The Impact of Concept Inventories On Physics Education and It's Relevance For Engineering Education†*

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

ABSTRACT

I review the:
(a) Before Concept Inventory (BCI) *dark ages* of post-secondary introductory physics education;
(b) 1985 advent of the first physics CI, the *Mechanics Diagnostic* (MD) by Halloun & Hestenes (HH);
(c) 1987-90 early research use of the (MD) by HH, Hake, and Crouch & Mazur;
(d) 1992 *Force Concept Inventory* (FCI), successor to the MD, and the Factor Analysis Debate (1995);
(e) 1995 revision of the FCI by Halloun, Hake, Mosca, and Hestenes;
(f) 1998 meta-analysis of FCI/MD results on 62 introductory physics courses (N = 6542) showing about a two-standard-deviation superiority in average normalized gains <g> for "interactive engagement" over traditional passive-student lecture courses by Hake and subsequent confirmation by about 25 other physics education research studies.

I then indicate:
(a) fourteen hard lessons from the physics education reform effort;
(b) suggestions for the administration and reporting of CI's;
(c) listings of CI's, including those for physics and engineering; *and comment that*:
(d) for physics education the road to reform has been all *uphill*;
(e) the glacial inertia of the educational system, though not well understood, appears to be typical of the slow *Diffusion of Innovations* [Rogers (2003)] in human society;
(f) there are at least "Eleven Barriers to Change in Higher Education";
(g) but, even so, for physics education, Rogers' "early adopters" of reform have now appeared at Harvard, North Carolina State University, MIT, the Univ. of Colorado, California Polytechnic at San Luis Obispo, and the Univ. of British Columbia, possibly presaging a Rogers "take off" for physics education reform, about two decades ACI (After Concept Inventory).

I conclude that:
(a) CI's can stimulate reform, but judging from the results in physics it may take about two decades before even early adopters become evident;
(b) there are at least seven reasons why the rate of adoption of reforms may be greater in engineering education than in physics education.

In an Appendix I respond to criticisms of the FCI and the average normalized gain <g>.

---

† The reference is: Hake, R.R. 2011. "The Impact of Concept Inventories On Physics Education and It's Relevance For Engineering Education" invited talk, 8 August, second annual NSF-sponsored "National Meeting on STEM Concept Inventories," Washington, D.C., online as ref. 64 at <http://bit.ly/a6M5y0>.

Table of Contents

TOPIC PAGE

TOPIC	PAGE

TOPIC **PAGE**

I. The Dark Ages of Physics Education*

A. The Traditional Introductory *Passive-Student* Lecture Course As Seen By:

1. Eric Schocket†

Eric Schocket (1966-2006)

"*The lack of community, together with the lack of interchange between the professor and the students combines to produce a totally passive classroom experience*. . . .The best classes I had were classes in which I was constantly engaged, constantly questioning and pushing the limits of the subject and myself. The way this course is organized accounts for the lack of student involvement . . . . The students are given premasticated information simply to mimic and apply to problems. *Let them rather, be exposed to conceptual problems, try to find solutions to them on their own, and then help them to understand the mistakes they make along the way*." (My *italics*.)

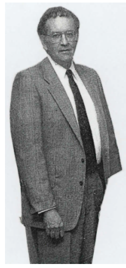
Eric Schocket, as quoted by Sheila Tobias (1994) in *They're Not Dumb, They're Different: Stalking the Second Tier*.

---

*This section (pp. 6-8) is derived from a talk "What Can We Learn from the Physics Education Reform Effort?" [Hake (2000)] at the ASME Mechanical Engineering Education Conference in Fort Lauderdale, Florida, 26-29 March 2000.

† At the time Schocket was quoted by Tobias, he was a graduate student - BA summa cum laude, literature, UC Berkeley. According to Wikipedia <http://bit.ly/qqctXs>, Schocket received a PhD in American Literature from Stanford University, was then an Associate Professor of American literature at Hampshire College, wrote primarily on issues of class, and (sadly) died of leukemia in 2006.

2. John Rigden* (1984)



John Rigden

<u>THE INTODUCTORY COURSE</u>

THE STUDENTS: Most of them come because they MUST - they are premeds, **aspiring engineers**, budding chemists. For them physics is required.

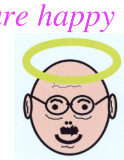
THE FACULTY: They realize that the introductory course is a formidable challenge: the density of concepts is high, it quickly builds upon itself adding a logical layer upon another logical layer. Most of all the pace is relentless. There are no pauses for reflection; tomorrow another chapter will be started!

THE STUDENTS: Free-body diagrams are a bane to their sanity and the normal force is orthogonal to their sense of rationality. Newton's second law is an algebraic relation to solve for one quantity when good fortune provides the other two. Action-reaction pairs get confused: they act in the same direction, on the same body, but they show why a horse can pull a cart.

THE FACULTY: The objective of this course is to uproot the common-sense ideas based on everyday experience and to sow the seeds that will flower into the Newtonian view - airless, frictionless, idealized, abstract. Moreover the faculty have the somber realization that this conceptual transformation must be consummated in a few class periods. Quite a challenge: after all, it required the physicists of earlier eras years and years to assimilate the same ideas.

THE STUDENTS: They do not regard the introductory course as an opportunity to understand a few general principles that can be applied across a range of natural phenomena; they do not see that the course bears any significance to their day-to-day activities or to their future. *They leave the introductory course with a disjointed, collage-like idea as to the content of physics and they leave with no idea whatsoever how it is that we know what we know.*

THE FACULTY: *They leave with a rekindled and deepened awareness of the conceptual richness of the introductory course. They are happy with the way they have brought the ideas together.*



The Introductory Course …….. **<u>ILLUSION !</u>**

---

*John Rigden <http://en.wikipedia.org/wiki/John_Rigden> is, among other things, a former editor of the *American Journal of Physics*.

2. Arnold Arons*

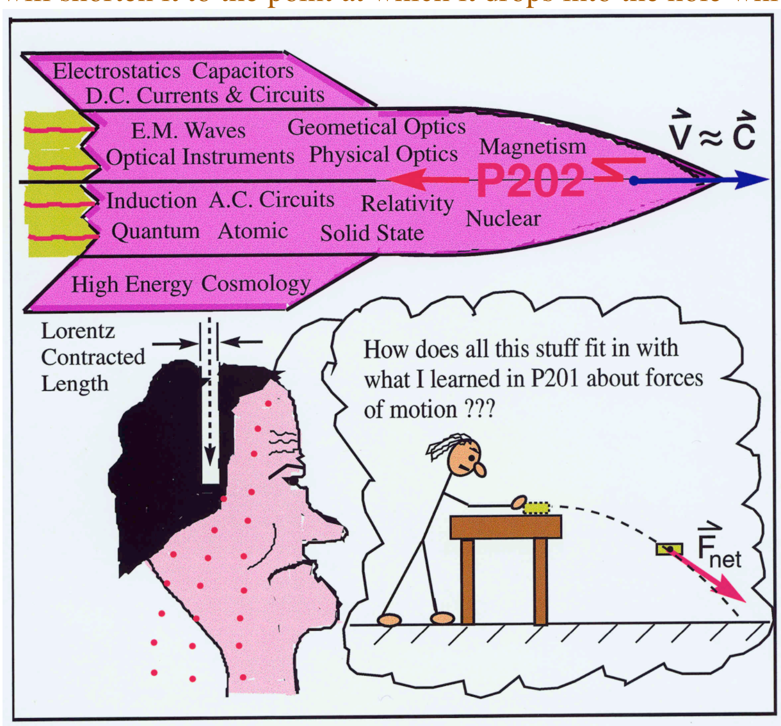
Arnold Arons (1916 - 2001)

THE STANDARD RELATIVISTIC MODEL OF INSTRUCTION [Arons (1986)]

"The relativistic model is based on the premise that, if one starts with an

E - N - O - R - M - O - U - S

breadth of subject matter but passes it by the student at sufficiently high velocity, the Lorentz contraction will shorten it to the point at which it drops into the hole which is the student mind."



*Arnold Arons, along with Robert Karplus, can fairly be called one of the founding fathers of U.S. Physics Education Research - see "The Arons Advocated Method" [Hake (2004a)].

B. Most American Physics Nobelists Are Products of Traditional Introductory Physics Courses, So Such Courses Must Be OK?

1. Response to Physicist Jack Uretsky

Jack Uretsky (2009) of the Phys-L list wrote: "American education has produced a number of physics Nobelists. How many were products of physics courses that would be approved by PER enthusiasts?"

In response I wrote [Hake (2009d), slightly edited)]; bracketed by lines "HHHHHH. . . ."]:

HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Probably near zero American physics Nobelists were products of "Interactive Engagement" courses. Instead, *most Nobelists are products of traditional introductory physics courses.* BUT SO WHAT? The hard facts are that the *average* present-day student of introductory physics - *even at Harvard and MIT* - reacts passively to lectures and learns very little from them. This has been demonstrated, for example, by:

(a) the very low average <<g>> of the class-average *normalized* gains <g> of 0.23 [plus or minus 0.04 (std dev)] on tests of conceptual understanding of Newtonian Mechanics in 14 traditional lecture courses surveyed in "Interactive-engagement vs traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses" (Hake1998a,b); and

(b) corroborative pre/post test results in about 25 other Physics Education Research (PER) papers as listed in Hake (2008a).

Long aware of the deficiencies of the average introductory physics courses - see e.g., the review by McDermott & Redish (1999) - PER's have attempted to design courses - see e.g., Hake (1998b) - which enhance the learning of the vast majority of AVERAGE students, not potential Nobelists.

Why the emphasis on the "average student" rather than the "exceptional student"? Because most exceptional students will learn on their own, with or without the (for them) usually helpful but unnecessary "interactive engagement." On the other hand, *U.S. competitiveness* stressed in e.g., *Rising Above the Gathering Storm* [NRC (2007a)], *Is America Falling Off the Flat Earth?* [NRC (2007b)], and *Rising Above the Gathering Storm, Revisited* [NRC (2010) (not to mention the *fate of life on Planet Earth*) is in the hands and minds of the masses of "average students" who provide the work force and, at least in democracies, control national policy - see e.g., "The Threat to Life on Planet Earth Is a More Important Issue Than David Brooks' Skills Slowdown" [Hake (2009b)].

A quote from the late physics education guru Arnold Arons (1997, p. vii) seems appropriate:

> "I point to the following unwelcome truth: much as we might dislike the implications, research is showing that didactic exposition of abstract ideas and lines of reasoning (however engaging and lucid we might try to make them) to passive listeners yields pathetically thin results in learning and understanding - except in the very small percentage of students who are specially gifted in the field."

HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

C. Likewise, Many Accomplished Scientists, Technologists, Engineers, & Mathematicians Are Products of Traditional Introductory Physics Courses, So Such Courses Must Be OK?

1. ***Academically Adrift**→**G.B. Trudeau – Doonsberry-complete cartoon <http://bit.ly/mOweRZ>. [Arum & Roksa (2011)]

Arum & Roksa (2011, p. 21) wrote [bracketed by lines "A&R-A&R-A&R-. . . . ."; my ***italics***]:
A&R-A&R-A&R-A&R-A&R-A&R-A&R-A&R-A&R
With regard to the quality of research, we tend to evaluate faculty the way the Michelin guide evaluates restaurants," Lee Shulman, former president of the Carnegie Foundation for the Advancement of Teaching, recently noted. "We ask, 'How high is the quality of this cuisine relative to the genre of food? How excellent is it?' With regard to teaching, the evaluation is done more in the style of the Board of Health. The question is, 'Is it safe to eat here?'" Our research suggests that for many students currently enrolled in higher education, the answer is: not particularly. Growing numbers of students are sent to college at increasingly higher costs, but ***for a large proportion of them the gains in critical thinking, complex reasoning and written communication are either exceedingly small or empirically nonexistent***. At least 45 percent of students in our sample did not demonstrate any statistically significant improvement in CLA . . . .[[College Learning Assessment]]. . . . performance during the first two years of college. While these students may have developed subject-specific skills that were not tested for by the CLA, in terms of general analytical competencies assessed, ***large numbers of U.S. college students can be accurately described as academically adrift***. They might graduate, but they are failing to develop the higher-order cognitive skills that it is widely assumed college students should master. These findings are sobering and should be a cause for concern.
A&R-A&R-A&R-A&R-A&R-A&R-A&R-A&R-A&R

2. Academically Adrift College Students Can Degrade Pre-College Education

In "Should the Culture of University Science Education Be Changed ?" [Hake (2011b)] I wrote (paraphrasing):

> "Unfortunately, some of the students in U.S. universities are prospective K-12 teachers. They will have learned little from their university science/math courses and will teach as they have been taught thus leading to the general population's ignorance of science-related societal issues, election by them of leaders who are similarly ignorant, and consequent threat to life on Planet Earth."

____

***** For reviews see Louis Menand's (2011) "Live and Learn" in the *New Yorker*, and the discussion-list post "Academically Adrift?" [Hake (2011c). *Qualitative* evidence of higher education's failure to promote student learning that's consistent with the *quantitative* evidence of Arum & Roksa (2011) is marshaled in *Higher Education?: How Colleges Are Wasting Our Money and Failing Our Kids – and What We Can Do About It* [Hacker & Dreifus (2010)]. For similar criticism by former university presidents Derek Bok (Harvard), James Duderstadt (Univ. of Michigan), and Richard Cyert (Carnegie Mellon) see "Can Scientific Research Enhance the Art of Teaching?" [Hake (2007a)].

II. A Glimmer of Hope for Physics Education: The Concept Inventory*

A. First Physics Concept Inventory: *Mechanics Diagnostic* by:

Ibrahim Halloun & David Hestenes

1. Advantages of the Concept Inventory

As indicated in "Lessons from the Physics Education Reform Effort" [Hake (2002a):

"For more than three decades, physics education researchers have repeatedly shown that *traditional* introductory physics courses with passive student lectures, recipe labs, and algorithmic problem exams are of limited value in enhancing students' conceptual understanding of the subject [McDermott and Redish (1999)]. Unfortunately, this work was largely ignored by the physics and education communities until Halloun and Hestenes (1985a,b) devised the *Mechanics Diagnostic* (MD) test of conceptual understanding of Newtonian mechanics. Among many other virtues, the MD and the subsequent *Force Concept Inventory* (FCI) (Hestenes et al., 1992, Halloun et al., 1995) tests have two major advantages:

(a) the multiple-choice format facilitates relatively easy administration of the tests to thousands of students;

(b) the questions probe for a conceptual understanding of the basic concepts of Newtonian mechanics in a way that is understandable to the novice who has never taken a physics course (and thus can be given as an introductory course pretest), yet at the same time are rigorous enough for the initiate."

---

*Might Concept Inventories <http://en.wikipedia.org/wiki/Concept_inventory> also be a "Glimmer of Hope" for Higher Education generally? See, e.g., "The Physics Education Reform Effort: A Possible Model for Higher Education?" [Hake (2005e)] and the 1,610 Google hits on that title (as of 13 Sept 2011 12:30-0700) at <http://bit.ly/ocZg95>.

2. Pre/Post Test Results Shown in "The initial knowledge state of college physics students" [Halloun & Hestenes (1985a)]

ABSTRACT: *An instrument to assess the basic knowledge state of students taking a first course in physics* has been designed and validated. Measurements with the instrument show that the student's initial qualitative, common sense beliefs about motion and causes has a large effect on performance in physics, but *conventional instruction induces only a small change in those beliefs.*

Of considerably interest is this excerpt from Halloun & Hestenes (1985a):

"**Professor A** is a theoretical physicist; his lectures emphasize the conceptual structure of physics, with careful definitions and orderly logical arguments.

Professor B incorporates many demonstrations in his lectures, and he expends great time and energy preparing them; he strives especially to help students develop physical intuition.

Professor C emphasizes problem solving, and he teaches by example, solving one problem after another in his lectures.

Professor D is an experimental physicist teaching introductory physics for the first time; he followed the book closely in his lectures.

All four professors are known as good teachers according to informal peer opinion and formal evaluations by students. Indeed, *Professor B has twice received awards for outstanding teaching*. . . . [[My *italics* – see e.g., "SET's Are Not Valid Gauges of Teaching Performance" [Hake (2006,b,c,d,e.f)] and "SET's Are Not Valid Gauges of Students' Higher-Level Learning" [Hake (2011g)]. In Hake (2011g) I contest engineer Don Woods' (2011) contention that SET's are a "valid method of measuring teaching productivity."
Here SET's = "Student Evaluations of Teaching," usually the primary evidence in support of teaching awards.]]*. . . .

Now, Table I shows that ***the basic knowledge gain is the same for all four of the classes in University Physics.*** All four classes used the same textbook (Tipler), and covered the same chapters in it. Considering the wide differences in the teaching styles of the four professors, we conclude that the ***basic knowledge gain under conventional instruction is essentially independent of the professor.*** "

---

*Here and henceforth my inserts into quoted material are indicated by ". . . . . [[insert]]. . . . ."

3. Early Pre/post Test Results with Mechanics Diagnostic

a. "**Modeling instruction in mechanics**" [Halloun & Hestenes (1987)]
ABSTRACT: Modeling theory. . . . .[[Hestenes (1987, 1992), Halloun (2004)]]. . . . was used in the design of a method to teach problem solving in introductory mechanics. A pedagogical experiment . . . . .[[**pre/post testing with the *Mechanics Diagnostic***]]. . . . to evaluate the effectiveness of the method found positive results.

b. "**Promoting Student Crossover to the Newtonian World**" [Hake (1987)]
ABSTRACT: A six-week non-calculus-based introductory course was designed and taught with the intention of promoting students' conceptual understanding of Newtonian mechanics. Primary emphasis was placed on (1) laboratories structured to induce Socratic dialogue; (2) lectures stressing a qualitative approach to problem solving, and contrasting Newtonian and students' non-Newtonian concepts of motion; videotapes from *The Mechanical Universe* series. That the course was effective in promoting student crossover to the Newtonian world was suggested by student performance on **pre- and post-course mechanics exams . . . .[[the Mechanics Diagnostic test**]]. . . .

c. "**Peer Instruction: Ten years of experience and results**" [Crouch & Mazur (2001):

Catherine Crouch & Eric Mazur

ABSTRACT: We report data from ***ten years** of teaching*. . . . .[[1990-2000]]. . . *with Peer Instruction* (PI) in the calculus- and algebra-based introductory physics courses for non-majors; our results indicate increased student mastery of both conceptual reasoning and quantitative problem solving upon implementing PI. We also discuss ways we have improved our implementation of PI since introducing it in 1991. Most notably, we have replaced in-class reading quizzes with pre-class written responses to the reading, introduced a research-based mechanics textbook for portions of the course, and incorporated cooperative learning into the discussion sections as well as the lectures. These improvements are intended to help students learn more from pre-class reading and to increase student engagement in the discussion sections, and are accompanied by further increases in student understanding . . . .[[as shown **in 1990-91 by pre/post testing with the *Mechanics* Diagnostic** and thereafter with the FCI. For an excellent review of the Harvard work see the UTube video "Confessions of a Converted Lecturer" (Mazur, 2009)]]. . . .

B. *Force Concept Inventory* (FCI) (1992 & 1995) (successor to the *Mechanics Diagnostic*)

1. *Force Concept Inventory* [Hestenes, Wells, & Swackhamer (1992)]

David Hestenes   Malcolm Wells   Greg Swackhammer

2. The 1959 Factor Analysis Debate

a. Initiating the debate was "What Does the Force Concept Inventory Actually Measure?" [Huffman & Heller (1995)]. They opened with:

"Since publication of the Force Concept Inventory (FCI) in this journal in 1992 [Hestenes et al. (1992), the inventory has been widely used by high school, college, and university instructors across the country. The FCI itself is a 29-question . . . . [[30 in the 1995 version of Halloun et al. (1955)]]. . . . , multiple-choice test***** designed to assess students' Newtonian and non-Newtonian conceptions of force. In practice, it has provided an informative look at students' non-Newtonian views of physics. Numerous physics instructors have found the inventory a useful tool in helping to identify the concepts students understand and do not understand. . . . .The consistency of the results obtained with the FCI has been quite impressive. Comparable physics courses across the country have obtained very similar results using the inventory with both secondary and post-secondary students [Hestenes et al. (1992)]. . . . [[and more recently Hake (1998a,b).]]. . . . The authors have also gone to considerable length to conduct follow-up interviews with students in order to confirm the validity of responses to individual items. ***All of these findings lead to the general conclusion that the FCI is one of the most reliable and useful physics tests currently available for introductory physics teachers***. What, however, does the FCI actually measure?. . . . . "

____

*Can multiple choice tests measure higher-order learning? Psychometricians 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, Wells, & Swackhamer, 1992] . . . is an assessment that uses multiple-choice items to tap into higher-level cognitive processes."

Since Huffman and Heller (1995) state above that "*the FCI is one of the most reliable and useful physics tests currently available for introductory physics teachers,*" they must evidently regard it a reasonably good measure of course effectiveness, as is made clear by Heller & Huffman (1995) who wrote on p. 510:

> "We also agree with Hestenes and Halloun (1995) that there is a need for a 'nationally normed' test to evaluate the effectiveness of instruction in introductory physics courses. The FCI, which has reasonable face and content validity, is the best test currently available but we believe a more valid test from the point of view of the students' responses is needed."

Since no "more valid test from the point of view of the students' responses," has, as far as I know, ever been developed, the latter statement that the FCI is "the best test currently available [to evaluate the effectiveness of instruction in introductory physics courses]" is consistent with the conclusion of Hake (1998a,b) (Section IIC8) that:

> "The above conceptual (FCI) and problem solving test [Mechanics Baseline (MB)] results strongly suggest that **the use of IE strategies can increase mechanics-course effectiveness well beyond that obtained with traditional methods.**"

Thus that conclusion is unaffected by the outcome of the "1995 Factor Analysis Debate."

