6). Online to subscribers at <<http://www.ntlf.com/FTPSite/issues/v15n6/diary.htm>>.

OERL. 2006. *Online Evaluation Resource Library*, online at <<http://oerl.sri.com/>>. This resource, though technically sophisticated, is evidently in the early stages of development. Searches on 6 December 2006 at <<http://oerl.sri.com/search/instrSearch.jsp>> gave *zero* hits for science-related resources from these hotbeds of physics education research, development, and evaluation: Dickinson College, Harvard University, Indiana University, University of Maryland, North Carolina State University, University of Colorado at Boulder, University of Massachusetts at Amherst, and the University of Washington. [For a list of URL's for physics education websites see Meltzer (2006).] Thus, the introduction below to OERL (2006) may be an indication of *potential* rather than present capability.

Welcome to OERL, the Online Evaluation Resource Library. This library was developed for professionals seeking to design, conduct, document, or review project evaluations. The purpose of this system is to collect and make available evaluation plans, instruments, and reports for NSF projects that can be used as examples by Principal Investigators, project evaluators, and others outside the NSF community as they design proposals and projects. OERL also includes professional development modules that can be used to better understand and utilize the materials made available. OERL's mission is to support the continuous improvement of project evaluations. Sound evaluations are critical to determining project effectiveness. To this end, OERL provides: (a) a large collection of sound plans, reports, and instruments from past and current project evaluations in several content areas, and (b) guidelines for how to improve evaluation practice using the Web site resources.

Pavelich, M., B. Jenkins. J. Birk, R. Bauer, & S. Krause. 2004. "Development of a Chemistry Concept Inventory for Use in Chemistry, Materials and other Engineering Courses," *Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition* online at <http://www.asee.org/acPapers/2004-1907_Final.pdf> (124 kB).

Pellegrino, J.W., N. Chudowsky, R. Glaser, eds. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. National Academy Press; online at <<http://www.nap.edu/catalog/10019.html>>. As is common in the psychometric literature, the focus is on K-12.

Phillips, D.C. 2000. *Expanded social scientist's bestiary: a guide to fabled threats to, and defenses of, naturalistic social science*. Rowman & Littlefield - information at <<http://tinyurl.com/ycmlvy>>. See especially Chapter 9 on "Positivism."

Phillips, D.C. & N.C. Burbules. 2000. "*Postpositivism and Educational Research*." Rowman & Littlefield - information at <<http://tinyurl.com/yncvls>>. See especially "Mistaken accounts of positivism," pp. 11-14.

Pister, K. 1996. "Renewing the research university," *University of California at Santa Cruz Review* (Winter), quoted in NSF (1996).

PKAL. 2002. Project Kaleidoscope, *Report on Reports I, 2002: Recommendations for Action in Support of Undergraduate Science, Technology, Engineering, and Mathematics*, online at <<http://www.pkal.org/documents/ReportOnReports.cfm>>.

PKAL. 2006a. Project Kaleidoscope, *Report on Reports II, 2006: Report on Reports II, 2006: Recommendations for Urgent Action, Transforming America's Scientific and Technological Infrastructure*, online at <<http://www.pkal.org/documents/ReportOnReportsII.cfm>>:

PKAL's *raison d'être* is strengthening undergraduate STEM. Yet from the reports cited here (and from the others that seem to be issued daily) what is urgently needed is fundamental change of the entire system. Short-term, piece-meal, sector by sector, underfunded and uncoordinated efforts will not move America confidently and creatively forward in the next, challenging decades of the 21st century.

PKAL. 2006b. Project Kaleidoscope, Volume IV: *What works, what matters, what lasts: Assessing Student Achievement*, November 7; online at <<http://www.pkal.org/collections/AssessingStudentAchievement.cfm>>. A report on the NSF's *Assessing Student Achievement (ASA) Conference* of 19-21 October 2006. See also PKAL (2002, 2006a). It is stated that "As projects mature, they are being placed on two web sites for easy access to the increasing number of STEM faculty looking for tested tools to adapt in their own setting: OERL (2006), 'The Online Evaluation Resource Library,' and FLAG (2006), the 'Field-tested Learning Assessment Guide.'" But as indicated in OERL (2006) the coverage there neglects the extensive higher-education assessment carried out by physics education researchers.

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>>.