That debate was continued (with, as far as I know, no winner being proclaimed by independent experts) in:

b. "Interpreting the Force Concept Inventory: A Response to March 1995 Critique by Huffman and Heller" [Hestenes & Halloun (1995)];

c. "Interpreting the Force Concept Inventory: A Reply to Hestenes and Halloun" [Heller & Huffman (1995)]; and

d. "The Search for Conceptual Coherence in FCI data" [Halloun & Hestenes (1995, evidently unpublished)]

For discussions of "Factor Analysis" see e.g., Kim & Mueller (1978), Gorsuch (1983, 1990), Junker (1999), and Ding & Beichner (2009).

3. ***Force Concept Inventory (1995 Revision)*** [Halloun, Hake, Mosca, Hestenes (1995)]: **Currently available in 20 languages**: Arabic, Chinese, Croatian, Czech, English, Finnish, French, French (Canadian), German, Greek, Italian, Japanese, Malaysian, Persian, Portuguese, Russian, Spanish, Slovak, Swedish, & Turkish. This provides an opportunity for cross-national and cross-cultural studies of mechanics-course effectiveness, but that possibility, as far as I know, has not been exploited – please correct me if I'm wrong.

Since 1995 almost all education research involving the FCI has been conducted using the revised 1995 version. Unfortunately, literature reports of most of that research (other than by the Harvard group) *unscientifically* fail to indicate that the 1995 version of the FCI was used, referencing only the original version by Hestenes et al. (1992). As a result few teachers are aware of the 1995 version and continue to use the 1992 version to gauge the cognitive impact of their courses **:-(**. See e.g., "Re: Errata for FCI?" [Hake (2010h)].

C. Meta-Analysis of FCI/MD Results: “Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses” [Hake (1998a)]

1. ABSTRACT: A survey of pre/post test data using the Halloun-Hestenes *Mechanics Diagnostic* test or more recent *Force Concept Inventory* is reported for 62 introductory physics courses enrolling a total number of students N = 6542. A consistent analysis over diverse student populations in high schools, colleges, and universities is obtained if a rough measure of the average effectiveness of a course in promoting conceptual understanding is taken to be the average normalized gain <g>. The latter is defined as the ratio of the actual average gain (<%post> – <%pre>) to the maximum possible average gain (100 – <%pre>).

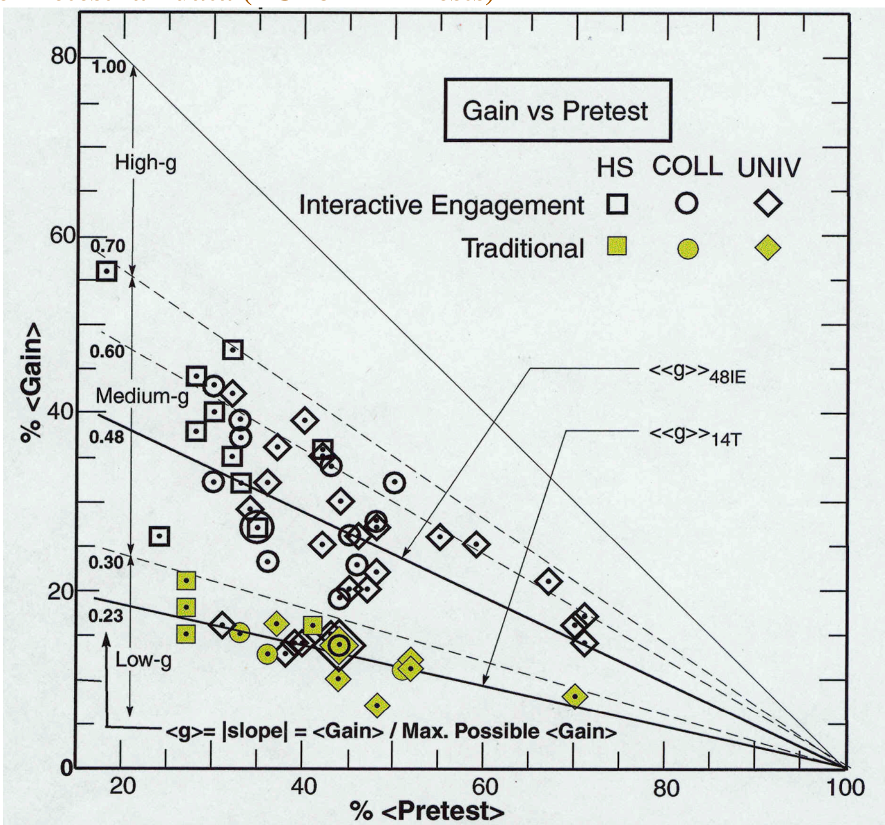
Fourteen “traditional” (T) courses (N = 2084) which made little or no use of interactive-engagement (IE) methods achieved an average gain $\langle\langle g \rangle\rangle_{14T} = 0.23 \pm 0.04$ (std dev). In sharp contrast, forty-eight courses (N = 4458) which made substantial use of IE methods achieved an average gain $\langle\langle g \rangle\rangle_{48IE} = 0.48 \pm 0.14$ (std dev), *almost two standard deviations of* $\langle\langle g \rangle\rangle_{IE}$ *above that of the traditional courses**.

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

---

* The Cohen Effect Size *d*, as calculated from the square root of the mean square deviations of the IE and T courses, is given as 2.43 in Eq. (9) of “Lessons from the Physics Education Reform Effort” [Hake (2002x)]. Nevertheless, Deslauriers, Schelew, & Wieman (2011) in “Improved Learning in a Large-Enrollment Physics Class” state: “The standard deviation calculated for both sections . . .[[in this study]]. . . was about 13%, giving an effect size for the difference between the two sections of 2.5 standard deviations. As reviewed by Froyd (2007) *other science and engineering classroom studies report effect sizes less than 1.0*. . . . .[[***not so***! – Froyd quoted the above abstract and discussed the similar two sd difference results of Buck & Wage (2005)]]. . . . An effect size of 2, obtained with trained personal tutors, is claimed to be the largest observed for any educational intervention [Bloom (1984)].”

2. Gain vs Pretest- all data (FCI or MD Tests)



Institution: UNIV = university, COLL = college, HS = high school; Method: IE = interactive engagement; T = traditional;

Symbols: <g> = course average *normalized* gain; max = maximum.

Fig. 1. %<Gain> versus %<Pretest> scores on the conceptual *Mechanics Diagnostic* (MD) or *Force Concept Inventory* (FCI) tests for 62 courses enrolling a total N = 6542 students: 14 traditional (T) courses (n = 2084), which made little or no use of interactive engagement (IE) methods, and 48 IE courses (n = 4458), which made considerable use of IE methods. Slope lines for the average of the 14 T courses $\langle\langle g\rangle\rangle_{14T}$ = 0.23 ± 0.04 (std dev) and the 48 IE courses $\langle\langle g\rangle\rangle_{48IE}$ = 0.48 ± 0.14 (std dev)are shown. The negative-slope straight lines are lines of constant, normalized average gain <g> = <Gain>/maximum possible <Gain> =(<%post > – <%pre >) / (100 – <%pre>). Thus, for example, if a class averaged 40% on the pretest and 60% on the posttest, then the class-average normalized gain <g> = (60% – 40%)/(100% – 40%) = 20%/60% = 0.33. (The random guessing score is 20%.)

3. Histogram of the Fraction of Courses vs <g>

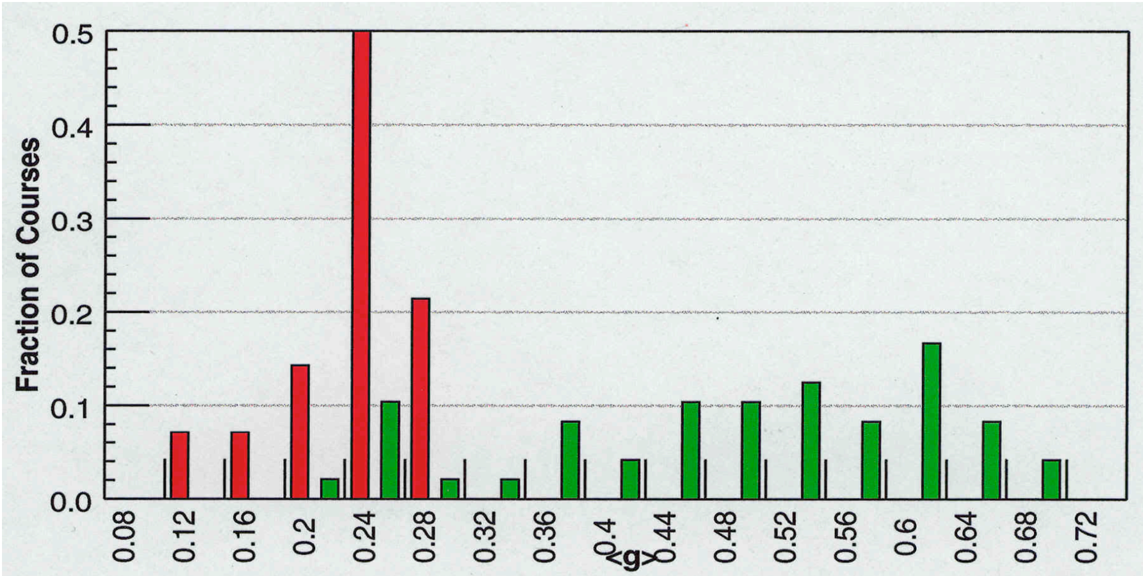


Fig. 2. Average normalized gain <g>: red bars show the fraction of 14 traditional courses (N = 2084), and green bars show the fraction of 48 interactive engagement courses (N = 4458), both within bins of width δ<g> = 0.04 centered on the <g> values shown.

4. Operational Definitions* of Interactive Engagement (IE) and Traditional (T) Courses

a. IE courses are defined *operationally* as those designed at least in part to promote conceptual understanding through the active engagement of students in heads-on (always) and hands-on (usually) activities that yield immediate feedback† through discussion with peers and/or instructors;

b. T courses are defined *operationally* as those reported by instructors to make little or no use of IE methods, relying primarily on passive-student lectures, recipe laboratories, and algorithmic problem examinations.

---

*The importance of *operational* definitions in educational research has been emphasized by Klahr & Li (2005) and Hake (2008d, 2010d, 2010e, 2011b).

†Thus a hallmark of IE courses is their use of "formative assessment" in the sense used by Black & Wiliam (1998) and Shavelson (2008): "All those activities undertaken by teachers -- and by their students in assessing themselves -- *that provide information to be used as feedback to modify teaching and learning activities*." Note that this meaning of "formative assessment" is different from "formative evaluation" as defined by JCSEE (1994): "Formative evaluation is evaluation designed and used to improve an object, especially when it is still being developed."

5. Systematic Errors Are Probably Insignificant [For a detailed treatment of systematic and random errors see Hake (1998a)]

a. Question Ambiguities and Isolated False Positives

Very few due to the careful work of the ASU group.

b. Teaching to the Test and Test-question Leakage

No evidence of this.

c. Fraction *f* of Course Time Spent on Mechanics

No apparent correlation of <g> with f for $0.7 \leq f \leq 1$ *as discussed in Hake (1998a).*

d. Post and Pretest Motivation of Students

No apparent correlation of <g> with amount of grade credit for post-test FCI score – see e.g. Henderson (2002) – nevertheless I recommend giving grade credit for the posttest FCI – see Section IIE1e: "Giving course credit probably motivates students to take the posttest more seriously and thereby demonstrate more adequately their understanding. . . . . If no grade credit is given for performance on the posttest then selective removal of 'no-effort' tests [see e.g., Henderson (2002), Mallinckrodt (2001)] by different investigators with different no-effort criteria will lead to uncertainty in comparing normalized gains of different courses."

e. Hawthorne/John Henry Effects*

Neither is likely to have affected the conclusions of Hake (1998a) as discussed in that paper.

---

* See e.g., *Research Methods in Education* [Slavin (1992)]

6. Posttest Scores: Mechanics Baseline vs FCI

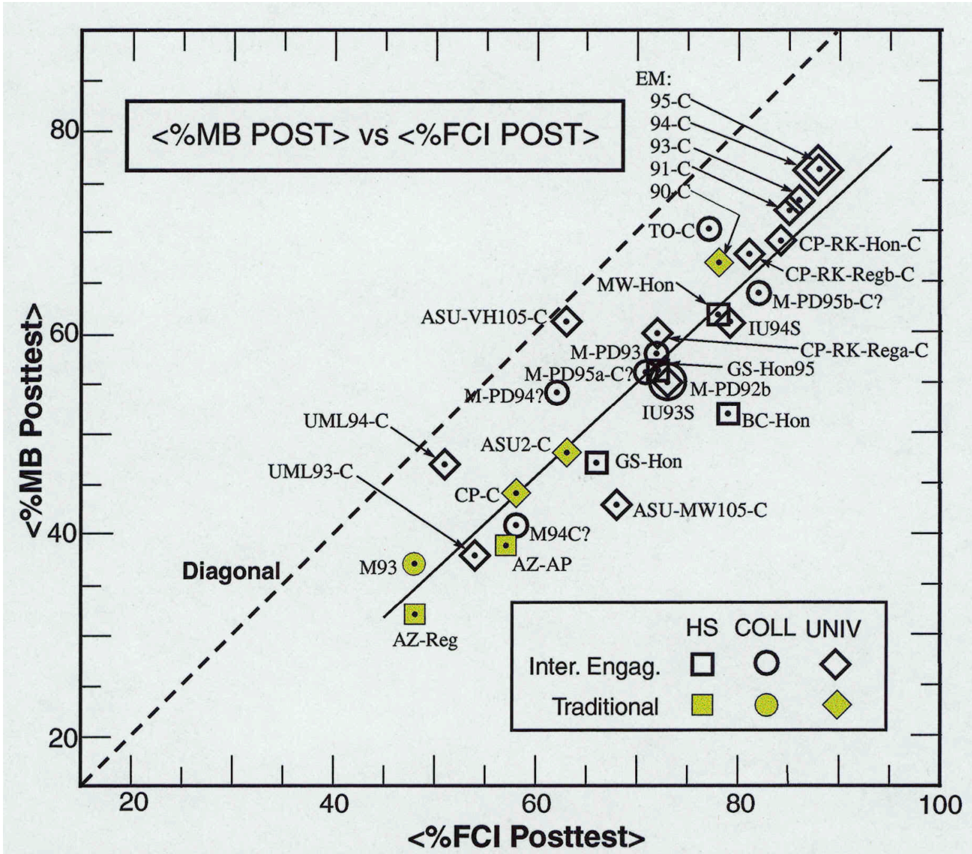


Fig. 3. Post-course MB (problem-solving) vs FCI (conceptual understanding) test scores for the 30 high school, college, and university courses (N = 3259) for which both sets of data were available. The solid line is a least squares fit. The correlation coefficient is r = + 0.91.

7. A Crucial Companion Paper to Hake (1998a)

"Interactive-engagement methods in introductory mechanics courses" [Hake (1998b):

Average pre/post test scores, standard deviations, instructional methods, materials used, institutions, and instructors for each of the survey courses of are tabulated and referenced. In addition the paper includes:

(a) case histories for the seven IE courses of Hake (1998a,b) whose effectiveness as gauged by pre-to-post test gains was close to those of T courses,

(b) advice for implementing IE methods, and

(c) suggestions for further research.

Submitted on 6/19/98 to the "Physics Education Research Supplement" (PERS) of the *American Journal of Physics*, but ***rejected*** **: – (** by its editor on the grounds that the very transparent *Physical Review* type data tables [see e.g., Table II of Hake (1967)] were "impenetrable"!

8. Summary And Conclusions

a. Fourteen T courses (N = 2084) which made little or no use of IE methods achieved an average normalized gain $<<g>>_{14T} = 0.23 \pm 0.04$sd. In sharp contrast, 48 IE courses (N = 4458) that made substantial use of IE methods achieved an average gain $<<g>>_{48IE} = 0.48 \pm 0.14$sd. It is extremely unlikely that systematic errors play a significant role in the nearly two-standard deviation difference of the T and IE courses.

b. A plot of average course scores on the problem-solving MB test vs those on the conceptual FCI show a strong positive correlation r = + 0.91. Comparison of IE and traditional courses implies that IE methods **enhance** problem-solving ability.

c. The above conceptual and problem solving test results strongly suggest *that* ***the use of IE strategies can increase mechanics-course effectiveness well beyond that obtained with traditional methods.***

9. Research Consistent With the Above Meta-analysis

As of 2006, average normalized gain differences between T and IE courses that are consistent with the work of Hake (1998a,b) and Figure 1 had been reported in at least 25 other research papers, as given and referenced in "Design-Based Research in Physics Education Research: A Review" [Hake (2008a, p.12)]:

Redish, Saul, & Steinberg, 1997;
Saul, 1998;
Francis, Adams, & Noonan, 1998;
Heller, 1999;
Redish & Steinberg, 1999;
Redish, 1999;
Beichner et al., 1999;
Cummings, Marx, Thornton, & Kuhl, 1999;
Novak, Patterson, Gavrin, & Christian, 1999;
Bernhard, 2000;
Crouch & Mazur, 2001;
Johnson, 2001;
Meltzer, 2002a, 2002b;
Meltzer & Manivannan, 2002;
Savinainen & Scott, 2002a, 2002b);
Steinberg & Donnelly, 2002;
Fagan, Crouch, & Mazur, 2002;
Van Domelen & Van Heuvelen, 2002;
Belcher, 2003;
Dori & Belcher, 2004;
Hoellwarth, Moelter, & Knight, 2005;
Lorenzo, Crouch, & Mazur, 2006; &
Rosenberg, Lorenzo, & Mazur, 2006.

As indicated in Hake (2008b), this consistency of the results of many investigators in various institutions working with different student populations with the results of Hake (1998a,b)) constitutes the most important single warrant for the validity of conclusion in Hake (1998a) that: "*The conceptual and problem-solving test results strongly suggest that the classroom use of IE methods can increase mechanics-course effectiveness well beyond that obtained in traditional practice*."

Such gradual buildup of an agreed-upon "community map" (Redish, 1999; Ziman, 2000) is characteristic of the progress of traditional science, but it seems to be consistently undervalued in educational research.

Furthermore, that interactive engagement courses would be more effective in enhancing conceptual understanding of counterintuitive Newtonian laws than traditional courses with their passive-student lectures, recipe laboratories, and algorithmic problem sets certainly would be expected from (a) previous physics education research (McDermott & Redish, 1999), including the astute ethnographically-based insights of the late Arnold Arons (1997) - see "The Arons Advocated Method" [Hake (2004a)]; and (b) neuroscience – see pp. 13-14 of "Should We Measure Change? Yes!" [Hake (2011b) for a discussion.

10. Do the Above Results Have Anything To Do With Engineering Education?

Note the similarity of Fig. 1 with Fig. 4, shown below, due to Buck and Wage (2005) for the *Signals and Systems Concept Inventory* (SSCI). The SSCI is a 25- question, multiple-choice exam designed to be given in one hour, and is available in both discrete-time (DT) and continuous-time (CT) form.

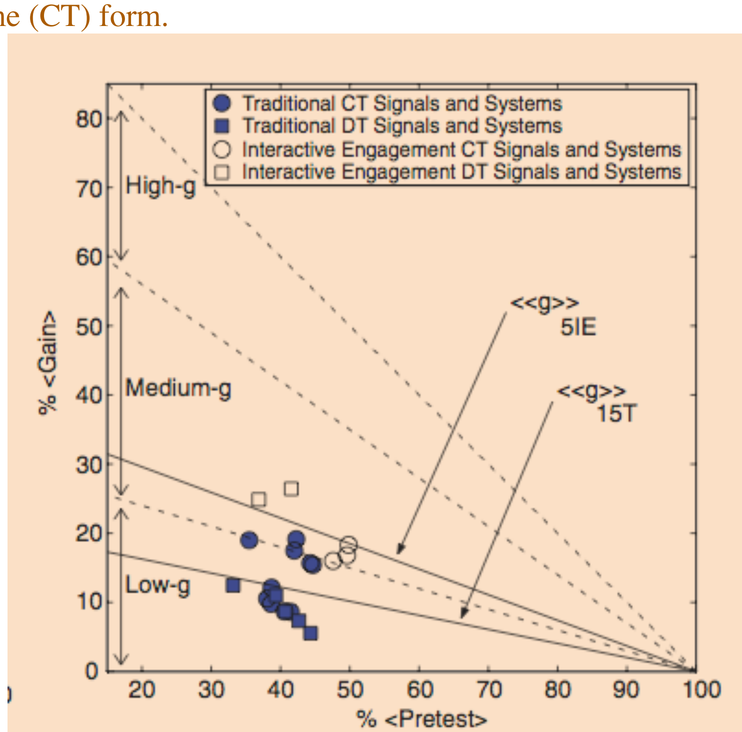


Fig. 4. The 15 lecture format courses (shaded points) had an average of the average normalized gains <<g>>**15 T** = 0.20 ± 0.07, while the five active and cooperative learning (ACL) courses (open points) achieved <<g>>**5 IE** = 0.37 ± 0.06, more than two standard deviations above the lecture courses.. . . .[[**bold** text indicates my additions]]. . .

11. Does the Normalized Gain on the FCI/MD Tell All?

Does the class average normalized gain <g> for the FCI or MD, provide a definitive assessment of the *overall* effectiveness of an introductory physics class? ***No!*** It assesses "*only the attainment of a **minimal** conceptual understanding of mechanics* (Hestenes, 1997)]. [In some first-semester or first- quarter introductory physics courses, subjects other than mechanics are often covered. The effectiveness of the course in promoting student understanding of those topics would not, of course, be assessed by the average normalized gain on the FCI or MD.] Furthermore, as indicated in my AJP rejected **:-(**, and therefore almost universally ignored, "Interactive-engagement methods in introductory mechanics courses" [Hake (1998b)] I wrote:

"Among desirable outcomes of the introductory course that the [Mechanics Diagnostic and FCI] tests DO NOT MEASURE DIRECTLY are e.g., students':

(a) satisfaction with and interest in physics;

(b) understanding of the nature, methods, and limitations of science;

(c) understanding of the processes of scientific inquiry such as experimental design, control of variables, dimensional analysis, order-of-magnitude estimation, thought experiments, hypothetical reasoning, graphing, and error analysis;

(d) ability to articulate their knowledge and learning processes;

(e) ability to collaborate and work in groups;

(f) communication skills;

(g) ability to solve real-world problems;

(h) understanding of the history of science and the relationship of science to society and other disciplines;

(i) understanding of, or at least appreciation for, "modern" physics;

(j) ability to participate in authentic research."

12. *Empirical* Justification of <g> as a Comparative Measure of Course Effectiveness

[not to be confused with the average *amount of learning* in a course as measured by the average actual gain <G> = <%post> - <%post>].

a. In Hake (1998a,b) a consistent analysis – see Fig. 1 and 2 above - over diverse student populations in high schools, colleges, and universities was obtained by using <g> as a rough measure of the average effectiveness of a course in promoting conceptual understanding of Newtonian mechanics. A major advantage of <g> was:

(1) The correlation of the average normalized <g> with <%pre> for the 62 survey courses was a very low + 0.02.*

(2) In contrast, the average posttest score %<post > and the average *actual* gain %<G> are less suitable for comparing course effectiveness over diverse groups since the correlation of:

<%post> with <%pre> is + 0.55, and
<%G> with <%pre> is – 0.49,

as is reasonable. Note that in the absence of instruction, a high positive correlation of <%post > with <%pre> would be expected.

b. A similar consistent analysis using <g> is obtained in over 25 physics education research papers as listed above in Section IIC9 "Research Consistent With the Above Meta-analysis"

---

*A relatively low correlation of <g> with <%pre> may not always be the case. The abstract of Coletta & Phillips (2005) reads (with inserts to distinguish the single student normalized again g_i from the class average normalized gain <g>: "We examined . . . . [[*single student*]]. . . normalized gains and preinstruction scores on the force concept inventory FCI for students in interactive engagement courses in introductory mechanics at four universities and found a significant, positive correlation for three of them.. . . .[[For 670 students in 4 classes at Harvard no such correlation was found. Similarly no such correlation was found by: (a) Cummings et al. (1999; p. S44) for 347 students at Rensselaer Polytechnic Institute given the FCI; (b) Meltzer (2002, Table I) for 219 Southeastern Louisiana University students given the Conceptual Survey of Electricity (CSE) test derived from the Conceptual Survey of Electricity and Magnetism (CSEM) test of Maloney et al. (2001); and (c) Hestenes (2005) for 12,000 high-school students who took the FCI test - this according to Coletta et al. (2007a)]]. . . . . .We also examined class average FCI scores of 2948 students in 38 interactive engagement classes, 31 of which were from the same four universities and 7 of which came from 3 other schools. We found a significant, positive correlation. . . . . .[[+0.63]]. . . between class average normalized FCI gains. . . .[[Unfortunately, Coletta & Phillips (2005) do not indicate *how* they calculated "class average normalized gains," but I shall *assume* they used the same method as in Hake (1998a) for which "class average normalized gains" = <g>]]. . . . and class average preinstruction scores. . . [[<%pre>]]. . . . To probe this correlation, we administered Lawson's. . . .[[Lawson (1978, 1992, 1995); Hake (2009c)]]. . . . . classroom test of scientific reasoning to 65 students and found a significant, positive correlation. . . . [[r =+ 0.51]]. . . . between these students' normalized FCI gains and their Lawson test scores. This correlation is even stronger than the correlation between. . . .[single student]]. . . FCI gains and preinstruction FCI scores. Our study demonstrates that differences in student populations are important when comparing normalized gains in different interactive engagement classes. We suggest using the Lawson test along with the FCI to measure the effectiveness of alternative interactive engagement strategies."

D. Lessons From The Physics Education Reform Effort*

1. Six Lessons On Interactive Engagement

(1) The use of IE strategies can increase the effectiveness of conceptually difficult courses well beyond that obtained by traditional methods.

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

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

(4) Education Research and Development (R&D) by disciplinary experts (DE's), and of the same quality and nature as traditional science/engineering R&D, is needed to develop potentially effective educational methods within each discipline. But the DE's should take advantage of the insights of

(a) DE's doing education R&D in other disciplines,
(b) cognitive scientists,
(c) faculty and graduates of education schools, and
(d) classroom teachers.

(5) The development of effective educational methods within each discipline requires a redesign process of continuous long-term classroom use, feedback, assessment, research analysis, and revision.

(6) Although non-traditional IE methods appear to be much more effective than T methods, *there is need for more research to develop better strategies for enhancing student learning*.

2. Eight Lessons On Implementation

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

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

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

(10) A major problem for undergraduate education in the United States is the inadequate preparation of incoming students, in part due to the inadequate university education of K–12 teachers.

(11) Interdisciplinary cooperation of instructors, departments, institutions, and professional organizations is required for synthesis, integration, and *change in the entire chaotic educational* ***system*** . . . . . .[[(a) *perhaps only engineers*, the "master integrators" who "get things done" *are up to that monumental task* **-** see the Conclusion of this talk, Section VD7; (b) for a compilation of references on "systems thinking" see "Over Two-Hundred Annotated References on Systems Thinking" [Hake (2009a). For related posts go to "Hake'sEdStuff" at <http://HakesEdStuff.blogspot.com/>, and then scroll down to and click on "Systems Thinking" under "Labels" in the right-hand column.

____

* Derived from "Lessons from the Physics Education Reform Effort" [Hake (2002a)] and "Six Lessons From the Physics Education Reform Effort" [Hake (2007c)].

(12) Various institutional and political factors, including the culture of research universities slow educational reform . . . .[[see "Changing the Culture of Science Education at Research Universities (Anderson et al. (2011) and "Should the Culture of University Science Education Be Changed" (Hake, 2011f)]]. . . . . .

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

(14) *"Education is not rocket science, it's much harder*."

George (Pinky) Nelson, astronaut, astrophysicist, and former director of the AAAS Project 2061, as quoted by E.F. (Joe) Redish (1999).

E. Suggestions for the Administration and Reporting of Concept Inventories*

I list below some administrative and reporting suggestions that might assist instructors in using *formative*‡ diagnostic testing. These suggestions reflect the hard lessons I have learned in the pre/post testing of 1263 pre-med introductory-physics-course students at Indiana University and the compilation of a 6542 student (62-course) survey (Hake 1998a, b) of pre/post test results. Although they are undoubtedly *not* appropriate for all classroom situations, they may at least indicate some of the problems that should be anticipated. A dagger† preceding a suggestion indicates that the suggestion is intended, at least in part, to promote confidentiality§ of the test.

1. Administration of Diagnostic Tests (DT's)†

†a. When administering DT's to students, refer to the tests by home-made generic titles rather than the specific titles designated by the authors (e.g.,"Mechanics Familiarity Survey" rather than *Force Concept Inventory* or *Force Motion Concept Evaluation*).

b. To enable meaningful pre/post comparison, maximum time intervals ΔT given to students to complete the pretest and the posttest should be same. To facilitate more meaningful meta-analysis, maximum time interval ΔT should be specified by the test designers and that interval should be rigidly enforced by all examiners.

---

* Derived from "Assessment of Physics Teaching Methods" [Hake (2002b)].

‡Here *formative* is used in the sense defined by the "Joint Committee on Standards for Educational Evaluation" [JCSEE (1994)]: "Formative evaluation is evaluation designed and used to improve an object, especially when it is still being developed." If pre/post testing were to be used for high-stakes summative purposes such as rating instructors, then Campbell's (1976) and Dunkenfield's Laws [see Kleiman (2010)] would probably rear their ugly heads so as to distort and corrupt the testing – see Hake (2010f,g) for a discussion. These two laws are:
Dukenfield's Law: If a thing is worth winning, it's worth cheating for.
Campbell's Law: 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. . . . . . .

§ See "Why Password Protection of Concept Tests is Critical: A Response to Klymkowsky" [Hake (2010b)].

†The sensitivity of concept inventory test results to administrative procedures has been emphasized by Ding, Reay, Lee, & Bao (2008), evidently either unaware or dismissive of Hake (2002b).

c. Do *not* allow students to take either the pretest or the posttest anonymously, because non-anonymity allows:

(1) Proper incentive for students to exert effort on the test.

(2) Analysis of "matched" pre/post test data, i.e., obtaining the average class pretest score by counting only the scores of those students who took the posttest, and thus allowing a more rigorous calculation of the class average normalized gain <g>. [For a discussion of "matched data" see Hake (2002a); and also Hake (1998b), Table I, footnote "c" on page 7.]

(3) Knowledge of the normalized gain g for each single student in the class, thus allowing a calculation of the average of the single-student gains: "g-ave." [See Hake (1998a - Sec. V and also footnote #46), for a discussion of systematic and random errors in pre/post testing and the connection between low correlation of single students g's with their pretest scores, and the small difference between values of g-ave and <g>.]

(4) Analyses of single student normalized or actual gains in terms of single-student characteristics or performance on other tests.

(5) Calculation of the correlation of individual student g's with their pretest scores.

†d. If possible, give the pretest on the *first* day of class. Take great care that all question sheets and answer sheets are returned and verify such return by counting those given out and those returned. *In order to promote serious effort on the pretest by students, explain that although their scores on the pretest will not count towards the course grade, their scores will be confidentially returned to them and will assist both themselves and their instructors to know the degree and type of effort required for them to understand the class topic.*

†e. Give the posttest *unannounced* near the final day of classes, and preferably as part of the final exam with significant course credit given for posttest performance. Giving course credit probably motivates students to take the posttest more seriously and thereby demonstrate more adequately their understanding, especially if time devoted to the posttest subtracts from time spent on the rest of the final exam. *If no grade credit is given for performance on the posttest then* selective removal of "no-effort" tests [see e.g., Henderson (2002), Mallinckrodt (2001)] by different investigators with different no-effort criteria will lead to uncertainty in comparing normalized gains of different courses.

†f. Do *not* return DT's to students after either the pretest or the posttest.

†g. Post DT scores by ID without posting or disseminating questions or answers.

†h. Avoid in-class discussion of questions identical or almost identical to DT questions (an example of "teaching to the test").

†i. For the posttest, announce that instructors be willing to discuss the questions and/or problems only *privately* with students.

†j. *Do not make DT questions or problems available on the web unless they are password protected* such that only authorized instructors may gain access. *Do not publish DT's in the open literature,* as has been the common practice. Carefully constructed DT's are international assets whose confidentiality should be as well protected as the MCAT (The U.S. Medical College Admission Test).

†k. because of the almost unavoidable slow diffusion of test questions and answers to student files, *replace each DT at approximately 5- or 10-year intervals*, such that it can be meaningfully calibrated against the previous test(s). [So far this has NOT been done for the now overused 1992/95 versions of the FCI; in my opinion, *as time goes on, research results based on the 1992/95 FCI will become more and more doubtful*.]

2. Reporting of Diagnostic Tests (DT's)

a. Report at least the class average:

(1) <%pre> *with its standard deviation* (sd),

(2) <%post> *with its sd*, and

(3) average normalized gain <g>.

ASIDES:

(a) Unless standard deviations are reported, the effect size and errors in <g> cannot be ascertained.

(b) As a statistic for comparison of courses and for meta-analyses, the class average <g> is better, in my opinion, than g-ave (the average of the single-student normalized gains) because the latter:

(1) must exclude students who score 100% on the pretest and thus achieve an infinite or indeterminate g; and

(2) may introduce skewing due to outliers who score near 100% on the pretest and less on the posttest such their <g>'s are large and negative. The selective removal of outliers so as to avoid outliers by various different investigators with different outlier criteria will lead to a degree of uncertainty in comparing normalized gains of different courses.

b. Report if possible:

(1) *Cohen's "effect size" d* (Cohen 1988)

ASIDE: The effect size is commonly used in meta-analyses, and strongly recommended by many critics as a preferred alternative (or at least addition) to the *usually inappropriate* t-tests and *p* values associated with null-hypothesis testing. [See Hake (2002a) for references to the effect size and the extensive anti-p literature.] But as indicated in Hake (2002a): ". . . . it should be noted that *for pre/post comparisons of course effectiveness over diverse student populations with widely varying average pretest scores, d is a relatively poor metric* because unlike <g> :

(a) *d* depends on the actual bare (unrenormalized) gain that tends to be negatively correlated with the <pre>;

(b) <Gain>'s are confounded with sd's: given two classes both with identical *statistically* significant <Gain>'s, the more homogeneous class with the smaller rms pre/post sd's will be awarded the higher *d.*

(2) The *Kuder-Richardson reliability coefficients* KR-20 (or equivalent) - for tests in which the answers are either correct or incorrect as in most Concept Inventories. If that's not the case then *Cronbach's alpha* can be utilized. [See e.g., Beichner (1994), Slavin (1992), Allen et al. (2008), and the Wikipedia entry <http://bit.ly/qkWWdc> for discussions.

(3) The *estimated systematic and random errors*. [See e.g., Hake (1998a).]

(4) The *correlation of individual student g's and pretest scores*.
ASIDE: A significant *positive* correlation would suggest that the instruction tends to favor students who have *more* prior knowledge of the subject as judged by the pretest score ("Matthew effect"*); a significant *negative* correlation would suggest that the instruction favors students judged by the pretest score ("anti–Matthew effect"); and an insignificant correlation would suggest that the instruction is at about the right level for students who have an average prior knowledge of the subject as judged by the pretest score.

---

*Matthew, *First Gospel of the New Testament* (Gutenberg edition) "to him that hath shall be given, but from him that hath not shall be taken away even that which he hath."

F. Concept Inventory Listings Including Physics and Engineering

1. Field-tested Learning Assessment Guide [FLAG (2011)]

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 agriculture, astronomy & physics, biology, chemistry, computer science, **engineering**, general science, interdisciplinary courses, and mathematics are online at <http://www.flaguide.org/tools/tools.php>.

2. NCSU Assessment Instrument Page [NCSU (2011)]

Online at line at <http://bit.ly/9gfUpY>. North Carolina State University Physics Education R & D Group. Inventories on many different aspects of various disciplines including physics and Engineering

3. Engineering-Related Concept Inventories

See Table 4 of "Concept Inventories in Engineering Education" [Reed-Rhoads & Imbrie (2008)] for a listing of 21 Engineering Related Concept Inventories. See also the new Concept Inventory HUB at <http://ciHUB.org/>, especially <http://bit.ly/qeEtTN>.

4. Concept Inventories in Higher Education Science [Libarkin (2008)]

Table 1 of Libarkin (2008) gives a "Comprehensive list of published concept inventories in science." Table 2. Gives a "List of unpublished science concept inventories under development." Among the CI's under development but not listed in Libarkin's Table 2 are

"Setting the Scope of Concept Inventories for Introductory Computing Subjects" [Goldman, et al. (2010)]; and

"First Results From The Science Literacy Concept Inventory: The Reasoning We Don't Produce Through Gen-Ed" [Nuhfer (2011)].

III. For Physics Education the Road to Reform Has Been All Uphill!

A. Slow Diffusion of Innovations

1. *Diffusion of Innovations* [Rogers (2003)]

Rogers begins Chapter 1 "Elements of Diffusion" by quoting Niccolo Machiavelli (1513):

> "There is nothing more difficult to plan, more doubtful of success, nor more dangerous to manage than the creation of a new order of things.... Whenever his enemies have the ability to attack the innovator they do so with the passion of partisans, while the other defend him sluggishly, so that the innovator and his party alike are vulnerable."

And then writes:

> "Getting a new idea adopted, even when it has obvious advantages, is difficult. Many innovations require a lengthy period of many years from the time they become available to the time when they become widely adopted. Therefore a common problem for many individuals and organizations is how to speed up the rate of diffusion of an innovation."

2. Inertia of the Educational System

Lesson #13 of Hake (2002a) is

> "The monumental inertia of the educational system may thwart long-term national reform. The glacial inertia of U.S. educational system is not well understood. An issue of *Daedalus* (1998) contains essays by researchers in education and by historians of more rapidly developing institutions such as power systems, communications, health care, 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?' As far as I know, no definitive answer has yet been forthcoming."

In "Should We Measure Change? Yes!" [Hake (2011b, pp. 22-24) I list the following "Eleven Barriers to Change in Higher Education" (the first four of which were suggested by Sheila Tobias (2000) in her "Guest comment: From innovation to change: Forging a physics education agenda for the 21st century":

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.

5. Universities fail to think of education in terms of student learning rather than the delivery of instruction [Barr and Tagg (1995), Tagg (2003)] - for a discussion see Hake (2005e).

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."

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 (2002e)]; "SET's Are Not Valid Gauges of Teaching Performance" [Hake 2006 (a,b,c,d,e]; "SET's Are Not Valid Gauges of Students' Higher-Level Learning" [Hake (2011g)].

8. The provincialism of educational researchers in various disciplines, as discussed in: (a) Section II, "The Insularity of Educational Research" of an article "Design-Based Research: A Primer for Physics Education Researchers" [Hake (2004c)]; (b) "Cognitive Science and Physics Education Research: What We've Got Here Is a Failure to Communicate" [Hake (2007b)]; and (c) "Language Ambiguities in Education Research" [Hake (2008d).

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."

10. The schizophrenia of academic scientists (counterpart to the pre/post paranoia of PEP's – see barrier #11 below) 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 ("Should We Measure Change? Yes"):

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

3. Eleven Quotes in Honor of Inertia [Hake (2006a)]

In addition to Rogers' Machiavelli quote above, these two from Hake (2006a) seem especially relevant to the inertia of higher education:

> "Langdell's innovations. . . .[[case studies in the Harvard Law School, circa late 1800's]]. . . . initially met with enormous resistance. Many students were outraged. During the first three years of his administration, as word spread of Harvard's new approach to legal education, enrollment at the school dropped from 165 to 117 students, leading Boston University to start a law school of its own. Alumni were in open revolt."
>
> David Gavrin (2003)

> "Difficulties of Change: . . . 9. 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 MacVicart, W.T. Martin, Edwin Taylor, and Jerrold Zacharias (1977).

More recently I became aware of another quote that's especially germane to the inertia of higher education:

> "If you try to introduce people to a paradigm shift, they will hear what you have to say and then interpret your words in terms of their old paradigm. What does not fit, they will not hear."
>
> Myron Tribus (2001), former director of the Center for Advanced Engineering Study at MIT – see <http://en.wikipedia.org/wiki/Myron_Tribus>

4. Dancy & Henderson's (2010)] Self-Report Survey Suggests That Reform Methods Haven't Led To Fundamental Changes in Instruction

Dancy & Henderson wrote: ". . . . self-reports. . . .[[according to Henderson & Dancy (2009), of 722 physics faculty across the United States with a 50.3% response rate]]. . . . . of actual classroom practices indicate that the availability of . . .[physics education research curricula and pedagogies]. . . *has not led to fundamental changes in instruction*. . . [suggesting]. . . a need for research-based dissemination that accounts for the complexity of instructional change."

Henderson and his co-workers, practicing what they preach, have been researching the "complexity of instructional change," see e.g.:

a. "Promoting instructional change in new faculty: An evaluation of the physics and astronomy new faculty workshop" [Henderson (2007)];

b. "Barriers to the Use of Research-Based Instructional Strategies: The Influence of Both Individual and Situational Characteristics" [Henderson & Dancy (2007)];

c. "Physics Faculty and Educational Researchers: Divergent Expectations as Barriers to the Diffusion of Innovations" [Henderson & Dancy (2008)];

d. "Improving 'Educational Change Agents' Efficacy in Science, Engineering, and Mathematics Education" [Froyd, Beach, Henderson, & Finkelstein (2008)];

e. "Impact of physics education research on the teaching of introductory quantitative physics in the United States" [Henderson & Dancy (2009)];

f. "Beyond Dissemination in College science teaching: An Introduction to Four Core Change Strategies" [Henderson, Finkelstein, & Beach (2010)];

g. "Facilitating Change in Undergraduate STEM Instructional Practices: An Analytic Review of the Literature" [Henderson, Beach, & Finkelstein (2011)]; and

h. "Four Categories of Change Strategies for Transforming Undergraduate Instruction," [Henderson, Beach, & Finkelstein (2011)]

i. "The Use of Research-Based Instructional Strategies in Introductory Physics: Where do Faculty Leave the Innovation-Decision Process?" [Henderson, Dancy, & Niewiadomska-Bugaj (submitted)].

But the six "early adopters" of Section IV "Impact of Physics Concept Inventories on Undergraduate Physics Education" – see below - show that fundamental changes in instruction *have* occurred even though they're not reflected in the survey of Dancy & Henderson (2010). Fig. 1-2, p. 12, of Rogers (2003) shows that curves of "percent of adopters" a vs time t typically rise in an S shaped curve. It would appear that introductory physics education reform in higher education† is now near the bottom of the S with a relatively small slope da/dt. Judging from Rogers' curves, now that there have been a few "early adopters," da/dt for introductory physics education reform in higher education *might* be expected to increase rapidly into a "take off" region within a few years.