Powell, K. 2003. "Spare me the lecture: US research universities, with their enormous classes, have a poor reputation for teaching science. Experts agree that a shake-up is needed, but which strategies work best? Kendall Powell goes back to school," *Nature* **425**, 18 September, pp. 234-236; online as a 388 kB pdf at <<http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v425/n6955/index.html>>, scroll down about 1/3 of the page to "News Features." For a critique of Powell's review see Hake (2003b). Powell wrote:

Evidence of . . . [the failure of the passive-student lecture]. . . is provided by assessments such as the Force Concept Inventory (FCI), a multiple-choice test designed to examine students' understanding of Newton's laws of mechanics. Developed around a decade ago by David Hestenes, a physicist turned education researcher at Arizona State University in Tempe, the FCI has changed some researchers' opinions of their teaching techniques.

Powers, E. 2006. "New Critique of Teacher Ed," *Inside Higher Ed*. 19 September, online at <<http://insidehighered.com/news/2006/09/19/teachered>>. A report on Levine's (2006) criticism of teacher education.

Ravitch, D. 2002. "Education after the culture wars," in *Daedalus* (2002), online at <<http://www.amacad.org/publications/summer2002/ravitch.pdf>> (704 kB).

Redish, E. F. 1994. "Implications of cognitive studies for teaching physics," *Am. J. Phys.* **62**(9): 796-803, online at <<http://www.physics.umd.edu/perg/papers/redish/cogsci.html>>.

Redish, E.F. 1999. "Millikan Award Lecture 1998: Building a Science of Teaching Physics," *Am. J. Phys.* **67**(7): 562-573; online at <http://www.physics.umd.edu/perg/papers/redish/millikan.htm>. See also Redish (1994).

Redish, E.F. 2006. "Re: Can Neuroscience Benefit Classroom Instruction?" PhysLrnR post 15 Oct 2006 22:12:17-0400; online at <http://tinyurl.com/tqzhe>. In the last sentence, the URL for Redish's American Association of Physics Teachers (AAPT) talk "How having a theory of learning changes what I do in class on Monday," should have been given as <http://www.physics.umd.edu/perg/talks/redish/Monday.pdf> (2.1 MB).

Reif, F. 1974. "Educational Challenges for the University," *Science* **184**, 537-542.

Rhem, J. 2006. "The High Risks of Improving Teaching," *National Teaching and Learning Forum* **15**(6), online to subscribers at <http://www.ntlf.com/FTPSite/issues/v15n6/risks.htm>, also disseminated by the *Tomorrow's Professor* list <http://ctl.stanford.edu/Tomprof/postings.html> as TP Msg. #760 on 20 Nov 2006.

Rogers, E.M. 2003. *Diffusion of Innovations*, 5th edition. Free Press.

Rogosa, D.R., D. Brandt, & M. Zimowski. 1982. "A growth curve approach to the measurement of change," *Psychological Bulletin* **92**: 726-748.

Rogosa, D.R., & J.B. Willett. 1983. "Demonstrating the reliability of the difference score in the measurement of change," *Journal of Educational Measurement* **20**: 335- 343.

Rogosa, D. R. & Willet, J. B. 1985. "Understanding correlates of change by modeling individual differences in growth," *Psychometrika* **50**: 203-228. Abstract online at <http://www.springerlink.com/content/ru93345312046086/>.

Rogosa, D.R. 1995. "Myth and methods: 'Myths about longitudinal research' plus supplemental questions," in J.M. Gottmann, ed. *The Analysis of Change*, pp. 3-66. Erlbaum; examples from this paper are online at <http://www.stanford.edu/~rag/Myths/myths.html>.

Rothman, F.G. & J.L. Narum. 1999. *Then, Now, and in the Next Decade: A Commentary on Strengthening Undergraduate Science, Mathematics, Engineering, and Technology Education*, Project Kaleidoscope, online at <http://www.pkal.org/documents/ThenNowAndInTheNextDecade.cfm>. Frank Rothman is Provost Emeritus of Brown University and Jeanne Narum is the director of Project Kaleidoscope <http://www.pkal.org/>.

Rudner, L.M. 2001. "Item Response Theory," online at <http://edres.org/irt/>. This site offers a goldmine of information on IRT and its practical use, including free software.