5. Research on Effective Practices vs Research on Implementing Change

Consistent with the claim of Dancy and Henderson (2010) that " the availability of . . .[physics education research curricula and pedagogies]. . . *has not led to fundamental changes in instruction*. . . [suggesting]. . . a need for research-based dissemination that accounts for the complexity of instructional change," Fairweather (2008), who, according to Labov et al. (2009), "was asked to review and synthesize . . . . the articles submitted for the October workshop". . . .[[see National Academies (2008), and NRC (2011a)]]. . . .." wrote:

> "NSF- and association-funded reforms at the classroom level, however well intentioned, have not led to the hoped for magnitude of change in student learning, retention in the major, and the like *in spite of empirical evidence of effectiveness* [*italics* in the original]. Among the most important elements of a successful change strategy to promote the improvement of undergraduate STEM education. . . .[is recognizing]. . . that *more effort needs to be expended on strategies to promote the adoption and implementation of STEM reforms rather than on assessing the outcomes of these reforms* [my *italics*]. Additional research can be useful but the problem in STEM education lies less in not knowing what works and more in getting people to use proven techniques."

---

† Physics education using the "modeling method" in high schools is evidently diffusing more rapidly than IE methods in higher education, evidently due in part to university-based graduate programs dedicated to professional development of in-service physics teachers, as described by Hestenes et al. (2011). It seems unlikely that in-service college and university faculty would be willing to enroll in similar graduate programs.

B. Opposition of Psychologists, Education Specialists, & Psychometricians (PEP's)

1. Pre/Post Paranoia

In "Should We Measure Change? Yes!" [Hake (2011b), I wrote:

"Formative pre/post testing is being successfully employed to improve the effectiveness of courses in undergraduate astronomy, economics, engineering, biology, chemistry, economics, geoscience, 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. . . . . . . Cronbach & Furby (1970) wrote:

'*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.*'

This dour appraisal of pre/post testing has echoed down though the literature to present day texts on assessment such as that by Suskie (2004a). . . .[[for an antidote to Suskie(2004a) see Maki (2004)]]. . . . . 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 (2005e)], 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 . . . . . .[[see Hake (2011b) for the following references]]. . . . . 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."

2. The Randomized Control Trial "Gold Standard"

In "Will the No Child Left Behind Act Promote Direct Instruction of Science?" [Hake (2005a)], I gave, as one of the seven reasons why Direct Science Instruction threatens to predominate nationally under the aegis of the "No Child Left Behind Act," the following [bracketed by lines "HHHHHH. . . .",]:

HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Most interactive engagement and guided inquiry methods have not been tested in randomized control trials (RCT's), the "gold standard" of the U.S. Dept. of Education (USDE).

That a single research method should be designated as the "gold standard" for evaluating an intervention's effectiveness appears antithetical to the report of the NRC's Committee on Scientific Principles for Education Research [Shavelson & Towne (2002) - ST]. ST state that scientific research should "pose significant questions that can be investigated empirically," and "use methods that permit direct investigation of the questions."

Furthermore, the USDE's RCT gold standard is considered problematic by a wide array of scholars. Taking issue with the RTC gold standard are [see that article for references other than Zaritsky, Kelly, Flowers, Rogers, Patrick (2003), and Scriven (2008)]: philosophers Dennis Phillips [Shavelson, Phillips, Towne, & Feuer (2003)] and Michael Scriven (2004, 2008); mathematicians Burkhardt & Schoenfeld (2003); **engineer Woodie Flowers** [Zaritsky, Kelly, Flowers, Rogers, Patrick (2003)]; and physicist Andre deSessa [Cobb, Confey, diSessa, Lehrer, & Schauble (2003)].

In addition, the following organizations oppose the RCT "gold standard":

(a) American Evaluation Association (AEA) <http://www.eval.org/doestatement.htm>,

(b) American Education Research Association (AERA) <http://www.eval.org/doeaera.htm>, and

(c) National Education Association <http://www.eval.org/doe.nearesponse.pdf> (88 kB).

HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

And in "Re: Should Randomized Control Trials Be the Gold Standard of Educational Research? " [Hake (2005b) I wrote:

"Could physics education researchers (PER's) whose work is predominately in UNDERGRADUATE education utilize RCT's? PER's deal with populations of UP (university professors) and US (Undergraduate Students). Most UP's demand autonomy in the way they teach courses since they obviously know best how to lecture. Most of the US's (or their parents) paid good money to be lectured at. No one that I know of has been insane enough to even suggest that subjects from populations UP and US be randomly assigned to different curricula in an RCT, especially if one curriculum deemphasizes lectures. Also the average UP, thrown into an IE course would be a total disaster. If anyone has some ideas on how to accomplish an RTC among UP's and US's while avoiding dismissal or execution please let me know. Of course one could PAY the subjects, but this might bias the results towards the greedy and/or impecunious."

C. Physics Education Research (PER) Is Not Widely Known in Higher Education (Even Among Some Physicists- See "1" Below)§

1. NRC's CUSE – Oblivious or Dismissive of Halloun & Hestenes (1985a)* and Hake (1998a,b).

In "NRC's CUSE: Oblivious of the Advantage of Pre/Post Testing With High Quality Standardized Tests?" [Hake (2003a)], I wrote [bracketed by lines "HHHHH. . . . .":

HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

[Discussion list subscribers] may or may not be aware of the recent valuable NRC CUSE (Committee On Undergraduate Science Education) report "Improving Undergraduate Instruction in Science, Technology, Engineering, and Mathematics: Report of a Workshop" [NRC (2003b)] by McCray et al. - henceforth called the "McCray Report" (MR). This is the fourth report under the aegis of CUSE. Earlier reports were NRC (1997), NRC (1999), and NRC (2003a). The MR preface states on page viii (my CAPS):

> "In 2002, with new leadership** and largely new membership, CUSE set about to build upon this background [NRC (1997, 1999, 2003a)] by convening a 2-day workshop covering instruction in four major scientific disciplines (biology, chemistry, physics, geosciences) WITH THE PRIMARY GOAL OF DEVELOPING CRITERIA AND BENCHMARKS FOR EVALUATING UNDERGRADUATE STEM PROGRAM EFFECTIVENESS. . . ."

Despite that goal, MR failed to even mention what I regard as one of the primary "Lessons from the Physics Education Reform Effort" (Hake 2002a): Lesson #3. "**High-Quality Standardized Tests Of The Cognitive And Affective Impact Of Courses Are Essential For Gauging The Relative Effectiveness Of Non-Traditional Educational Methods.**"

HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

____

§Derived in part from "Physics Education Research - Not Widely Known in Higher Education" [Hake (2011d,e)]

*An invited paper "Using the RTOP To Evaluate Reformed Science and Mathematics Instruction," in the McCray report NRC (2003b) by biologist Anton Lawson (2003)] *did* reference Halloun & Hestenes (1985a)l, but only in an offhand way with regard to validation the "Reformed Teaching Observation Protocol" (RTOP).

**Richard McCray, Bonnie Brunkhorst, Sarah Elgin, Ronald Henry, John Jungck, Alan Kay, Ishrat Khan, Ramon Lopez, Lillian McDermott, Robert Olin, James Serum, Susan Singer, Carl Wieman; Robert DeHaan, Study Director, Mary Ann Kasper, Senior Project Assistant, Julie Anne Schuch, Research Associate. Physicists participating in the workshop were Paula Heron, Priscilla Laws, John Layman, Ramon Lopez, Lillian McDermott, Carl Wieman, and Jack Wilson.

In "NRC Weighs In On States' Science Assessments" [Hake (2005c)], I wrote (slightly edited):

"It is heartening that, following the lead of Donovan & Pellegrino† (2003), and departing from the dismissal of the mass-testing use of the "Force Concept Inventory" [Hestenes et al. (1992)] in previous NRC reports [e.g., NRC (1997, 1999, 2003a, 2003b), Labov (2003)]; Wilson & Bertenthal (2005) have recognized that multiple-choice testing:

(a) CAN serve to measure student understanding in conceptually difficult areas, and

(b) is therefore of value in large-scale testing of the effectiveness of courses in promoting students' conceptual understanding, as has been demonstrated by the pre/post testing carried out by physics education researchers [for a review see Hake (2002a,b; 2005e)].

2. Cognitive Science and Physics Education Research: What We've Got Here Is Failure to Communicate

The abstract of "Cognitive Science and Physics Education Research: What We've Got Here Is Failure to Communicate" [Hake (2007b)], rejected by the *Journal of Learning Sciences,* reads:

"In 1999 cognitive scientist (CS) Allan Collins (1999) wrote: "Recently researchers have begun to study teaching and learning in the context of real-world learning environments," evidently unaware that Physics Education Researchers (PER's) had been doing classroom research for about three decades. Then in 2004: (a) David Olsen (2004) maintained that the search for 'what works' in education is folly, contradicting PER results; and (b) CS's David Klahr & Milena Nigam (2004) purported to show the superiority of what they called "direct instruction" (DI), defining DI - in a sense unknown to PER's - as a limiting form of what PER's call "Interactive Engagement" (IE). But CS's Kirschner, Sweller, & Clark (2006) outdid the non- recognition of PER by Collins, Olsen, and Klahr & Nigam by not only defining DI as a form of IE, but also proclaiming the "failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching," despite PER evidence for the relative effectiveness of all but unguided "discovery." Will articles such as this in the multidisciplinary *Journal of the Learning Sciences* assuage, to any extent, past failures of CS's to communicate with PER's and vice versa?"

Judging from JLS's rejection, the answer to the above question is "**NO!**"

3. Psychologists Banta & Blaich (2011) Find Few Cases of Improved Learning After a Teaching Innovation

The alert economist Bill Goeff (2011) wrote:

"As a result of increased accountability pressures, U.S. colleges and universities are now directed to 'assess' their students' learning and to take action if that learning is found wanting. It may surprise physicists that assessment experts are not familiar with the role that Physics Education Research (PER) has played in improving student learning. Indeed, Banta and Blaich (2011) lament that they can find virtually no examples of improved learning in higher education after a teaching innovation. Their paper does not mention physics at all! As a non-physicist familiar with some elements of PER, such as the work of Hestenes et al. (1992), Hake (1998a), Crouch et al. (2007) and Deslauriers et al. (2011) (of course, after Banta and Blaich (2011), but notable nonetheless), ***this lack of awareness of PER is remarkable.*"** [My *italics*.]

____

† Donovan & Pellegrino reference Hake (1998a), Halloun & Hestenes (1985a), Hestenes (1987), Hestenes et al. (1992), Maloney et al. (2001), Mazur (1997), and Thornton & Sokoloff (1998).

IV. Impact of Physics Concept Inventories on Undergraduate Physics Education – "EarlyAdopters"

According to Rogers in *Diffusion of Innovation*, p. 283: "*Early adopters* are a more integrated part of the local social system than innovators. . . . This adopter category, more than any other, has the highest degree of opinion leadership in most systems. Potential adopters look to early adopters for advice and information about an innovation. The early adopter is considered by many to be 'the individual to check with' before adopting a new idea. This adopter category is generally sought by change agents as a local missionary for speeding the diffusion process. Because early adopters are not too far ahead of the average individual in innovativeness, they serve as a role model for many other members of the social system. Early adopters help trigger the critical mass when they adopt an innovation."

Herewith are abstracts of articles by six "early adopters" whose efforts were motivated, at least in part, by Concept Inventories:

A. Havard [Crouch & Mazur (2001)]

See the abstract on page x.

B. North Carolina State University [Beichner & Saul (2004)]

ABSTRACT: The SCALE-UP Project has established a highly collaborative, hands-on, computer-rich, interactive learning environment for large-enrollment courses. Class time is spent primarily on hands-on activities, simulations, and interesting questions as well as hypothesis-driven labs. Students sit in three groups of three students at round tables. Instructors circulate and work with teams and individuals, engaging them in Socratic-like dialogues. Rigorous evaluations of learning have been conducted in parallel with the curriculum development effort. Our findings can be summarized as follows: Ability to solve problems is improved, conceptual understanding is increased, attitudes are improved, failure rates are drastically reduced (especially for women and minorities), and performance in follow up physics and engineering classes is positively impacted.

C. MIT [Dori & Belcher (2004)]

ABSTRACT: Educational technology supports meaningful learning and enables the presentation of spatial and dynamic images, which portray relationships among complex concepts. The Technology-Enabled Active Learning (TEAL) Project at the Massachusetts Institute of Technology (MIT) involves media-rich software for simulation and visualization in freshman physics carried out in a specially redesigned classroom to facilitate group interaction. These technology-based learning materials are especially useful in electromagnetism to help students conceptualize phenomena and processes. This study analyzes the effects of the unique learning environment of the TEAL project on students' cognitive and affective outcomes. The assessment of the project included examining students' conceptual understanding before and after studying electromagnetism in a media-rich environment. We also investigated the effect of this environment on students' preferences regarding the various teaching methods. As part of the project, we developed pre- and posttests consisting of conceptual questions from standardized tests, as well as questions designed to assess the effect of visualizations and experiments. The research population consisted of 811 undergraduate students. It consisted of a small- and a large-scale experimental groups and a control group. TEAL students improved their conceptual understanding of the subject matter to a significantly higher extent than their control group peers. A majority of the students in the small-scale experiment noted that they would recommend the TEAL course to fellow students, indicating the benefits of interactivity, visualization, and hands-on experiments, which the technology helped enable. In the large-scale implementation students expressed both positive and negative attitudes in the course survey.

D. University of Colorado at Boulder [Pollock (2004)]

ABSTRACT: We examine the effects of, and interplay among, several proven research-based reforms implemented in an introductory large enrollment (500+) calculus-based physics course. These interventions included Peer Instruction with student response system in lecture, Tutorials with trained undergrad learning assistants in recitations, and personalized computer assignments. We collected extensive informal online survey data along with validated pre/post content- and attitude-surveys, and long answer pre/post content questions designed to assess learning gains and near transfer, to investigate complementary effects of these multiple reforms, and to begin to understand which features are necessary and effective for high fidelity replication. Our average [median] normalized gain was 0.62 [0.67] on the FCI, 0.66 [0.77] on the FMCE, yet we find we cannot uniquely associate these gains with any individual isolated course components. We also investigate the population of students with low final conceptual scores, with an emphasis on the roles played by demographics, preparation, and self-reported attitudes and beliefs about learning.

E. California Polytechnic at San Luis Obispo

1. [Hoellwarth, Moelter, & Knight (2005)]

   ABSTRACT: We present data on student performance on conceptual understanding and on quantitative problem-solving ability in introductory mechanics in both studio and traditional classroom modes. The conceptual measures used were the Force Concept Inventory and the Force and Motion Conceptual Evaluation. Quantitative problem-solving ability was measured with standard questions on the final exam. Our data compare three different quarters over the course of 2 years. In all three quarters, the normalized learning gain in conceptual understanding was significantly larger for students in the studio sections. At the same time, students in the studio sections performed the same or slightly worse on quantitative final exam problems.

2. [Hoellwarth & Moelter (2011)]

ABSTRACT: We have developed a curriculum for introductory mechanics that emphasizes interactive engagement and conceptual understanding using the studio format. As previously reported, we have shown in three different quarters that the curriculum much improved the students' conceptual understanding compared to the traditional course without significantly affecting the scores on a traditional final exam. Here we report the results for the entire three-year period during which the course was taught, 34 sections of the course were taught with 11 different instructors to over 1200 students. In each term, these sections had common exams, syllabus, and schedule. Student experiences were very similar in terms of activities. Student performance was measured using the force and motion conceptual evaluation or the force concept inventory; the average pre/post normalized gain was 0.59. There was no significant correlation with any instructor characteristics, including teaching experience, knowledge of interactive-engagement methods, and attitudes. Because the instructor characteristics are not important, it is the structure of the course that promotes the learning gains.

F. University of British Columbia [Deslauriers, Schelew, & Wieman (2011)]

ABSTRACT: We compared the amounts of learning achieved using two different instructional approaches under controlled conditions. We measured the learning of a specific set of topics and objectives when taught by 3 hours of traditional lecture given by an experienced highly rated instructor and 3 hours of instruction given by a trained but inexperienced instructor using instruction based on research in cognitive psychology and physics education. The comparison was made between two large sections (N = 267 and N = 271) of an introductory undergraduate physics course. We found increased student attendance, higher engagement, and more than twice the learning in the section taught using research-based instruction.

V. CONCLUSIONS

A. Judging From Results in Physics, Concept Inventories *Can* Stimulate Reform But It May Take About Two Decades Before Even "Early Adopters" Become Evident

The research of Halloun & Hestenes (1985a,b) culminating in the first physics concept inventory, the *Mechanics Diagnostic,* was followed 13 years latter by a survey [Hake (1998a,b)] showing that "Interactive Engagement" (IE) courses achieved average course averaged normalized gains <<g>> that were about two standard deviations above those of "Traditional" (T) courses. Then from 2001 to 2011 (3 to 13 years) after that survey "Early Adopters," stimulated at least in part by the Concept Inventory results, published articles on successful reform efforts in large enrollment introductory courses at Harvard [Crouch & Mazur (2001)]; North Carolina State University [Beichner & Saul (2004)]; MIT [Dori & Belcher (2004)]; University of Colorado at Boulder [Pollock (2004)]; California Poly at San Luis Obispo [Hoellwarth et al. (2005, 2011)]; and the University of British Columbia [Deslauriers, Schelew, & Wieman (2011)].

B. Such Slow Adoption of Innovation Appears To Be the Norm in Human Society

On page 267 if his landmark book *Diffusion of Innovations*, Eric Rogers quotes what I choose to call Gabriel Tarde's (1903) Rule:

> "A *slow advance in the beginning*, followed by rapid and uniformly accelerated progress, followed again by progress that continues to slaken until it finally stops: These are the three stages of. . . . .invention. . . . . If taken as a guide by the statistician and by the sociologist, [they] would save many ***illusions***."………………….(*Tarde's Rule*)

Fairweather's comment on the National Academies "Evidence on Promising Practices in STEM Education" [National Academies (2008), Labov et al. (2009), NRC (2011a)]:

> "NSF- and association-funded reforms at the classroom level, however well intentioned, have not led to the hoped for magnitude of change in student learning, retention in the major, and the like *in spite of empirical evidence of effectiveness* [*italics* in the original]"

may reflect only that Fairweather's "hoped for magnitude of change in student learning, retention in the major, and the like" is an ***illusion*** due to his unawareness of (a) typical %adapter vs time curves, such as those shown in Fig. 1-2 "The Diffusion Process" on p. 11 of Rogers' book, (b) Tarde's Rule, and (c) the time frame (2001-2011) during which "Early Adopters" published their evidence for successful use of IE methods in physics classrooms. Thus I am dubious of Fairweather's claim that:

> ". . . .more effort needs to be expended on strategies to promote the adoption and implementation of STEM reforms rather than on assessing the outcomes of these reforms. Additional research can be useful but the problem in STEM education lies less in not knowing what works and more in getting people to use proven techniques."

Instead, I agree with Susan Singer's comment [NRC (2011a, p; 68)]:

> ". . . . .several people view further research on effective practices and further research on implementing change as mutually exclusive.. . . similar to scientific research, the process of change is iterative and requires both types of research."

I also agree with Singer that there is a need "to develop a broader theoretical framework to guide STEM education research within and across disciplines" . . . [[requiring more research on effective practices]]. . . . Furthermore, as stated in Lesson #6 of Hake (2002a) :

> "[there is] a need for more research to develop better strategies for enhancing student learning. On a test as elemental as the FCI, it would seem that reasonably effective courses should yield <g>'s above 0.8, *but thus far none much above 0.7 have, to my knowledge, been reported*. This and the poor showing on the pre/post MPEX test of student understanding of the nature of science and education (Redish et al., 1998) indicate that ***more work needs to be done to improve IE methods***."

C. Will Differences Listed Below Between Engineering Education (EE) and Physics Education (PE) Change the Impact of Concept Inventories in EE From That in PE?

Reed-Rhoads & Imbrie (2008) in "Concept Inventories in Engineering Education," indicated three differences between engineering and physics education:

1. Little Is Known About Misconceptions in Engineering

Reed-Rhoads & Imbrie (2008) wrote: "There is very little known or published on the engineering concepts and subject matter misconceptions [so that] engineering faces the challenge of using the inventories as a tool to identify misconceptions, being only guided in a general sense of instructors' perceptions of potential misconceptions (Streveler et al., 2008). . . . . . [[student interviews and analysis of pretest scores (especially if trial pretests allow free responses) might help to identify misconceptions]]. . . . .

2. Concept Inventories May Be Regarded as Required for Summative Evaluation of Engineering Education

Reed-Rhoads & Imbrie (2008) wrote: "The engineering inventories are being developed in an era of evolving accreditation standards that focus on learning outcomes. . . . .[[see e.g. ABET (2011)]]. . . . This coincidence could infer the purpose of the CIs is for accreditation instead of instruments aimed at increasing student learning. The FCI, on the other hand, might be viewed as a more charitable contribution to existing research. This difference could infuse a tincture of skepticism regarding the researchers' motivation (i.e., *have-to* vs. *want-to*)." . . . . . .[[This should not be a problem IF Engineering CIs are used only for *formative* assessment to improve student learning - see footnote ‡ in page 27]]. . . . .

3. Engineering Education Can Build on the Physics Concept Inventory Evidence That Interactive Engagement Courses Are Much Effective Than Traditional *Passive Student* Lecture Courses

Reed-Rhoads & Imbrie (2008) wrote: "In part due to the FCI, evidence is beyond anecdotal that traditional teaching is ineffective providing a landscape ripe for encouraging faculty to improve their practices with conceptual assessment imperative to demonstrate the effectiveness of innovations." . . . . [[this, of course, could act to shorten the time between an engineering innovation and its widespread adoption.]]. . . . .

In my opinion the differences between engineering and physics education cited by Reed-Rhoads & Imbrie (2008) in "1" and "2" above should not slow the rate of adoption of reform methods in engineering education below that in physics education for the reasons noted above in the inserts at . . . . .[[insert]]. . . . Furthermore there are at least:

D. Seven Reasons Why the Rate of Adoption of Reform Methods in Engineering Education May Be Greater That in Physics Education:

1. Engineering Education Can Build on the Physics Concept Inventory Evidence That Interactive Engagement Courses Are Much Effective Than Traditional *Passive Student* Lecture Courses

The same point made by Reed-Rhoads & Imbrie (2008) in "3" above.