- Sadler, P.M. 1999. "The Relevance of Multiple Choice Testing in Assessing Science Understanding," in Mintzes et al. (1999, pp. 249-278).
- Savinainen A. & P. Scott. 2002a. "The Force Concept Inventory: a tool for monitoring student learning," *Phys. Educ.* **37**: 45-52; online at <<http://kotisivu.dnainternet.net/savant/>>.
- Savinainen A. & P. Scott. 2002b. "Using the Force Concept Inventory to monitor student learning and to plan teaching," *Phys. Educ.* **37**: 53-58; online at <<http://kotisivu.dnainternet.net/savant/>>.
- 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>>.
- Scriven, M. 2006. "Converting Perspective to Practice." *Journal of MultiDisciplinary Evaluation*, Number 6, November, online at <http://evaluation.wmich.edu/jmde/JMDE_Num006.html>. A response to Cook (2006).
- Shadish, W.R., T.D. Cook, & D.T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin - information at <<http://tinyurl.com/y3e7vw>>. A goldmine of references to social-science research.
- Shavelson, R.J. & L. Towne, eds., 2002. *Scientific Research in Education*, National Academy Press; online at <<http://www.nap.edu/catalog/10236.html>>.
- 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/>> as the first "highlighted" paper.
- Shulman, L. 1986. "Those who understand: knowledge growth in teaching. " *Educational Researcher* **15**(2):4-14, online "soon" (but don't hold your breath) at <<http://www.aera.net/publications/?id=331>>.
- Shulman, L. 1987. "Knowledge and teaching: foundations of the new reform." *Harvard Educational Review* 57:1-22.
- Singer J.D. & J.B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press - information at <<http://gseacademic.harvard.edu/alda/>>.

Smith, K.A., S. D. Sheppard, D.W. Johnson, & R.T. Johnson. 2005. "Pedagogies of Engagement: Classroom-Based Practices," *Journal of Engineering Education* **94**(1): 87-102; online as a 492 kB pdf at <<http://tinyurl.com/y939x2>>. They write [my inserts at "...[insert]. . ."]:

Research that has had a significant influence on the instructional practices of engineering faculty is Hake's (1998a,b) comparison of students' scores on the physics Force Concept Inventory (FCI), a measure of students' conceptual understanding of mechanics, in traditional lecture courses and interactive engagement courses. The results shown for high school (HS), college (COLL), and university (UNIV) students in. . .[Fig. 1 of Hake (1998a)]. . . show that student-student interaction during class time is associated with a greater percent . . . [normalized]. . . gain on the FCI. Further study of the figure shows that even the best lectures. . . [more accurately the "best traditional courses" since *non-passive* student lectures such those by Mazur [Crouch & Mazur (2001)] yield relatively high normalized gains]. . . achieve student gains that are at the low end of student . . . [normalized]. . . gains in interactive engagement classes.

Squire, L.R. & E.R. Kandel. 2000. *Memory: From Mind to Molecules*. W.H. Freeman. See also Kandel (2006).

Stein, S. 1997. "Preparation of Future Teachers," *Notices of the AMS* **44** (3); 311-312; online at <<http://www.ams.org/notices/199703/letters.pdf>> (108 kB).

Stokstad, E. 2001. "Reintroducing the Intro Course," *Science* **293**: 1608-1610, 31 August. Stokstad wrote: "Physicists are out in front in measuring how well students learn the basics, as science educators incorporate hands-on activities in hopes of making the introductory course a beginning rather than a finale."

Suskie, L. 2004a. *Assessing Student Learning: A Common Sense Guide*. Anker Publishing. Anker information is at <<http://tinyurl.com/yzgvo>>.

Suskie, L. 2004b. "Re: pre- post testing in assessment," ASSESS post of 19 Aug 2004 08:19:53-0400; online at <<http://tinyurl.com/akz23>>. See also Suskie (2004a).

Swartz, C. E. 1999. "Editorial: Demise of a Shibboleth," *The Physics Teacher* **37**: 330, online to subscribers at <<http://tinyurl.com/y33cto>>. Swartz wrote [my *italics*]:

There is a variety of evidence, and claims of evidence, that each of the latest fads . . . (constructivism, "group" and "peer" instruction, "interaction") . . . produces superior learning and happier students. In particular, students who interact with apparatus or lecture do better on the "Force Concept Inventory" exam [Hestenes et al., 1992]. The evidence of Richard Hake's (1998a) metastatistical study is so dramatic that the only surprising result is that many schools and colleges are still teaching in old-fashioned ways. Perhaps the interaction technique reduces coverage of topics, or perhaps the method requires new teaching skills that teachers find awkward. *At any rate the new methodology is not sweeping the nation.*

Thompson, B. 2006. *Foundations of Behavioral Statistics: An Insight-Based Approach*. Guilford Press - information at <<http://tinyurl.com/vftol>> and <<http://www.coe.tamu.edu/~bthompson/datasets.htm>>.

Thornton, R.K. & D.R. Sokoloff. 1998. "Assessing student learning of Newton's Laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula," *Am. J. Phys.* **66**(4): 338-352.