2. ABET's Incentives to Improve Engineering Education

According to ABET's (2011) "Criteria For Accrediting Engineering Programs: Effective for Evaluations During the 2011-2012 Accreditation Cycle," ABET accreditation requires that Engineering Departments demonstrate student outcomes that prepare graduates to attain the [following] educational objectives:

(a) an ability to apply knowledge of mathematics, science, and engineering;
(b) an ability to design and conduct experiments, as well as to analyze and interpret data;
(c) an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability;
(d) an ability to function on multidisciplinary teams ;
(e) an ability to identify, formulate, and solve engineering problems;
(f) an understanding of professional and ethical responsibility;
(g) an ability to communicate effectively;
(h) the broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context
(i) a recognition of the need for, and an ability to engage in life-long learning;
(j) a knowledge of contemporary issues;
(k) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

The above objectives would seem to require reform methods of instruction similar to those developed in physics [for a listing see Hake (1998b)], together with formative Concept Inventories to gauge the effectiveness of those methods, both among themselves and in comparison with traditional methods.

3. National Studies Call For a New Type of Engineering Graduate

Redish & Smith (2008) in "Looking beyond content: Skill development for engineers," wrote: "Numerous studies and reports now call for a different type of engineering graduate. Thoughtful studies abound. . . . ." Studies referenced by Redish & Smith are:

a. A 1997 Engineering Futures Conference co-sponsored by Boeing and Rensselaer Polytechnic Institute, which resulted in a provocative set of characteristics of a global engineer (Boeing, 1997). I've found no online version of Boeing (1997), but Karl Smith (2007) in a talk "Preparing Students for an Interdependent World: Role of Cooperation and Social Interdependence Theory" gives a list of the "Desired Attributes of a Global Engineer" from that report:

(1) A good grasp of engineering science fundamentals, including:
(a) Mechanics and dynamics
(b) Mathematics (including statistics)
(c) Physical and life sciences
(d) Information science/technology

(2) A good understanding of the design and manufacturing process (i.e., understands engineering and industrial perspective)

(3) A multidisciplinary, systems perspective, along with a product focus

(4) A basic understanding of the context in which engineering is practiced, including:
(a) Customer and societal needs and concerns
(b) Economics and finance
(c) The environment and its protection
(d) The history of technology and society

(5) An awareness of the boundaries of one's knowledge, along with an appreciation for other areas of knowledge and their interrelatedness with one's own expertise

(6) An awareness of and strong appreciation for other cultures and their diversity, their distinctiveness, and their inherent value

(7) A strong commitment to team work, including extensive experience with and understanding of team dynamics

(8) Good communication skills, including written, verbal, graphic, and listening

(9) High ethical standards (honesty, sense of personal and social responsibility, fairness, etc)

(10) An ability to think both critically and creatively, in both independent and cooperative modes

(11) Flexibility: the ability and willingness to adapt to rapid and/or major change

(12) Curiosity and the accompanying drive to learn continuously throughout one's career

(13) An ability to impart knowledge to others]]. . . . . .

b. *The Engineer of 2020: Visions of Engineering in the New Century* [NAE (2004)]; The Executive Summary reads, in part:

"In the past, changes in the engineering profession and engineering education have followed changes in technology and society. Disciplines were added and curricula were created to meet the critical challenges in society and to provide the workforce required to integrate new developments into our economy. Today's landscape is little different; society continually changes and engineering must adapt to remain relevant. But we must ask if it serves the nation well to permit the engineering profession and engineering education to lag technology and society, especially as technological change occurs at a faster and faster pace. Rather, should the engineering profession anticipate needed advances and prepare for a future where it will provide more benefit to humankind? Likewise, should engineering education evolve to do the same?"

c. *Educating the engineer of 2020: Adapting engineering education to the new century* [NAE (2005a)];

The description reads:

"Educating the Engineer of 2020 is grounded by the observations, questions, and conclusions presented in the best-selling book *The Engineer of 2020: Visions of Engineering in the New Century*. This new book offers recommendations on how to enrich and broaden engineering education so graduates are better prepared to work in a constantly changing global economy. It notes the importance of improving recruitment and retention of students and making the learning experience more meaningful to them. It also discusses the value of considering changes in engineering education in the broader context of enhancing the status of the engineering profession and improving the public understanding of engineering. Although certain basics of engineering will not change in the future, the explosion of knowledge, the global economy, and the way engineers work will reflect an ongoing evolution. If the United States is to maintain its economic leadership and be able to sustain its share of high-technology jobs, it must prepare for this wave of change."

d. "The National Engineering Education Research Colloquies" Special Report [NEERC (2006)]. The first paragraph reads [slightly edited]:

"Rapid changes in the worldwide engineering enterprise are creating a compelling rationale for us to rethink how we should educate future generations of engineers [NSF (1995), NRC (2007a), RAE (2006), IEAC (1996). According to "The Engineer of 2020" [NAE (2004)], tomorrow's graduate will need to collaboratively contribute expertise across multiple perspectives in an emerging global economy that is fueled by rapid innovation and marked by an astonishing pace of technological breakthroughs. Deteriorating urban infrastructures, environmental degradation, and the need to provide housing, food, water, and health care for eight billion people will challenge the analytical skills and creativity of engineers. From a U.S. perspective, a continuing decline in interest by American youths in engineering, a shrinking capacity for technological innovation, and an engineering research infrastructure in distress are early warning signs that the nation's prosperity and security are at stake if we fail to take action. Our leadership and capacity for innovation are destined to erode unless current trends are reversed [NAE (2005), TFFAI (2005), COC (2004)].

e. Commissioned papers for the National Center on Education and the Economy, ...[[online at <http://bit.ly/pKs6fD>]]. . . . .including, "Rethinking and Redesigning Curriculum, Instruction and Assessment" (Pellegrino, 2006). . . . .[[and "Sources of Innovation and Creativity: A Summary of the Research" (Adams, 2005)]]. . . .

Among Publications *not* mentioned by Redish & Smith (2008), but of possible interest to those interested in engineering education are:

f. *Engineering Research and America's Future* [NAE (2005b)].

The second paragraph of the preface reads:
"To assess and document the current state of the U.S. engineering research enterprise and to raise awareness of the critical role of engineering research in maintaining U.S. technological leadership, the National Academy of Engineering (NAE) initiated the current study. The focus of the study is primarily on academic research because of its importance to long-term basic engineering research and to educating future engineers and engineering researchers. The study is based on the opinions and judgments of a 15-member committee of experts from industry and universities. The committee's deliberations were informed by testimony from key decision makers and policy makers in the federal government, as well as a detailed review of many recent studies on national R&D policy, investment patterns, needs, and shortcomings."

g. *Is America Falling Off the Flat Earth* [NRC (2007b)].

The description reads: "The aviation and telecommunication revolutions have conspired to make distance increasingly irrelevant. An important consequence of this is that US citizens, accustomed to competing with their neighbors for jobs, now must compete with candidates from all around the world. These candidates are numerous, highly motivated, increasingly well educated, and willing to work for a fraction of the compensation traditionally expected by US workers. If the United States is to offset the latter disadvantage and provide its citizens with the opportunity for high-quality jobs, it will require the nation to excel at innovation -- that is, to be first to market new products and services based on new knowledge and the ability to apply that knowledge. *This capacity to discover, create and **market*** . . . . [[my *italics*]]. . . . will continue to be heavily dependent on the nation's prowess in science and technology. Indicators of trends in these fields are, at best, highly disconcerting. While many factors warrant urgent attention, the two most critical are these: (1) America must repair its failing K-12 educational system, particularly in mathematics and science, in part by providing more teachers qualified to teach those subjects, and (2) the federal government must markedly increase its investment in basic research, that is, in the creation of new knowledge. Only by providing leading-edge human capital and knowledge capital can America continue to maintain a high standard of living--including providing national security--for its citizens."

h. *Restructuring Engineering Education: A Focus on Change* [NSF (1995)].

i. *Shaping the Future, Volume II: Perspectives on Undergraduate Education in Science, Mathematics,Engineering, and Technology* [NSF (1996)].

j. *Making the Case for Engineering: Study and Recommendations* [NSF (2005)].

The above national studies (just as the Boeing and ABET criteria above) would seem to require reform methods of instruction similar to those developed in physics [for a listing see Hake (1998b)], together with formative Concept Inventories to gauge the effectiveness of those methods, both among themselves and in comparison with traditional methods.

4. Many Engineering Educators Are Interested In Concept Inventories and Their Potential To Improve Engineering Education - See, e.g.:

"Progress on Concept Inventory Assessment Tools" [Evans et al. (2003)];

"Concept Inventories: Tools For Uncovering STEM Students' Misconceptions" [Richardson (2004)];

"The statistics concept inventory: Developing a valid and reliable instrument" [Allen et al. (2004)];

"Does Active Learning Work? A Review of the Research" [Prince (2004)];

"A statics concept inventory: Development and psychometric analysis" [Steif & Dantzler (2005)];

"Active and Cooperative Learning in Signal Processing Courses" [Buck & Wage (2005)];

"New practices for administering and analyzing the results of concept inventories" [Steif & Hansen (2007)];

"Tools to facilitate better teaching and learning: Concept inventories" [Reed-Rhoades et al. (2007)];

"Learning Conceptual Knowledge in the Engineering Sciences: Overview and Future Research Directions" [Streveler et al. (2008)];

"Concept Inventories in Engineering Education" [Reed-Rhoads & Imbrie (2008)];

"White Paper on Promising Practices in Undergraduate STEM Education" [Froyd (2011)].

5. Establishment in 2002 of the "Engineering Center for the Advancement of Scholarship on Engineering Education (CASEE)

"The Center for the Advancement of Scholarship on Engineering Education (CASEE) is the first operating center of the National Academy of Engineering. We are dedicated to achieving excellence in engineering education, education that is effective, engaged, and efficient. As a reflection of our on-going research and implementation activities, these pages will be dynamic and frequently updated."

Redish & Smith (2008) point out that according to CASEE <http://bit.ly/pfjrDS> "there are five U.S. schools with Ph.D. programs in engineering education . . . . [[now 10 according to http://bit.ly/p0CfL7> . . . and more than 20 schools with engineering education centers <http://bit.ly/qxRjek>, and interest in engineering education research

6. Recent Initiation of the Concept Inventory Hub <http://ciHUB.org/> at Purdue

This a community for Concept Inventory developers, researchers, and users should give a tremendous boost to the improvement of engineering education – see e.g. the Merlot page at <http://bit.ly/puI0PI>,

7. Engineers Are "Adept at Correlating Exactitude With Chaos to Bring Visions into Focus."

"The essence of engineering . . . .is *integrating* all knowledge for some purpose. As society's 'master integrators,' engineers must provide leadership in the concurrent and interactive processes of innovation and wealth creation. The engineer must be able to work across many different disciplines and fields - and make the connections that will lead to deeper insights, more creative solutions, and ***getting things done***. In a poetic sense, paraphrasing the words of Italo Calvino (1988), ***the engineer must be adept at correlating exactitude with chaos to bring vision bring visions into focus***."

Joseph Bordogna (1997)
former Head of NSF's Directorate for Engineering

VI. Appendix

A. Criticisms of the FCI

a. *Jose Mestre* (2008) wrote (correcting his currently dead URL): "The good news is that many other STEM disciplines are developing their version of the FCI <http://bit.ly/qmIf1b> . . . . .[[this is the engineering "Foundation Coalition"]]. . . The bad news is that this is a very narrow focus and we should look beyond identifying and eradicating misconceptions as the goal of STEM instruction, but at least it starts a much needed dialog toward curricular improvement."

In my opinion, the FCI is focused on more than just "identifying and eradicating misconceptions," as Mestre (2005) himself indicated in his earlier "Facts and myths about pedagogies of engagement in science learning." There Mestre wrote [my *italics*]:

> "Figure 1 shows two questions from [the FCI], which should make it evident that the FCI *measures understanding of basic conceptual knowledge that physicists expect students to grasp after taking an introductory physics course*."

Furthermore, as Hestenes (1997) states [my *italics*]: The FCI was developed to assess the *effectiveness of mechanics courses in meeting a minimal performance standard*: to teach students to reliably discriminate between the applicability of scientific concepts and naïve alternatives in common physical situations."

Similarly the questions in any good multiple-choice test of introductory students' conceptual understanding in any STEM area would employ distractors which feature naïve alternatives in common situations.

b. *Susanne Singer* (2008) wrote: "While valuable in moving research forward, there is more to understanding and enhancing student learning in STEM fields than addressing alternative conceptions uncovered in concepts inventories, underscoring the need for multiple modes of evidence in undergraduate learning in STEM fields."

See my response above to Mestre's similar denigration of the FCI.

b. The FCI appraisals of *Kenneth Heller* and *Karen Cummings* appear on page 41 of *Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics Education: Summary of Two Workshops* [NRC (2011a)]:

> "*Kenneth Heller pointed out that the FCI is not about forces and is not a concept inventory.* [My *italics*.] Rather, it is an instrument about misconceptions that is based on the misconception research. Although the instrument is reliable, Heller stressed that it is not a predictor of students' success in introductory physics. He asked the presenters whether they are trying to replicate the success of the FCI or develop concept inventories that may or may not have the same properties as the FCI. Libarkin and Reed-Rhoads said their respective communities (geosciences and engineering) are trying to do both. *Cummings agreed with Heller's assessment of the FCI* and emphasized the importance of being clear about what these instruments measure."

I suspect that Ken Heller's remark may reflect the influence of "What Does the Force Concept Inventory Actually Measure?" [Huffman & Patricia Heller (1995)] and "Interpreting the Force Concept Inventory: A Reply to Hestenes and Halloun" [Patricia Heller and Huffman (1995)]. But, as indicated on pp. 14-15 of this article, even Huffman and Patricia Heller have stated that "the FCI is one of the most reliable and useful physics tests currently available for introductory physics teachers" [Huffman & Heller (1995)] and that the FCI is "the best test currently available [to evaluate the effectiveness of instruction in introductory physics courses]" [Heller & Huffman (1995)]. How could the latter statement be correct if the FCI failed to gauge students' conceptual understanding of mechanics and thus be a form of "concept inventory"? See also my responses above to criticisms of the FCI by Mestre and Singer.

c. *Robin Millar* and *Jonathan Osborne*† in *Quality Research in Literacy and Science Education: International Perspectives and Gold Standards* (Shelley, Yore, & Hand, eds., 2009) wrote [bracketed by lines "M&O-M&O-. . . ."; my ***bold italics***]*:

M&O-M&O-M&O-M&O-M&O-M&O-M&O

One striking feature of science education is that ***no standard or commonly agreed outcome measures exist for any major topic.*** The need for them is demonstrated by the way in which some published assessment tools, like the Force Concept Inventory (FCI, Hestenes, Wells, & Swackhamer, 1992), have been used by researchers.

But such tools are also subject to quite significant criticism - not the least because ***the construction of any test instrument requires choices about the learning outcomes of most worth***, which inevitably involve values. . . .[[Yes, Hestenes et al. (1992) evidently valued students' learning the basic concepts of Newtonian mechanics - obviously arbitrary, arguable, and suspect **;-)**]] . . . . .

Also instruments like the FCI have not been subjected to the kind of rigorous scrutiny of factorial structure . . . .[[***not so*** **!** - see e.g. Huffman & Heller (1995), Hestenes & Halloun (1995), Heller & Huffman (1995), Halloun & Hestenes (1995)]]]. . . . . . ***and content validity*** . . . .[[***not so*** **!** - see section IIB. "Validity and reliability of the mechanics test" (Mechanics Diagnostic) in Halloun & Hestenes (1985a) - that verification of validity applies also to the FCI since it's almost the same as the Mechanics Diagnostic – and see also Kuder-Richardson reliability coefficients (KR-20) in the Tables Ia,b,c footnotes of Hake (1998b)]]. . . . . that would be standard practice for measures of attainment or learning outcome in other subject areas, in particular ***the kinds of standard measure used by psychologists.*** . . . [[so PER's should emulate psychologists?? See e.g.: (a) "Do Psychologists Research the Effectiveness of Their Courses? Hake Responds to Sternberg" (Hake (2005d), and (b) "Possible Palliatives for the Paralyzing Pre/Post Paranoia that Plagues Some PEP's" (PEP's = Psychologists, Education Specialists, and Psychometricians) (Hake, 2006k)]] . . . .

---

† Osborne was on the committee that produced the NRC's (2011c) *A Framework for K-12 Science Education* which, for the most part, ignores physics education research, as does NRC (2011b). Osborne (2007), in his "In praise of armchair science education" and either unaware or dismissive of physics education research, wrote: "50 years of research, curriculum development, and implementation have not presented consistent and compelling patterns of outcomes."

* These comments are derived from a discussion-list post "Re: Quality Research in Literacy and Science Education: International Perspectives and Gold Standards" [Hake (2010a)]

So using a trusted, off-the-shelf, outcome measure is not an option for the science education researcher . . . . [[So s(he) should *ignore* formative diagnostic tests such as (a) the FCI; (b) those listed: (1) at NCSU's (2010) "Assessment Instrument Information," (2) at FLAG (2011), (3) by Reed Rhoades and Imbry (2008), (4) by Libarkin (2008), and (5) those referenced for astronomy, biology, chemistry, economics, engineering, geoscience, and math in "Should We Measure Change? Yes!" (Hake, 2011b)]]. . . . . . .
M&O-M&O-M&O-M&O-M&O-M&O-M&O

> Millar and Osborne appear to be either dismissive or oblivious of the fact that "Conceptual Inventories" <http://en.wikipedia.org/wiki/Concept_inventory>, developed through arduous quantitative and qualitative research by disciplinary experts, are currently being used to improve undergraduate - and some high-school - courses in science, technology, engineering, and math (STEM) disciplines (but not psychology!) - see, e.g., (a) "Design-Based Research in Physics Education Research: A Review" (Hake (2008b) ; (b) "Should We Measure Change? Yes!" (Hake, 2011b); and (c) "Workshop on Linking Evidence and Promising Practices in STEM Undergraduate Education" [National Academies (2008)].

B. Criticisms of <g>

a. Psychometric guru *Robert Mislevy* (2006), while acknowledging the value of <g> in analyzing pre/post test gains Hake (1998a,b), states that he is *uninterested* in <g> because it, unlike "Item Response Theory" [see e.g., Junker (1999), Rudner (2001)] is not "grounded in the framework of probability-based reasoning."*

But, in my opinion, Mislevy's objection must be balanced against the:

(1) *empirical* justification of <g> 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, as indicated in Sect. IIC12 "*Empirical* Justification of <g> as a Comparative Measure of Course Effectiveness ";

(2) evidently unaddressed problem of how to employ IRT *to compare the effectiveness of courses* in which the initial average knowledge state of students is highly variable – IRT use appears to be concentrated instead on test analysis, see e.g., Junker (1999), Pellegrino et al. (2001), Morris et al. (2006), Wang & Bao (2009), Ding & Beichner (2009), and Wallace & Bailey (2010);

(3) difficulties that average faculty members might experience in using IRT to improve the effectiveness of their courses.

---

*For an unpublished attempt to ground normalized gain "in the framework of probability-based reasoning" see Bao (2008).

b. *Susan Ramlo* (2008a) wrote:

> "... there is a more statistically acceptable method for including pretest scores' relationship with the posttest scores than using either ..... [[average normalized gain <g> or the average of single student normalized gains gain calculations]]. . . . .The general linear model (GLM... or, basically, linear regression) is a more stable method for including the pretest scores' relationship to the posttest score."

But, as stated in Hake (2008c):

> "Ramlo's statement . . . . [[see above]]. . . suggests that she may be unaware of the research question that was posed in Hake (1998a,b) and for which <g> was empirically set forth to assist an answer: 'Can the classroom use of Interactive Engagement methods increase the effectiveness of introductory mechanics courses well beyond that attained by traditional methods?' I think *Phil Sadler's* (2008a,b) posts, also touting the general linear model, also betray an ignorance of the motivation for <g> in Hake (1998a,b). Can Sadler or Ramlo explain how the GLM could be used to answer the above research question in a survey of 62 courses with widely varying average pretest scores?"

Neither Sadler of Ramlo responded to this challenge.

c. *Redish & Hammer* (2009) wrote (converting their nomenclature to that used in this article):

> "An often used figure of merit for pre-post testing is the average normalized gain <g>. . . . . This measure is typically used to permit the comparison of classes with different pre-test scores. Some care must be taken, as the score is particular to the specific test used and may distort the result when either the pre-or post-test averages are high, producing end effects. We are grateful to Robert Mislevy (private communication) for this comment."

Since <g> is an *empirically* derived parameter (see the above Sect. IIC12 "*Empirical* Justification of <g> as a Comparative Measure of Course Effectiveness" that showed no evidence of "end effects" in the survey of Hake (1998a,b), it's not clear why Meslevy claims that <g> "may distort the result when either the pre-or post-test averages are high." In fact, <g> avoids a "ceiling effect" if the test (as does the FCI) poses a "performance ceiling effect," as opposed to an "instrumental ceiling effect" – for a discussion see Hake (2006g,h,i,j).

d. *Miller et al.* (2010) wrote (converting their nomenclature to that used in this article): "Hake's single-student gain for the ith student, g_i, is defined the actual gain ($\%post_i - \%pre_i$) divided by the maximum possible gain ($100\% - \%pre_i$). When calculating the average g for an entire class, the instructor has two choices: calculate g_i for each student and then average the values to obtain $g_{ave} = (\Sigma_{(\text{over i from 1 to N})}\, g_i)\,/N$, or average *pre* and *post* for the class to obtain the average normalized gain $<g> = (<\%post> - <\%pre>)/\,(100\% - <\%pre>$. * The former approach is more rigorous . . . . .[[on what grounds do Miller et al. base this claim?]]. . . . though a perfect pretest score results in division by zero. The latter approach is often simpler to calculate, though it masks important information about individual students. In this paper, we focus on individual student gains, and therefore rely solely on the former approach."