Tobias S. & R.R. Hake. 1988. "Professors as physics students: What can they teach us?" *Am. J. Phys.* **56**: 786-794, online at <<http://www.physics.indiana.edu/~hake/ProfsAsStudents.pdf>> (1.1 MB).

Tobias, S. 2000. "Guest comment: From innovation to change: Forging a physics education agenda for the 21st century," *Am. J. Phys.* **68**(2): 103-104; online to subscribers at <<http://scitation.aip.org/dbt/dbt.jsp?KEY=AJPIAS&Volume=68&Issue=2>>.

Tuma, D.T. & F. Reif, eds. 1980. *Problem Solving and Education: Issues in Teaching and Research*, Lawrence Erlbaum.

USDE. 2003. U.S. Department of Education, *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*. Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. The entire guide is online at <<http://www.ed.gov/rschstat/research/pubs/rigorousvid/rigorousvid.pdf>> (140 KB). The Guide's authoring group, the Coalition for Evidence-Based Policy (CEBP) <<http://coexgov.securesites.net/index.php?keyword=a432fbc34d71c7>> was formerly a part of the Institute of Education Sciences [IES (2006)], in turn a part of the USDE [for the structure of this bureaucratic colossus see <<http://www.ed.gov/about/offices/or/index.html?src=ln>>]. The CEBP is now sponsored by the "council for excellence in government" <<http://coexgov.securesites.net/index.php>>, with "the mission to promote government policymaking based on rigorous evidence of program effectiveness." The CEBP's Board of Advisors <<http://coexgov.securesites.net/index.php?keyword=a432fbc71d7564>> includes luminaries such as famed Randomized Control Trial (RCT) authority Robert Boruch (University of Pennsylvania); political economist David Ellwood (Harvard); former FDA commissioner David Kessler (Univ. of California - San Francisco); past American Psychological Association president Martin Seligman (University of Pennsylvania); psychologist Robert Slavin (Johns Hopkins); economics Nobelist Robert Solow (MIT); and progressive-education basher Diane Ravitch. Unfortunately, no physical scientists, mathematicians, philosophers, or K-12 teachers are members of the CEBP.

USDE. 2005. U.S. Department of Education, *No Child Left Behind Act*, online at <<http://www.ed.gov/nclb/landing.jhtml?src=pb>>.

USDE. 2006. U.S. Dept. of Education, *A Test Of Leadership: Charting the Future of U.S. Higher Education*, pre-publication copy, September, online at <http://www.ed.gov/about/bdscomm/list/hiedfuture/reports/pre-pub-report.pdf> (7.1 MB). The report states [my italics]:

We believe that improved accountability is vital to ensuring the success of all the other reforms we propose. *Colleges and universities must become more transparent about cost, price, and student success outcomes, and must willingly share this information with students and families. Student achievement, which is inextricably connected to institutional success, must be measured by institutions on a "value-added" basis that takes into account students' academic baseline when assessing their results.* This information should be made available to students, and reported publicly in aggregate form to provide consumers and policymakers an accessible, understandable way to measure the relative effectiveness of different colleges and universities.

Vlastos, G. 1990. Private communication to R.R. Hake, September 17. Vlastos wrote: "Though Socrates was not engaged in physical inquiry, your program . . . [Socratic Dialogue Inducing Labs Labs - Hake (1992, 2002 f,g,h,j)]. . . is entirely in his spirit."

Vlastos, G. 1991. *Socrates, Ironist and Moral Philosopher*. Cornell Univ. Press and Cambridge University Press, esp. Chap. 2, "Socrates contra Socrates in Plato." Cambridge University Press information is at <http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521314503>:

This long-awaited study of the most enigmatic figure of Greek philosophy reclaims Socrates' ground-breaking originality. Written by a leading historian of Greek thought, it argues for a Socrates who, though long overshadowed by his successors Plato and Aristotle, marked the true turning point in Greek philosophy, religion and ethics. *The quest for the historical figure focuses on the Socrates of Plato's earlier dialogues, setting him in sharp contrast to that other Socrates of later dialogues, where he is used as a mouthpiece for Plato's often anti-Socratic doctrine.* [My italics.] At the heart of the book is the paradoxical nature of Socratic thought. But the paradoxes are explained, not explained away. The book highlights the tensions in the Socratic search for the answer to the question "How should we live?" Conceived as a divine mandate, the search is carried out through elenctic argument, and dominated by an uncompromising rationalism. The magnetic quality of Socrates' personality is allowed to emerge throughout the book. Clearly and forcefully written, philosophically sophisticated but entirely accessible to non-specialists, this book will be of major importance and interest to all those studying ancient philosophy and the history of Western thought.