But as explained in Section IIE2-Aside(b): "As a statistic for comparison of courses and for meta-analyses, the class average <g> is better, in my opinion, than g_{ave} (the average of the single-student normalized gains) because the latter:

(1) must exclude students who score 100% on the pretest and thus achieve an infinite or indeterminate g; and (2) may introduce skewing due to outliers who score near 100% on the pretest and less on the posttest such their <g>'s are large and negative. The selective removal of outliers so as to avoid outliers by various different investigators with different outlier criteria will lead to a degree of uncertainty in comparing normalized gains of different courses."

With regard to Miller et al.'s use of term "Hake's gain," neither the single-student gain g_i, nor the average gain $<g>$ is "Hake's." In Hake (2008) I wrote: "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 indicate that *g* is an example of "POMP" (Percentage Of Maximum Possible).

____

* Footnote #46 of Hake (1998a) states that: (1) for N > 20, g_{ave} is usually within 5% of <g>; and (2) low [g_{ave} – <g>] implies a low correlation between g_j and $pre_j \equiv (S_i)_j$ for individual students, just as there is a low correlation between <g> and <Si> for courses. Bao (2006) has also discussed the relationship between g_{ave} and <g>, but was either oblivious or dismissive of footnote #46 of Hake (1998a).

e. The abstract of *Dowd et al. (2011)* reads, in part [my counters at ". . . .[[counter]]. . . ":

"The normalized gain, g, is among the most widely- and easily-used metrics in physics education research, particularly as it is applied to the Force Concept Inventory (FCI). However, its ease and ubiquity mask several pitfalls associated with the metric, particularly when it is used to directly compare a small number of courses. Specifically, these pitfalls are: the ambiguities between 'class' and 'student average' gain,

. . . .[[this "pitfall" in *not* due to the "normalized gain metric." There were no such ambiguities in Hake (1998a,b) – see especially footnote #46 of Hake (1998a). Furthermore, the ambiguities that exist in some PER articles – see e.g., Coletta & Phillips (2005) - are due *only* to the carelessness of their authors in not stating exactly how they calculate the "average normalized gain."]]. . . . .

the tendency not to report error margins,

. . . .[[this is not the fault of the "normalized gain metric"]]. . . .,

the effect of losses in initially high-scoring populations

. . . . [[in my opinion, since there's about a two-standard deviation differences between the average <<g>> of the average normalized gains <g> between IE and T courses, it's doubtful that not taking into account the "effect of losses" would have changed the conclusions of Hake (1998a,b) that "The above conceptual (FCI) and problem solving (MB) test results strongly suggest that the use of IE strategies can increase mechanics-course effectiveness well beyond that obtained with traditional methods.]]. . . . . .,

and the extent to which details about classes are washed out by reporting a single number. .

. . . .[[the influence of "details about classes" evidently is of minor importance in the comparison of IE and T courses in Hake (1998a,b)]]. . . . .

f. *Docktor & Mestre* (2011) wrote: "The 'normalized gain' is a commonly reported measure for comparing pretest and posttest scores across populations (Hake, 1998a), but the statistical origin for this measure is unclear and alternatives have been suggested (such as 'normalized change'). It is unclear why normalized gain is still favored, and the PER community should reach an agreement about how to analyze and report scores from concept inventories."

The statistical origin is unclear?? As indicated in Section IIC11 above, "Experimental Justification of <g> as a *Comparative Measure of Course Effectiveness,"* the "average normalized gain" <g> was utilized in Hake (1998a,b) as a strictly *empirically-based* parameter with no claim to a statistical pedigree. Nevertheless <g> provided a consistent analysis of pre/post test results – see Fig. 1 and 2 above - over diverse student populations in high schools, colleges, and universities. Furthermore, similar consistent analyses using <g> have been obtained in over 25 physics education research papers as listed above in Section IIC9 "Research Consistent With the Above Meta-analysis."

In my opinion, it is unclear why Docktor & Mestre (2011): (1) choose to discount the above evidence for the value of <g>, and (2) think that "It is unclear why normalized gain is still favored."

As regards the “normalized change” [Marx & Cummings (2007)], in a footnote on p. 1 of “Should We Measure Change? Yes!” [Hake (2011b)], I wrote:

> “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 (2002d), I think that. . . .[[for comparing the effectiveness of different courses]]. . . . replacing a class average $<g_i>$ with a class average $<c_i>$, as advocated by Marx & Cummings, has little, if any, advantage. . . .[[over simply using the average normalized gain <g>, as in Hake (1998a,b) and most subsequent studies by other investigators.]]. . . . . .”

g. Lasry, Rosenfield, Dedic, Dahan, & Reshef (2011) wrote: “The internal reliability of the *Mechanics Diagnostic Test* was previously determined, and KR□20 was found to equal 0.86 for the pretest and 0.89 for the post-test [Halloun & Hestenes (1985a)]. To our knowledge, *no assessment of the internal consistency of the FCI has been reported*. [My *italics*.]

In the above Section IIC7 “A Crucial Companion Paper to Hake (1998a)” I pointed out that the AJP rejected paper “Interactive-engagement methods in introductory mechanics courses” [Hake (1998b)] contained, among other valuable information, the average pre/post test scores, standard deviations, instructional methods, materials used, institutions, and instructors for each of the survey courses. In addition, the following posttest Kuder-Richardson reliability coefficients (KR-20) were contained in footnote ††† of Tables Ia,b,c of Hake (1998b) - nearly identical to the 1992 version: KR-20 = 0.81 and nearly identical to the 1995 version: KR-20 = 0.86).

The baleful effects of inadequate reviewing of articles submitted for publication in the field of Physics Education Research (PER) have been discussed in “Re: Learning Physics with a Statistical Gold Standard” [Hake (2011m)].

VI. REFERENCES

NOTES:

1. The present reference formatting is a blend of the *best* formatting features from the style manuals of the AIP (American Institute of Physics <http://bit.ly/d8hJgp>), APA (American Psychological Association <http://apastyle.apa.org/>), and CSE (Council of Science Editors <http://bit.ly/aqIKXb>)]. Such formatting is not commonly employed *but should be*.

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

3. References preceded by † are not tied to text material or to other references listed here but are included because they might be of interest to some readers.

4. All URL's were accessed on 12-13 Sept 2011; most are shortened by <http://bit.ly/>.

†AAAS. 2004. *Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics (STEM) Education*. AAAS, publisher's information at <http://bit.ly/pwmF8C>.

ABET. 2011. "Criteria For Accrediting Engineering Programs: Effective for Evaluations During the 2011-2012 Accreditation Cycle," Incorporates all changes approved by the ABET Board of Directors as of October 30, 2010, Accreditation Board for Engineering and Technology, online as a 90 kB pdf at <http://bit.ly/q08zP5>.

Adams, K. 2006. "The Sources of Innovation and Creativity," a paper commissioned by the National Center on Education and the Economy for the *New* Commission on the Skills of the American Workforce [NCEE (2007)]; online as a 815 kB pdf at <http://bit.ly/qrSPgO>. Compare this with DeHaan's (2009) "Teaching Creativity and Inventive Problem Solving in Science."

Allen, K., A. Stone, T. Reed-Rhoads, & , T.J. Murphy. 2004. *The statistics concept inventory: Developing a valid and reliable instrument.* Paper presented at the American Society for Engineering Education Annual Conference and Exposition, Salt Lake City, UT; online as a 193 kB pdf at <http://bit.ly/ogz9mL>. See also Allen (2006), Stone et al. (2003), Stone (2006), and Allen et al. (2008).

Allen, K.C. 2006, *The Statistic Concept Inventory: Development and Analysis of a* Cognitive *Assessment Instrument in Statistics*," Unpublished Doctoral Dissertation, University of Oklahoma; online as a 3.4 MB pdf at <http://bit.ly/oBKFeq>.

Allen, K., T. Reed-Rhoads, R.A. Terry, T.J. Murphy, & A.D. Stone. 2008. "Coefficient Alpha: An Engineer's Interpretation of Test Reliability," *Journal of Engineering Education* **97**(1): 87-94; online as a 164 kB pdf at <http://bit.ly/pvwrRb>.

Anderson, W.A., U. Banerjee, C.L. Drennan, S.C.R. Elgin, I.R. Epstein, J. Handelsman, G.F. Hatfull, R. Losick, D.K. O'Dowd, B.M. Olivera, S.A. Strobel, G.C. Walker, I.M. Warner. 2011. "Changing the Culture of Science Education at Research Universities: Universities must better recognize, reward, and support the efforts of researchers who are also excellent and dedicated teachers," *Science*, 14 January, **331**(6014): 152-153: online as a 172 kB pdf at <http://bit.ly/eSLoCl>. For supporting online references see Hake (2011a) at <http://bit.ly/g24Iqm>. Anderson et al. reference the following physics education research sources:

(a) University of Maryland Physics Education Research Group <http://www.physics.umd.edu/perg/>;
(b) Physics Education Research at Colorado University - Boulder, <http://www.colorado.edu/physics/EducationIssues/>;
(c) "Six Lessons From the Physics Education Reform Effort" [Hake (2007c)],
(d) "Peer instruction: From Harvard to the two-year college" [Lasry, Mazur, & Watkins (2008)];
(e) "Transforming Science Education at Large Research Universities: A Case Study in Progress" [Wieman, Gilbert, & Perkins (2010)]

Arons, A.B. 1973. "Toward wider public understanding of science," *Am. J. Phys.* **41**(6): 769-782; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v41/i6>.

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://ajp.aapt.org/resource/1/ajpias/v42/i2>.

Arons, A.B. 1986. "Conceptual Difficulties of Science," *Proceedings of the Chicago Conferences on Liberal Education, No. 1*," edited by M.R. Rice, Univ. of Chicago, p. 23-32.

Arons, A.B. 1997. *Teaching Introductory Physics*. Wiley, publisher's information at <http://bit.ly/jBcyBU>. Amazon.com information at <http://amzn.to/nIiPGh>.

Arum, R. & J. Roksa. 2011. *Academically Adrift: Limited Learning on College Campuses*. University of Chicago Press, publisher's information, including a synopsis and bio, are online at <http://bit.ly/gPYBHj>. Amazon.com information at <http://amzn.to/f1f45O>, note the searchable "Look Inside" feature. See also "Academically Adrift?" [Hake (2011c)].

Atkinson, D., E. Churchill, T. Nishino, and H. Okada. 2007. "Alignment and interaction in a sociocognitive approach to second language acquisition," *Modern Language Journal* **91**(2): 169-188. An ERIC abstract is online at <http://1.usa.gov/n2JcTD>.

Bao, L. 2006. "Theoretical comparisons of average normalized gain calculations," *Am. J. Phys.* **74**(10): 917-922; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v74/i10>.

Bao, L. 2008. "Dynamic Models of Learning and Education Measurement," evidently unpublished; online as a 291 kB pdf at <http://arxiv.org/pdf/0710.1375v2>. The last two sentences read:

"For all courses. . . . .[[surveyed by Hake (1998a,b)]]. . . . the average normalized gains had a very low correlation +0.02 with average pretest scores. This feature of the normalized gain has allowed researchers to investigate the effectiveness of instruction using data collected from classes with widely different average pretest scores. However, the question of why the average normalized gain has this feature and to what extent this feature is generally present is not well understood. In addition, there have been debates as to what the normalized gain actually measures, and concerns that it lacks a probability framework that undergirds psychometric methods such as Item Response Theory (IRT). The present model leads to an explanation of the observed features of the normalized gain, connects to other models such as IRT, and shows that the normalized gain does have a probability framework but one different from that emphasized by IRT."

Banta, T.W. & C. Blaich. 2011. "Closing the Assessment Loop," *Change Magazine*, January/February; online at <http://bit.ly/lQyEYp>.

Barr, R.B. & J. Tagg. 1995. "From Teaching to Learning: A New Paradigm for Undergraduate Education," *Change Magazine* **27**(6): 13-25, November/December; online as a 111 kB pdf at <http://bit.ly/8XGJPc>. See also Tagg (2003).

Beichner, R.J. 1994. "Testing student interpretation of kinematics graphs." *Am. J. Phys.* **62**(8): 750-762. A good introduction for physicists to some of the basic statistics of educational research; online as a 594 kB pdf at <http://bit.ly/qI2WBt>.

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 as a 1 MB pdf at <http://bit.ly/aBLkGW>. See also Beichner (2011).

Beichner, R.J. 2011. "The SCALE-UP Project: A Student-Centered Active Learning Environment for Undergraduate Programs," paper commissioned by the National Academies, online as a 3.8 MB pdf at <http://bit.ly/rrQyOE>. See also National Academies (2008).

Black, P. & D. Wiliam. 1998. "Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan* **80**(2): 139-144, 146-148; online as a 233 kB pdf at <http://bit.ly/c7ZQlr>.

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 as a 2.6 MB pdf at <http://bit.ly/r4mEYk>. See also the Wikipedia entry at <http://bit.ly/nmmYf9>.

Boeing. 1997. "A Manifesto for Global Engineering Education: Summary of the Engineering Futures Conference," January 22-23, the Boeing Company and Rensselaer Polytechnic Institute. I've found no online version, but Karl Smith (2007) in a talk "Preparing Students for an Interdependent World: Role of Cooperation and Social Interdependence Theory" gives a list of the "Desired Attributes of a Global Engineer" from that report as reproduced in Section "VD3a" of this article.

Bok, D. 2005a. “Are colleges failing? Higher ed needs new lesson plans” *Boston Globe*, 18 December, freely online (possibly only for a short time) at <http://bo.st/ffRQFa>, and to educators at <http://bit.ly/fF954u> (scroll to the APPENDIX). Bok wrote:

> “. . . studies indicate that problem-based discussion, group study, and other forms of active learning produce greater gains in critical thinking than lectures, yet the lecture format is still the standard in most college classes, especially in large universities.”

See also Bok (2005b).

Bok, D. 2005b. *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/8648.html>. Amazon.com information is at <http://amzn.to/kXOKbF>, note the searchable “Look Inside” feature. See especially pages 115-116 on the research of Halloun & Hestenes (1985a,b) and pages 132-134 on the work of Eric Mazur (1997, 2008).

Bordogna, J. 1997. “Making Connections: The Role of Engineers and Engineering Education,” *The Bridge* **27**(1); on the web at <http://bit.ly/rgLNdh>.

Borenstein, S. 2011. “Study: It's not teacher, but method that matters,” *Associated Press*, 12 May, 14:27-0400; online at <http://bit.ly/rn3lFY>. Evidently Borenstein was the only reporter of the *Science* article by Deslauriers et al. who had sense enough to seek the opinion of a *physicist*, viz., Bob Beichner (instead of *psychologist*s or *cognitive scientists,* most of whom are blissfully unaware of what’s going on in physics education research).

†Bransford, J.D., A. L. Brown, & R.R. Cocking, eds. 2000. *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*, National Academy Press, online at <http://bit.ly/fVCQ6M>. The description reads:

"This popular trade book, originally released in hardcover in the Spring of 1999, has been newly expanded to show how the theories and insights from the original book can translate into actions and practice, now making a real connection between classroom activities and learning behavior. This paperback edition includes far-reaching suggestions for research that could increase the impact that classroom teaching has on actual learning. Like the original hardcover edition, this book offers exciting new research about the mind and the brain that provides answers to a number of compelling questions. When do infants begin to learn? How do experts learn and how is this different from non-experts? What can teachers and schools do-with curricula, classroom settings, and teaching methods--to help children learn most effectively? New evidence from many branches of science has significantly added to our understanding of what it means to know, from the neural processes that occur during learning to the influence of culture on what people see and absorb. *How People Learn* examines these findings and their implications for what we teach, how we teach it, and how we assess what our children learn. The book uses exemplary teaching to illustrate how approaches based on what we now know result in in-depth learning. This new knowledge calls into question concepts and practices firmly entrenched in our current education system. Topics include:

1. How learning actually changes the physical structure of the brain.
2. How existing knowledge affects what people notice and how they learn.
3. What the thought processes of experts tell us about how to teach.
4. The amazing learning potential of infants.
5. The relationship of classroom learning and everyday settings of community and workplace.
6. Learning needs and opportunities for teachers.
7. A realistic look at the role of technology in education."

See also *How Students Learn History, Mathematics, And Science In The Classroom* [Donovan & Bransford (2005)].

Breslow, L. 2010. "Wrestling With Pedagogical Change: The Teal Initiative at MIT," *Change Magazine*, September/October. A brief *Change Magazine* abstract is online at <http://bit.ly/nFCymK>.

Bressoud, D. 2011a. "The Worst Way to Teach," in the *Launchings* column of the "American Mathematical Association" of 1 July; online at <http://bit.ly/nmJJpj>. Bressoud wrote:

"This common belief . . . .[[in the efficacy of lectures]]. . . is also contradicted by the evidence that we have, the most recent and dramatic of which is 'Improved Learning in a Large-Enrollment Physics Class' [Deslauriers et al. (2011)] from the Carl Wieman Science Education Initiative (CWSEI) at the University of British Columbia (UBC). The CWSEI study compared lecture format with interactive, clicker-based peer instruction in two large (267 and 271 students, respectively) sections of introductory physics for engineering majors. The results were published in *Science* and made enough of a splash that they were reported by Carey (2011) in *The New York Times*, Mervis (2011) in *ScienceNOW*, and by *The Economist* (2011). What is most impressive is how well controlled the study was - ensuring that the two classes really were comparable - and how strong the outcome was: The clicker-based peer instruction class performed 2.5 standard deviations above the control group."

See also "Re: Lecture Isn't Effective: More Evidence" [Hake (2011h,i,j)

Bressoud, D. 2011b. “The Best Way to Learn,” in the *Launchings* column of the “American Mathematical Association” of 1 August; online at <http://bit.ly/qRHMCp>. Bressoud wrote:

“Last month, in ‘The Worst Way to Teach’ [Bressaud (2011a)] I wrote about some of the problems with instruction delivered by lecture. It stirred up a fair amount of discussion. Richard Hake (2011a) started a thread on the Math Forum list. . . . .[[more specifically the MathEdCC list]]. . . . He added several references to my own list and sparked a discussion that produced some heat and a lot of light. I do want to clarify that I recognize how important what I say in the classroom can be, as I will expound a bit later in this column. Nevertheless, I stand by my statement that ‘sitting still, listening to someone talk, and attempting to transcribe what they have said into a notebook is a very poor substitute for actively engaging with the material at hand, for doing mathematics.’ ”

Buck, J.R. & K.E. Wage. 2005. “Active and Cooperative Learning in Signal Processing Courses,” *IEEE Signal Processing Magazine* **22**(2): 76-81;online as a 217 kB pdf at <http://bit.ly/o3N1f9>.

Calvino, Italo. 1993. *Six Memos for the Next Millennium*, Harvard Univ. Press, publisher’s information at <http://bit.ly/pknCai>. Amazon.com information at <http://amzn.to/pNRAp4>, note the searchable “Look Inside” feature.

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 as a 196 kB pdf at <http://bit.ly/hMsyUr>.

Carey, B. 2011. “Less Talk, More Action: Improving Science Learning.” The *New York Times*, 13 May; online at <http://nyti.ms/mVDaWK>.

Chamberlin, T.C. 1890. “The Method of Multiple Working Hypotheses,” *Science* (old series) 15: 92-96; reprinted 1965, 148: 754-759; online at <http://bit.ly/p7Bcfd>.

An exception to the non-citing of Chamberlin by non-geologists is provided by the discerning biology education guru Anton Lawson. Lawson (1995), in his book *Science Teaching and the Development of Reasoning* features Chamberlin (1890) in his APPENDIX B, pp. 398-407 (but misspells Chamberlin as “Chamberlain”). Despite its *relative* neglect, Chamberlin's (1890) method is not forgotten as indicated by the fact that a Google <http://www.google.com/> search for [“The Method of Multiple Working Hypotheses” Chamberlin] (with the quotes but without the square brackets) turned up 14,800 hits at <http://bit.ly/l7xVwM> on 12 September 2011 13:13-0700.

COC. 2004. Council on Competitiveness, “Innovate America, National Innovation Initiative Summit and Report,” online as a 1.4 MB kB pdf at <http://bit.ly/oWGzFF>.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum, 2nd ed. Amazon.com information at <http://amzn.to/ogjYmj>.

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://bit.ly/nVizeU>. 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* [my *italics*].”

Coletta, V.P. and J.A. Phillips. 2005. “Interpreting FCI Scores: Normalized Gain, Preinstruction Scores, & Scientific Reasoning Ability,” *Am. J. Phys.* **73**(12): 1172-1182; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v73/i12>; an abstract is free to all. See also Coletta et al. (2007a,b) and Coletta & Phillips (2010).

Coletta, V.P., J.A. Phillips, & J.J. Steinert. 2007a. “Why You Should Measure Your Students' Reasoning Ability,” *Phys. Teach.* **45**(4): 235-238; online to subscribers at <http://tpt.aapt.org/resource/1/phteah/v45/i4>.

Coletta, V.P., J.A. Phillips, & J.J. Steinert. 2007b. “Interpreting force concept inventory scores: Normalized gain and SAT scores,” *Phys. Rev. ST Phys. Educ. Res.* **3**, 010106, Issue 1 - June; free online at <http://bit.ly/af0xQO>.

Coletta, V.P. & J.A. Phillips. 2010. “Developing Thinking and Problem Solving in Introductory Mechanics,” 2010 Physics Education Research Conference, p. 13-16; online as a 155 kB pdf at <http://bit.ly/o6d7oU>.

Collins, A. 1999. “The changing infrastructure of education research,” in E. C. Lagemann & L. S. Shulman, eds., *Issues in education research: Problems and possibilities*. Jossey-Bass, 1999; pp. 289-298.

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**(9): 970-977; online at <http://bit.ly/ppm3Bm>.

Crouch, C.H., J. Watkins, A. Fagen and E. Mazur. 2007. "Peer Instruction: Engaging students one-on-one, all at once" in *Reviews in Physics Education Research*, edited by E.F. Redish and P. Cooney; online at <http://bit.ly/qjR7TY>.

Cummings, K, J. Marx, R. Thornton, & D.Kuhl. 1999. “Evaluating innovations in studio physics,” *Phys. Educ. Res., Am. J.Phys. Suppl.* **67**, S38–S44; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v67/iS1>.

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>.

†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>.

Dancy, M. & C. Henderson. 2010. “Pedagogical practices and instructional change of physics faculty.” *Am. J. Phys.* **78**(10): 1056-1063; online as a 426 kB pdf at <http://bit.ly/9b4rX1>.

†De Graaf, E., G. Saunders-Smits, & M. Nieweg. 2006. *Research and Practice of Active Learning in Engineering Education*. Amsterdam University Press. Amazon.com information at <http://amzn.to/qKgh1l>.

DeHaan, R.L. 2005. “The Impending Revolution in Undergraduate Science Education,” *Journal of Science Education and Technology* **14**(2): 253-269; online as a 152 kB pdf at <http://bit.ly/ncAuQa>. DeHaan 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.”

DeHaan, R.L. 2009. “Teaching Creativity and Inventive Problem Solving in Science,” *CBE Life Sci Educ* **8**(3): 172-181; online at <http://bit.ly/btx32v>. Compare this with Adams’ (2006) “The Sources of Innovation and Creativity.”

Dellwo, D. 2009. “Reassessing Hake’s Gain,” preprint, Miller et al. (2010) state “available on request” (evidently on request to David R. Dellwo <dellwod@usmma.edu>).

Dellwo, D.R. 2010. “Course assessment using multi-stage pre/post testing and the components of normalized change,” *Journal of the Scholarship of Teaching and Learning* 10(1):, 55-67; online as a 1.0 MB pdf at <http://bit.ly/ckiuNJ>.

Deslauriers, L. E. Schelew, & C. Wieman. 2011. "Improved Learning in a Large-Enrollment Physics Class," *Science* **332** (6031): 862-864; an abstract is online at <http://bit.ly/oBxuWF>. Supplemental references are at <http://bit.ly/pDszky>.

This article has attracted widespread attention. For example, on 12 September 08:08-0700 a Google <http://www.google.com/> search for "Improved Learning in a Large-Enrollment Physics Class" (with the quotes) turned up 10,500 hits at <http://bit.ly/np18XA>. For news reports see Borenstein (2011), Carey (2010), *Economist* (2011), and Mervis (2011). For enthusiastic comment on by a respected mathematician see Bressoud 2011a,b). The non-physicists interviewed by the above news reporters appeared to be oblivious of any physics education research prior to Deslauriers et al., even despite references to such research: (1) in Deslauriers et al.; and (2) in articles in influential journals, e.g.:

(a) "Reintroducing the Intro Course" [Stockstad (2001)] in *Science*;
(b) "Teaching in a research context" [Wood & Gentile (2003)] in *Science*;
(c) "Scientific Teaching" [Handelsman et al. (2004)] in *Science*;
(d) "The Impending Revolution in Undergraduate Science Education" [DeHaan (2005)] in the *Journal of Science Education and Technology*;
(e) "Where's the evidence that active learning works?" [Michael (2006)] in *Advances in Physiology Education;*
(f) "Changing the Culture of Science Education at Research Universities" [Anderson et al. (2011)] in *Science*.

Ding, L., N.W Reay, A. Lee, & L. Bao. 2008. "Effects of testing conditions on conceptual survey results." *Physical Review Special Topics – Physics Education Research*, **4** (010112); online as a 299 kB pdf at <http://bit.ly/q41TVR>.

Ding, L., & R. Beichner. 2009. "Approaches to data analysis of multiple-choice questions," *Physical Review Special Topics – Physics Education Research*, **5** (020103); online as a 676 kB pdf at <http://bit.ly/qy3ER2>.

Docktor, J.L. & J.P. Mestre, 2011. "A Synthesis of Discipline-Based Education Research in Physics," online at <http://bit.ly/fOQWrc>. The online version carries no date and no indication of where it might be published. **:-(** . I suspect that it was commissioned by the NRC (2011d) Project: "Status, Contributions, and Future Directions of Discipline Based Education Research."

Donovan, M.S. & J. Pellegrino, eds. 2003. "Learning and Instruction: A SERP Research Agenda," National Academies Press; online at <http://books.nap.edu/catalog/10858.html>. The description reads:

"The Strategic Education Research Partnership (SERP) is a bold, ambitious plan that proposes a revolutionary program of education research and development. Its purpose is to construct a powerful knowledge base, derived from both research and practice, that will support the efforts of teachers, school administrators, colleges of education, and policy officials with the ultimate goal of significantly improving student learning. The proposals in this book have the potential to substantially improve the knowledge base that supports teaching and learning by pursuing answers to questions at the core of teaching practices. It calls for the linking of research and development, including instructional programs, assessment tools, teacher education programs, and materials. Best of all, the book provides a solid framework for a program of research and development that will be genuinely useful to classroom teachers."

†Donovan, M.S. & J.D. Bransford, eds. 2005. *How Students Learn History, Mathematics, And Science In The Classroom*. National Academies Press; online at <http://bit.ly/baJyu7>.

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://bit.ly/fbOeA8>. See also Rimer (2009) and Breslow (2010).

Dowd, J., K. Miller, B. Lukoff, N. Lasry, & E. Mazur. 2011. "'Stratagrams' for assessment of FCI performance: A 12,000-student analysis: Current conventions in assessing student performance can obscure 'gain' results," PERC 2011, an abstract is online at <http://bit.ly/rl6TOI>.

Duderstadt, J.J. 2000. *A University for the 21st Century*. Univ. of Michigan Press, publisher's information at <http://bit.ly/cvJ1yI>. Amazon.com information at <http://amzn.to/lxT8YU>. 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://bit.ly/nlsdX0>.

†Ebert-May, D. & J. Hodder. 2008. *Pathways to Scientific Teaching*. Sinauer Associates. Amazon.com information at <http://amzn.to/akFzgP>.

Economist. 2011. "An Alternative Vote: Applying Science to the Teaching of Science," *The Economist*, 14 May; online at <http://econ.st/qzIoPq>.

†Epstein, J. 2007. "Development and Validation of the Calculus Concept Inventory," in *Proceedings of the Ninth International Conference on Mathematics Education in a Global Community*, 7-12 September, edited by Pugalee, Rogerson, & Schinck; online as a 48 kB pdf at <http://bit.ly/bqKSWJ>.

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 InventoryAssessment Tools," *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference*; online as a 60 kB pdf at <http://bit.ly/p9k93i>.

Fairweather, J. 2008. "Linking Evidence and Promising Practices in Science, Technology, Engineering, and Mathematics (STEM) Undergraduate Education: A Status Report for The National Academies National Research Council Board of Science Education," online at <http://bit.ly/ePTL0W>. According to Labov et al. (2009), Fairweather "was asked to review and synthesize all of the additional articles submitted for the October workshop". . . .[[see National Academies (2008), and NRC (2011a)]]. . . .."

FLAG. 2011. “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>.

Flowers, W. 2000. “Why change, Been doin’ it this way for 4000 years!” ASME Mechanical Engineering Education Conference: Drivers and Strategies of Major Program Change, Fort Lauderdale, Florida, March 26-29, 2000; formerly online as a video at the ASU Center for Research on Education in Science, Mathematics, Engineering and Technology (CRESMET), but removed **:-(** after engineer Don Evans <http://bit.ly/oJ1ZzX>: retired from ASU and CRESMET was transformed by math educator Marilyn Carlson as described in a 2008 press release at <http://bit.ly/pgwy5W>.

Froyd, J.E. 2007. “Evidence for the Efficacy of Student-active Learning Pedagogies,” PKAL, online at <http://bit.ly/p0jD9Z>. Summary:

> “Although many resources have been published on improvements in student retention and/or learning as a result of using what can be referred to as student-active pedagogies, the resources are published in a variety of journals or on various websites. As a result, it may be difficult for an individual to locate and assemble these resources to support an argument in favor of using these alternative pedagogies. Over a period of eight years, including my time as the Project Director for the Foundation Coalition, one of the Engineering Education Coalitions supported by NSF, I have tried to assemble many of these resources in one place for easy reference.”

Froyd, J., A. Beach, C. Henderson, & N. Finkelstein. 2008. “Improving 'Educational Change Agents' Efficacy in Science, Engineering, and Mathematics Education,” in Hartman (2008, pp. 227-256); an abstract is online at <http://bit.ly/pkZC91>.

Froyd, J. 2011. “White Paper on Promising Practices in Undergraduate STEM Education,” online as a cc kB pdf at <http://bit.ly/lu1Q9b>. The abstract reads [see the paper for the references]:

> “Numerous calls for improvement in undergraduate science, technology, engineering, and mathematics (STEM) education have been made (National Academy of Engineering, 2004, 2005; National Research Council, 1999, 2003a, 2003b) and the Board of Science Education of the National Academies has established a committee to recommend a list of promising practices in undergraduate STEM education. To support the committee in its work, this white paper offers a set of promising practices using two sets of standards: implementation and student performance, against which each promising practice will be evaluated. Each promising practice offers an alternative for one or more of the decisions that faculty members or curriculum program committees make as they construct courses or curricula.”

Garvin, D.A. 2003. “Making the Case: Professional education for the world of practice” Harvard Magazine, September/October; online at <http://bit.ly/dVy9Uw>.

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

Goffe, W.L. 2011. “How and Why Academic Physicists Can Aid Learning on Their Campuses.” APS *Forum on Education*, Summer” online at <http://bit.ly/ojsHfV>.

Goldman, K.J., P.Gross, C. Heeren, G.L. Herman, L.C. Kaczmarczyk, M.C. Loui, & C.B. Zilles. 2010. “Setting the Scope of Concept Inventories for Introductory Computing Subjects,” *ACM Transactions on Computing Education* (TOCE) **10**(2); an abstract is online at <http://bit.ly/qlerXx>:

> “A concept inventory is a standardized assessment tool intended to evaluate a student’s understanding of the core concepts of a topic. In order to create a concept inventory it is necessary to accurately identify these core concepts. A Delphi process . . . . . . [[<http://en.wikipedia.org/wiki/Delphi_method> . . . . .]] is a structured multi-step process that uses a group of experts to achieve a consensus opinion. We present the results of three Delphi processes to identify topics that are important and difficult in each of three introductory computing subjects: discrete mathematics, programming fundamentals, and logic design. The topic rankings can not only be used to guide the coverage of concept inventories, but can also be used by instructors to identify what topics merit special attention.”

Gorsuch, R.L. 1983. *Factor Analysis*. Psychology Press, 2nd edition. Amazon.com information at <http://amzn.to/r1BY4H>, note the searchable “Look Inside” feature.

Gorsuch, R.L. 1990. “Common factor analysis versus component analysis: Some well and little known facts,” *Multivar. Behav. Res.* **25**: 33; the first page is online at <http://bit.ly/peISTU>; an abstract is online at <http://bit.ly/rqANrw>.

Hacker, A. & C. Dreifus. 2010. *Higher Education?: How Colleges Are Wasting Our Money and Failing Our Kids – and What We Can Do About It.* Holt/Times Books. Amazon.com information at <http://amzn.to/bunggt>. See also the Hacker/Dreifus blog <http://bit.ly/gJp4Pg>, carrying favorable comments from Barbara Ehrenreich, Diane Ravitch, Joseph Stiglitz, and Jonathan Kozol. At <http://bit.ly/nnnbZW> the publisher states:

> “In this provocative investigation, the renowned sociologist Andrew Hacker and *New York Times* writer Claudia Dreifus make an incisive case that the American way of higher education—now a $420 billion-per-year business—has lost sight of its primary mission: the education of our young people. . . . .*Higher Education?* is a wake-up call and a call to arms.”

Hake, R.R. 1967. “Paramagnetic Superconductivity in Extreme Type II Superconductors,” *Phys. Rev.* **158**(2): 356-376; an abstract is online at <http://bit.ly/nto3g4>.

Hake, R.R. 1987. “Promoting Student Crossover to the Newtonian World,” *Am. J. Phys.* **55**(10): 878-884; online as a 788 kB pdf at <http://bit.ly/a6vc3H>.

†Hake, R.R. 1992. “Socratic pedagogy in the introductory physics lab,” *Phys. Teach.* **30**: 546-552; updated version (4/27/98) online as a 88 kb pdf at <http://bit.ly/9tSTdB>.

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 as an 84 kB pdf at <http://bit.ly/9484DG> .

Hake, R.R. 1998b. "Interactive-engagement methods in introductory mechanics courses," online as a 108 kB pdf at <http://bit.ly/aH2JQN>. A crucial companion paper to Hake (1998a).

Hake, R.R. 2000. "What Can We Learn from the Physics Education Reform Effort?" ASME Mechanical Engineering Education Conference: "Drivers and Strategies of Major Program Change," Fort Lauderdale, Florida, March 26-29; online as a 446 kB pdf at <http://bit.ly/gDrN2T>.

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

Hake, R.R. 2002b. "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 as an 84 kB pdf at <http://bit.ly/9DPnL5>.

†Hake, R.R. 2002c. "References for Psychometrically Naive Physicists," PhysLrnR post of 6 Nov 2002 14:15:05-0800; online at <http://bit.ly/mRQjo7>.

Hake, R.R. 2002d. "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 as a 220 kB pdf at <http://bit.ly/aDOkEf>.

Hake, R.R. 2002e. "Re: Problems with Student Evaluations: Is Assessment the Remedy?" online as a 72 kB pdf at <http://bit.ly/hjt9ll>.

Hake, R.R. 2003a. "NRC's CUSE: Oblivious of the Advantage of Pre/Post Testing With High Quality Standardized Tests?" online on the OPEN! AERA-D archives at <http://bit.ly/pG7Bu2 >. Post of 26 Jul 2003 08:26:34-0700 to AERA-D and various other discussion lists. See also Hake (2003b).

Hake, R.R. 2003b. "NRC's CUSE: Stranded on Assessless Island?" online on the OPEN! AERA-D archives at <http://bit.ly/mScERO>. Post of 3 Aug 2003 15:07:25-0700 to AERA-D and various other discussion lists.

Hake, R.R. 2004a. "The Arons Advocated Method," submitted to the "*American Journal of Physics*" on 24 April 2004, but rejected by an editor who evidently believed a referee who erroneously claimed that ARONS DID NO PHYSICS EDUCATION RESEARCH ! (did ethnographer Margaret Mead <http://bit.ly/eSQat5> do no anthropological research?); online as a 144 kB pdf at <http://bit.ly/boeQQt>.

Hake, R.R. 2004b. "Design-Based Research: A Primer for Physics Education Researchers," submitted to the *Am. J. Phys*. on 10 June 2004; online as a 310 kB pdf at <http://bit.ly/kstWVC> (310kB); see Section II "The Insularity of Educational Research."

Hake, R.R. 2004c. “Re: pre-post testing in assessment,” online on the OPEN! AERA-J archives at <http://bit.ly/pmGWce>. POD post of 19 Aug 2004 13:56:07-0700 to AERA-J and various other discussion lists.

Hake, R.R. 2005a. "Will the No Child Left Behind Act Promote Direct Instruction of Science?" Am. Phys. Soc. 50: 851 (2005); APS March Meeting, Los Angles, CA. 21-25 March; online as a 25 kB pdf at <http://bit.ly/f26Da7>.

Hake, R.R. 2005b. “Re: Should Randomized Control Trials Be the Gold Standard of Educational Research? “ online on the OPEN! AERA-L archives at <http://bit.ly/qEnem3>. Post of 17 Apr 2005 20:30:46-0700 to AERA-L and various other discussion lists.

Hake, R.R. 2005c. “NRC Weighs In On States' Science Assessments,” online on the OPEN! AERA-L archives at <http://bit.ly/qI9piG>. Post of 14 Jul 2005 13:32:37-0700 to AERA-L and various other discussion lists.

Hake, R.R. 2005d. “Do Psychologists Research the Effectiveness of Their Courses? Hake Responds to Sternberg,” online on the OPEN! AERA-L archives at <http://bit.ly/mZaZKc>. Post of 21 Jul 2005 22:55:31-0700 to AERA-L and various other discussion lists.

Hake, R.R. 2005e. “The Physics Education Reform Effort: A Possible Model for Higher Education?” online as a 100 kB pdf at <http://bit.ly/9aicfh>; a slightly edited version of the article that was: (a) published in the *National Teaching and Learning Forum* (NTLF) **15**(1), December 2005, online to subscribers at <http://bit.ly/bvm8Ye> (If your institution doesn't subscribe to NTLF, it *should*); (b) disseminated in *Tomorrow's Professor* Msg. #698 on 14 Feb 2006 archived at <http://bit.ly/d09Y8r> - type the message number into the slot at the top of the page.

Hake, R.R. 2006a. “Eleven Quotes in Honor of Inertia,” online on the OPEN! AERA-L archives at <http://bit.ly/g7jdeU>. Post of 13 Jun 2006 15:01:14-0700 to AERA-L, PhysLrnR, and POD.

Hake, R.R. 2006b. “SET’s Are Not Valid Gauges of Teaching Performance," online on the OPEN! AERA-L archives at <http://bit.ly/p3uh2K>. Post of 20 Jun 2006 07:50:58-0700 to AREA-L and various other discussion lists.

Hake, R.R. 2006c. “SET’s Are Not Valid Gauges of Teaching Performance #2,” online on the OPEN! AERA-L archives at <http://bit.ly/poBg28>. Post of 21 Jun 2006 21:12:19-0700 to AREA-L and various other discussion lists, in which I respond to 12 points made by Michael Scriven in his response to to Hake (2006a).

Hake, R.R. 2006d. “SET’s Are Not Valid Gauges of Teaching Performance #3,” online on the OPEN! AERA-L archives at <http://bit.ly/oHuu27>. Post of 25 Jun 2006 20:58:34-0700 to AREA-L and various other discussion lists, in which I respond to 5 points made by Wilbert (Bill) McKeachie (WM) in his response to Hake (2006a).

Hake, R.R. 2006e. “SET’s Are Not Valid Gauges of Teaching Performance #3 - Correction and Clarification,” online on the OPEN! AERA-L archives at <http://bit.ly/oX2rBp>. Post of 26 Jun 2006 13:48:24-0700 to AREA-L and various other discussion lists, in which I correct some deficiencies brought to my attention by Ed Nuhfer of the POD list.

Hake, R.R. 2006f. “SET’s Are Not Valid Gauges of Teaching Performance #4,” online on the OPEN! AERA-L archives at<http://bit.ly/qtJvBd>. Post of 27 Jun 2006 17:24:18-0700 to AREA-L and various other discussion lists, in which I respond to 6 points made by Michael Theall in his response to Hake (2006a).

Hake, R.R. 2006g. "Re: Ceiling Effects: Performance and Instrumental, online on the OPEN! AERA-L archives at <http://bit.ly/q48WmP>. Post of 19 Jul 2006 10:43:57 -0700 to AERA-L and various other discussion lists.

Hake, R.R. 2006h. "Re: Ceiling Effects: Performance and Instrumental-Addenda & Apology, online on the OPEN! AERA-L archives at < http://bit.ly/p0KZ9A>. Post of 19 Jul 2006 17:37:36 -0700 to AERA-L and various other discussion lists.

Hake, R.R. 2006i. "Re: Ceiling Effects: Performance and Instrumental #2, online on the OPEN! AERA-L archives at <http://bit.ly/nbpFdG>. Post of 20 Jul 2006 12:49:08 -0700 to AERA-L and various other discussion lists. http://bit.ly/nbpFdG

Hake, R.R. 2006j. "Re: Ceiling Effects: Performance and Instrumental #3," online on the OPEN! AERA-L archives at <http://bit.ly/ncfyiL>. Post of 21 Jul 2006 14:16:12-0700 to AERA-L and various other discussion lists.

Hake, R.R. 2006k. "Possible Palliatives for the Paralyzing Pre/Post Paranoia that Plagues Some PEP's," *Journal of MultiDisciplinary Evaluation*, Number 6, November, online at <http://bit.ly/caWtWl>. 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.” This was a warmup for "Should We Measure Change? Yes!" [Hake (2011b)].

Hake, R.R. 2007a. “Can Scientific Research Enhance the Art of Teaching?” invited talk, AAPT Greensboro meeting, 31 July, online as a 1.2 MB pdf at <http://bit.ly/a7xJxR>. See esp. Sect. V. “University Leaders Bemoan the Inertia of Higher Education: Why Is It So Slow To Recognize the Value of Interactive Engagement Methods in Promoting Higher-Level Learning?”

Hake, R.R. 2007b. “Cognitive Science and Physics Education Research: What We've Got Here Is a Failure to Communicate,” submitted to the *Journal of Learning Sciences* (JLS) on 10 October 2007; online as a 602 kB pdf at <http://bit.ly/azCRG8>. Rejected by JLS editors and referees who objected to (a) “casting the issue specifically between Cognitive Science and Physics Education Research,” (b) my position that the “problems of ambiguity and inconsistency in the terms and constructs at play in education research can be addressed by researchers reading across fields.”