Vlastos, G. 1994. *Socratic Studies*, edited by Myles Burnyeat, Cambridge University Press - information at <http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521447355>.

Wells, H.G. 1920. *The Outline of History*. For an interesting history of this treatise see http://en.wikipedia.org/wiki/The_Outline_of_History. For Amazon.com information on a two volume set published in 1974 by Scholarly Press see <http://tinyurl.com/yjs83d>.

Wieman, C. & K. Perkins. 2005. "Transforming Physics Education," *Phys. Today* **58**(11): 36-41; online at <http://www.colorado.edu/physics/EducationIssues/papers/PhysicsTodayFinal.pdf> (292 kB). Carl Wieman was awarded the 2001 Nobel prize in physics.

Wiggins, G. 1991. "Toward Assessment Worthy of the Liberal Arts: The Truth May Make You Free, but the Test May Keep You Imprisoned," Appendix to the MAA's *Getting Started With Assessment*, online at <<http://www.maa.org/saum/gettingstarted.html>>.

Willis, J. 2006. "Research watch II: Add the Science of Learning to the Art of Teaching to Enrich Classroom Instruction," *National Teaching and Learning Forum* **15**(5), online to subscribers at <<http://www.ntlf.com/FTPSite/issues/v15n5/research2.htm>>.

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>.

Wittmann, W.W. 1997. "The reliability of change scores: many misinterpretations of Lord and Cronbach by many others; revisiting some basics for longitudinal research," online at <<http://www.psychologie.uni-mannheim.de/psycho2/psycho2.en.php3?language=en>> / "Publications" / "Papers and preprints" where "/" means "click on."

Wood, W.B., & J.M. Gentile. 2003. "Teaching in a research context," *Science* **302**: 1510; 28 November; online to subscribers at <<http://www.sciencemag.org/content/vol302/issue5650/index.shtml> - [policyforum](http://www.sciencemag.org/content/vol302/issue5650/index.shtml#policyforum)>. A summary is online to all at <<http://www.sciencemag.org/cgi/content/summary/302/5650/1510>>.

Wosilait, K., P.L. Heron, P.S. Shaffer, and L.C. McDermott. 1998. "Development and assessment of a research-based tutorial on light and shadow," *Am. J. Phys.* **66**(10): 906–913; an abstract is online at <<http://tinyurl.com/yb5luk>>.

Ziman, J. 1969. "Information, Communication, Knowledge," *Nature* **224** (52 17): 324. An abstract with references is online at <<http://www.nature.com/nature/journal/v224/n5217/abs/224318a0.html>>. The abstract is: "At the British Association meeting in Exeter last month, Professor Ziman addressed the section devoted to general topics on the question of how scientific information becomes public knowledge. The system of communication, he implied, is not as rotten as some like to think."

Ziman, J. 2000. *Real Science: What it is, and what it means*. Cambridge University Press. See, especially Sec. 9.3 "Codified knowledge," pages 258-266.

Zimmerman, D.W. 1997. "A geometric interpretation of the validity and reliability of difference scores," *British Journal of Mathematical and Statistical Psychology* **50**: 73-80.

Zimmerman, D.W. & R.H. Williams. 1982. "Gain scores in research can be highly reliable," *Journal of Educational Measurement* **19**: 149-154. The abstract at <http://mypage.direct.ca/z/zimmerma/earlier.html> reads:

The common belief that gain scores are unreliable is based on certain assumptions about the values of parameters in a well known formula for the reliability of differences. In this paper we show that a reliability coefficient calculated from the formula can be high, provided one makes other assumptions about the values of pretest and posttest reliability coefficients and standard deviations. Furthermore, there is reason to believe that the revised assumptions are more realistic than the usual ones in testing practice.

Zull, J.E. 2002. *The Art of Changing the Brain: Enriching the Practice of Teaching by Exploring the Biology of Learning*, Stylus Press. Stylus information at <http://styluspub.com/Books/BookDetail.aspx?productID=44780>.

Zull, J.E. 2003. "What is 'The Art of Changing the Brain?'" New Horizons for Learning, online at <http://www.newhorizons.org/neuro/zull.htm>. See also Zull (2002).

Zumbo, B. D. 1999. "The simple difference score as an inherently poor measure of change: Some reality, much mythology," in Thompson, ed., *Advances in Social Science Methodology*, Volume 5, pp. 269-304. JAI Press; online at <http://tinyurl.com/kuf3t> (2.3 MB).