Hake, R.R. 2007c. “Six Lessons From the Physics Education Reform Effort,” *Latin American Journal of Physics Education* **1**(1), September; online (with AIP style numbered references) as a 124 kB pdf at <http://bit.ly/bjvDOb> (124 kB). Also available with APA style references as a 684 kB pdf at <http://bit.ly/96FWmE> (684 kB).

Hake, R.R. 2008a. “Design-Based Research in Physics Education Research: A Review,” in Kelly, Lesh, & Baek (2008)]. A pre-publication version of that chapter is online as a 1.1 MB pdf at <http://bit.ly/9kORMZ>.

†Hake, R.R. 2008b. “Can Distance and Classroom Learning Be Increased?” *IJ-SoTL* **2**(1): January; online at <http://bit.ly/98dL0Y>. The *International Journal of Scholarship of Teaching and Learning* (IJ-SoTL) <http://www.georgiasouthern.edu/ijsotl/> is an open, peer-reviewed, international electronic journal containing articles, essays, and discussions about the scholarship of teaching and learning (SoTL) and its applications in higher/tertiary education today.

Hake, R.R. 2008c. “What's Statistically Acceptable? (was measure of gain),” online on the OPEN! AERA-D archives at <http://bit.ly/o7M807>. Post of 23 Jun 2008 17:02:44-0700 to AERA-D and various other discussion lists. See also Hake (2008e,f). Hake (2008c) misspells “Ramlo” as “Romlo” - the fault of my lousy spiel chequer ;-) [see e.g., “Candidate for a Pullet Surprise” [Zar (1994)].

Hake, R.R. 2008d. “Language Ambiguities in Education Research,” submitted to the *Journal of Learning Sciences* on 21 August but rejected - proof positive of an exemplary article; online at <http://bit.ly/bHTebD>.
David Klahr <http://en.wikipedia.org/wiki/David_Klahr> wrote to me privately (quoted by permission): “I liked the paper. I think it's very thoughtful and nuanced. However it is tough going, even for someone as familiar with the issues (and as favorably cited by you) as I am. It's a shame that it was rejected, but I wonder if the reviewer just wasn't up to the very careful reading necessary to really follow your arguments all the way through. Even though I know this area quite well, obviously, I did have to really focus to fully understand the distinctions you were making.”

Hake, R.R. 2008e. “Can Pre-to-posttest Gains Gauge Course Effectiveness?” online on the OPEN! AERA-D archives at <http://bit.ly/qhhyAx>. Post of 18 Oct 2008 12:05:53-0700 to AERA-D, and various other discussion list. The abstract reads:
“In a PhysLrnR post Irfaan Sorathia asked 'what constitutes a significant change in a students pre/post test results. . . [[on the “Force Motion Conceptual Evaluation”]]. . . .. I reference: (a) sources that explain the normalized gain g; (b) Sue Ramlo's (2008a,b) dismissal of *g* and my counter; (c) Ramlo's (2008c) recent AJP article that concludes ‘the FMCE is a valuable instrument for measuring student learning,’ but essentially ignores the crucial question as to whether or not pre-to-posttest gains on the FMCE can be used to gauge the instructional effectiveness of courses in promoting an understanding of Newtonian mechanics; and (d) Lissitz & Samuelsen’s (2007) recent *Educational Researcher* article that ‘raises a number of questions about the current unified theory of test validity that has construct validity at its center’ (as Ramlo appears to place it).”

Hake, R.R. 2008f. “Can Pre-to-posttest Gains Gauge Course Effectiveness? #2,” online on the OPEN! AERA-L archives at <http://bit.ly/mRBgdu>. Post of 19 Oct 2008 16:08:08-0700 to AERA-L, and various other discussion list.

†Hake, R.R. & J.V. Mallow. 2008. “Gender Issues in Science/Math Education (GISME),” over 700 Annotated References & 1000 URL’s: Part 1 - All references in alphabetical order; Part 2 - Some references in subject order; both online at ref. 55 at <http://www.physics.indiana.edu/~hake/>; see also my “Hake’sEdStuff” blog entry at <http://bit.ly/fSkr6W>.

Hake, R.R. 2009a. “Over Two-Hundred Annotated References on Systems Thinking,” online as a 1.7 MB pdf at <http://bit.ly/9gZdXU >. The abstract and link to the complete post are on my website “Hake's Ed Stuff” at <http://bit.ly/duhqLE> with a provision for comments For other related posts go to “Hake's Ed Stuff” at <http://HakesEdStuff.blogspot.com/> and scroll down to and click on “Systems Thinking” under “Labels” in the right-hand column.

Hake, R.R. 2009b. “The Threat to Life on Planet Earth Is a More Important Issue Than David Brooks’ Skills Slowdown,” online on my blog “Hake’sEdStuff” at <http://bit.ly/cj2zUS> with a provision for comments.

Hake, R.R. 2009c. “Re: Lawson and FCI,” online on the PhysLrnR archives at <http://bit.ly/qxrlUo>. Post of 12 May 2009 11:18:31-0700 to PhysLrnR. In the ERRATUM at <http://bit.ly/mYv2YA> I give the correct the reference to Hake (1998b) – see this reference list - that was garbled in Hake (2009c).

Hake, R.R. 2009d. “Re: Student Engagement,” online on the OPEN! AERA-L archives at <http://bit.ly/qB93tC>. Post of 8 Dec 2009 10:59:49-0800 to AERA-L, AP-Physics, Net-Gold, Physhare, Phys-L, PhysLrnR, & Physoc. The abstract was sent to several other discussion lists.

†Hake, R.R. 2009e. “Over Two-Hundred Education & Science Blogs,” 30 March; online as a 2.6 MB pdf at <http://bit.ly/adz77c>. The abstract is online with a provision for comments on my blog “Hake'sEdStuff” at <http://bit.ly/aNVVHq>. There are now 197 comments, mostly SPAM! which I’m currently trying to eliminate.

Hake, R.R. 2010a. “Re: Quality Research in Literacy and Science Education: International Perspectives and Gold Standards,” online on the OPEN! AERA-L archives at <http://bit.ly/qe1y8o>. Post of 22 Feb 2010 14:04:43-0800 to AERA-L and Net-Gold. The abstract and link to the complete post was sent to various discussion lists), and are also online on my blog “Hake's Ed Stuff” at <http://bit.ly/aAzTRT> with a provision for comments.

Hake, R.R. 2010b. “Why Password Protection of Concept Tests is Critical: A Response to Klymkowsky,” online on the OPEN! AERA-L archives at <http://bit.ly/pD2ULs>. Post of 4 Jun 2010 15:56:17-0700 to AERA-L and Net-Gold. The abstract and link to the complete post were also transmitted to various discussion lists and are also on my blog “Hake'sEdStuff” at <http://bit.ly/boAuCx> with a provision for comments.

Hake, R.R. 2010c. “Is Psychometrics Pathological Science?” online on the OPEN AERA-L archives at <http://bit.ly/9Fp6QS> . Post of 17 Jun 2010 15:34:11-0700 to AERA-L, Net-Gold, and PhysLrnR. The abstract and URL was transmitted to various discussion lists and are on my blog “Hake’sEdStuff” at <http://bit.ly/qhI9SH> with a provision for comments.

Hake, R.R. 2010d. “Education Research Employing Operational Definitions Can Enhance the Teaching Art,” invited talk, Portland, OR, AAPT meeting, 19 July; online as a 3.8 MB pdf at <http://bit.ly/aGlkjm> and as ref. 60 at <http://www.physics.indiana.edu/~hake>.

Hake, R.R. 2010e. “Helping Students to Think Like Scientists in Socratic Dialogue Inducing Labs,” submitted to *The Physics Teacher* on 19 August; online as a 446 kB pdf at <http://bit.ly/99yb7p> and as ref. 62 at <http://www.physics.indiana.edu/~hake>. Accepted for publication - to be published in Fall 2011.

Hake, R.R. 2010f. “Dukenfield's Law & Campbell's Law,” online on the OPEN! AERA-L archives at <http://bit.ly/9FWI9n>. Post of 14 Aug 2010 20:52:11-0700 to AERA-L, and various other discussion lists.

Hake, R.R. 2010g. “Dukenfield's Law & Campbell's Law #2,” online on the OPEN! AERA-L archives at <http://bit.ly/d3FrI8>. Post of 22 Aug 2010 15:31:31-0700 to AERA-L and Net-Gold. The abstract and link to the complete post were also transmitted to various discussion lists and are also on my blog “Hake'sEdStuff” at <http://bit.ly/f9Maey> with a provision for comments.

Hake, R.R. 2010h. “Re: Errata for FCI?” online on the OPEN! Phys-L archives at <http://bit.ly/oBDX2u>. Post of 3 Nov 2010 16:03:01-0700 to Phys-L and PhysLrnR.

Hake, R.R. 2011a. Supporting online references for Anderson et al. (2011) at <http://bit.ly/g24Iqm> [over 75 references and 90 hot links relevant to (a) Undergraduate Education Reform, and (b) Biology Education Reform].

Hake, R.R. 2011b. “Should We Measure Change? Yes!” online as a 2.5 MB pdf at <http://bit.ly/d6WVKO> (2.5 MB). To appear as a chapter in "Rethinking Education as a Science" [Hake (in preparation)]. A severely truncated version appears in Hake (2006e). The abstract reads:

> “Formative pre/post testing is being successfully employed to improve the effectiveness of courses in undergraduate astronomy, biology, chemistry, economics, engineering, geoscience, math, 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.”

Hake, R.R. 2011c. “Academically Adrift?” online on the OPEN! AERA-L archives at <http://bit.ly/gwJD0W>. Post of 29 Jan 2011 10:00:09-0800 to AERA-L and Net-Gold. The abstract and link to the complete post are being transmitted to various discussion lists are also online on my blog “Hake'sEdStuff” at <http://bit.ly/hVYzHI> with a provision for comments.

Hake, R.R. 2011d. “Physics Education Research - Not Widely Known in Higher Education,” online on the OPEN! AERA-L archives at <http://bit.ly/iT4YsN>. Post of 27 Apr 2011 17:07:07-0700 to AERA-L and Net-Gold. The abstract and link to the complete post were transmitted to various discussion lists and are also on my blog “Hake'sEdStuff” at <http://bit.ly/msoLwx> with a provision for comments.

Hake, R.R. 2011e. “Physics Education Research - Not Widely Known in Higher Education #2,” online on the OPEN! AERA-L archives at <http://bit.ly/lNZe6Z>. Post of 30 Apr 2011 13:49:34-0700 to AERA-L and Net-Gold. The abstract and link to the complete post are being transmitted to various discussion lists and are also on my blog “Hake'sEdStuff” at <http://bit.ly/mozc8W> with a provision for comments.

Hake, R.R. 2011f. Invited talk, “Should the Culture of University Science Education Be Changed ?” Southern California Section of the AAPT, 14 May 2011, Pierce College, Woodland Hills, CA; online as a 3.2 MB pdf at <http://bit.ly/iegznz> and as reference #63 at <http://bit.ly/b2UsK6>. The abstract and link to the complete post are on my blog “Hake'sEdStuff” at <http://bit.ly/lCyN97> with a provision for comments.

Hake, R.R. 2011g. “SET's Are Not Valid Gauges of Students’ Higher-Level Learning #2," online on the OPEN! AERA-L archives at <http://bit.ly/jLZaz5>. Post of 17 May 2011 09:47:36-0700 to AERA-L and Net-Gold. The abstract and link to the complete post are also being distributed to various discussion lists and are also on my blog “Hake'sEdStuff” at <http://bit.ly/ixcQxs> with a provision for comments.

Hake, R.R. 2011h. “Re: Lecture Isn’t Effective: More Evidence,” online on the OPEN! MathEdCC archives at <http://bit.ly/r80W5i> along with 10 responses (as of 22 July 2011). Post of 15 July, shamelessly cross-posted to Math-Teach, Math-Learn, MathEdCC, and RUME. See also Hake (2011b,c).

Hake, R.R. 2011i. “Re: Lecture Isn't Effective: More Evidence #2,” online on the OPEN! AERA-L archives at <http://bit.ly/mXiXoh>. Post of 20 Jul 2011 17:13:46-0400 to AERA-L and Net-Gold. The abstract and link to the complete post were transmitted to various discussion lists and are also on my blog “Hake'sEdStuff” at <http://bit.ly/rr2BQU> with a provision for comments. See also the precursor Hake (2011h).

Hake, R.R. 2011j. “Re: Lecture Isn't Effective: More Evidence #4,” online on the OPEN! AERA-L archives at <http://bit.ly/oVqvml>. Post of 22 Jul 2011 14:52:44 -0700 to AERA-L and Net-Gold. The abstract and link to the complete post are being transmitted to various discussion lists.

†Hake, R.R. 2011k. “Re: Research on Active Learning in Higher Education,” online on the OPEN! AERA-L archives at <http://bit.ly/pUM8jG>. Post of 28 Jul 2011 17:04:29 -0700 to AERA-L and Net-Gold. The abstract and link to the complete post wete transmitted to various discussion lists and are also on my blog “Hake'sEdStuff” at <http://bit.ly/pRxxqu> with a provision for comments.

Hake, R.R. 2011m. "Re: Learning Physics with a Statistical Gold Standard" online on the OPEN! AERA-L archives at <http://bit.ly/nlyfCC>. Post of 24 Jul 2011 11:12:13-0700 to AERA-L and Net-Gold. The abstract and link to the complete post were transmitted to various discussion lists and are also on my blog “Hake'sEdStuff” at <http://bit.ly/q9Snn6> with a provision for comments.

†Hake, R.R. 2011n. “A Guide to the ADLsphere: Over Eighty Academic Discussion Lists With Links To Their Archives & Search Engines,” online as a 3.9 MB pdf at <http://bit.ly/970OZr> and as ref. 61 at <http://bit.ly/a6M5y0>.

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 as a 123 kB pdf at <http://bit.ly/r4cF8Z>. The first paragraph of the Foreword by E.F. Taylor reads [my *italics*]:

> “It is difficult to report on some kinds of knowledge. For example, how do you summarize in useful form the wealth of lore and insight that has come from many recent attempts to install educational innovations in colleges and universities? In practice, *innovation is more an art than a science*, and the academic innovator must be at least as much a politician as a scholar. *Persons initially opposed to the birth of a new program are often unmoved by logical argument and studied rationality*. How then can passage to final acceptance be described as a logically rational process? We have concluded that it cannot.”

Halloun, I. & Hestenes, D. 1985a. “The initial knowledge state of college physics,” *Am. J. Phys.* **53**(11): 1043-1055; online at <http://bit.ly/b1488v>, scroll down to “Evaluation Instruments.”

Halloun, I. & D. Hestenes. 1985b. “Common sense concepts about motion.” *Am. J. Phys.* **53**(11) 1056-1065; online at <http://bit.ly/b1488v>, scroll down to “Evaluation Instruments.”

Halloun, I. & D. Hestenes. 1987. “Modeling instruction in mechanics,” *Am. J. Phys.* **55**(5): 455-462; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v55/i5>.

Halloun, I. & D. Hestenes. 1995. “The Search for Conceptual Coherence in FCI data,” evidently unpublished, online as a 16 kB pdf at <http://bit.ly/nsHI29>.

Halloun, I., R.R. Hake, E.P. Mosca, & D. Hestenes. 1995. “Force Concept Inventory (1995 Revision),” online (password protected) at <http://bit.ly/b1488v>, scroll down to "Evaluation Instruments." Currently available in 20! languages: Arabic, Chinese, Croatian, Czech, English, Finnish, French, French (Canadian), German, Greek, Italian, Japanese, Malaysian, Persian, Portuguese, Russian, Spanish, Slovak, Swedish, & Turkish.

Halloun, I. 2004. *Modeling Theory in Science Education*. Springer, publisher’s information at <http://bit.ly/orN7y1>. Amazon.com information at <http://amzn.to/pypJve>. Note the searchable “Look Inside” feature.

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 as a 100 kB pdf at <http://bit.ly/aQbF8G>. See also the supporting material online as a 344 kB pdf at <http://bit.ly/eOCJmo> [URL's are specified for some, but (unfortunately) not all, online materials].

†Handelsman, J., S. Miller, & C. Pfund, 2007. *Scientific Teaching*, Freeman. Amazon.com information at <http://amzn.to/aKssNI>.

Hartman, H., ed. 2008. "*Integrating the Sciences and Society: Challenges, Practices, and Potentials*, Emerald Group Publishing, publisher's information at <http://bit.ly/qVnmOI>. Amazon.com information at <http://amzn.to/nOcd8E>, note the searchable "Look Inside" feature.

†Heckler, A.F. & E. C. Sayre. 2010. "What happens between pre- and post-tests: multiple measurements of student understanding during an introductory physics course. " *Am. J. Phys.* **78**(7): 768-777; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v78/i7>.

Heller. P. & D. Huffman. 1995. "Interpreting the Force Concept Inventory: A Reply to Hestenes and Halloun," *Phys. Teach.* **33**(8): 503, 507-511 (1995); online to subscribers at <http://tpt.aapt.org/resource/1/phteah/v33/i8>.

Henderson, C. 2002. "Common Concerns About the Force Concept Inventory," *Phys. Teach.* **40**(9): 542-547; online to subscribers at <http://tpt.aapt.org/resource/1/phteah/v40/i9>.

Henderson, C. 2007. "Promoting instructional change in new faculty: An evaluation of the physics and astronomy new faculty workshop," *Am. J. Phys.* **76**(2): 179-187;online as a 397 kB pdf at <http://bit.ly/qieA6s>.

Henderson, C. & M. Dancy. 2007. "Barriers to the Use of Research-Based Instructional Strategies: The Influence of Both Individual and Situational Characteristics," Phys. Rev. ST Phys. Educ. Res. **3**(2), 020102); online as a 270 kB pdf at <http://bit.ly/qVqVNG>.

Henderson, C. & M. Dancy. 2008. "Physics Faculty and Educational Researchers: Divergent Expectations as Barriers to the Diffusion of Innovations," A*m. J. Phys. (Phys. Ed. Res.)* **76**(1): 79-91; online as a 270 kB pdf at <http://bit.ly/nuPJSn>.

Henderson, C. & M. Dancy. 2009. "Impact of physics education research on the teaching of introductory quantitative physics in the United States," *Phys. Rev. ST Phys. Educ. Res.* **5**(2), 020107; online at <http://bit.ly/oKdtCV>.

Henderson, C., N. Finkelstein, & A. Beach. 2010. "Beyond Dissemination in College science teaching: An Introduction to Four Core Change Strategies," *J. Coll. Sci. Teach.* **39**(5), 18-25l online as a 279 kB pdf at <http://bit.ly/nMHDAi>.

Henderson, C., A. Beach, & N. Finkelstein. 2011. "Facilitating Change in Undergraduate STEM Instructional Practices: An Analytic Review of the Literature," accepted for publication in the *J. Res. Sci. Teach.* ; online as a 197 kB pdf at <http://bit.ly/oBSu5v> (supplementary materials to accompany article: Spreadsheet_S1.xls)

Henderson, C., A. Beach, & N. Finkelstein. 2011. "Four Categories of Change Strategies for Transforming Undergraduate Instruction," submitted in September 2009 to *Transitions, Transformations, and Transgressions in Learning and Education*, edited by Päivi Tynjälä, Marja-Leena Stenström & Marjatta Saarnivaara; accepted for publication; online as a 250 kB pdf at <http://bit.ly/qwKdYC>.

Henderson, C., M. Dancy, & M. Niewiadomska-Bugaj. 2011. "The Use of Research-Based Instructional Strategies in Introductory Physics: Where do Faculty Leave the Innovation-Decision Process?" submitted; online as a 209 kB pdf at <http://bit.ly/p1o25t>.

†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 as a 56 kB pdf at <http://bit.ly/axznvY>.

Hestenes, D. 1987. "Toward a Modeling Theory of Physics Instruction," *Am. J. Phys.* **55**(5): 440-454; online at <http://bit.ly/b1488v>. Scroll down to "Modeling Instruction: Theory and Practice."

Hestenes, D. 1992. "Modeling Games in the Newtonian World," *Am. J. Phys.* **60**(8): 732-748; online at <http://bit.ly/b1488v>. Scroll down to "Modeling Instruction: Theory and Practice."

Hestenes, D., M. Wells, & G. Swackhamer. 1992. "Force Concept Inventory," *Phys. Teach.* **30**(3): 141-158; online as a 100 kB pdf at <http://bit.ly/foWmEb> [but without the test itself]. For the 1995 revision see Halloun et al. (1995).

Hestenes, D. 1997. "Modeling Methodology for Physics Teachers, " in *The Changing Role of Physics Departments in Modern Universities: Proceedings of the ICUPE,* ed. by E.F. Redish and J.S. Rigden, AIP, pp. 935- 957; online as a 98 kB pdf at <http://bit.ly/qMJWV9>.

Hestenes, D. & I. Halloun. 1995. "Interpreting the Force Concept Inventory: A Response to March 1995 Critique by Huffman and Heller," Phys. Teach. **33**, 502, 504-506; online to subscribers at <http://tpt.aapt.org/resource/1/phteah/v33/i8>.

Hestenes, D. 2005. Private communication to Vincent Coletta, August, as indicated in Coletta et al. (2007a).

Hestenes, D., C. Megowan-Romanowicz, S.E. Osborn Popp, J. Jackson, and R.J. Culbertson. 2011. "A graduate program for high school physics and physical science teachers," *Am. J. Phys.* **79**(9): 971-979; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias>.

Hilborn, R.C. 2005. “Academic Schizophrenia, The Scholarship of Teaching, and STEM Leadership,” Project Kaleidoscope: What works, what matters, what lasts; online as a 192 kB pdf at <http://bit.ly/nVe1DE>.

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): 459-462; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v73/i5>.

Hoellwarth, C. & M. J. Moelter. 2011. “The implications of a robust curriculum in introductory mechanics.” Am. J. Phys. **79**(5): 540-545; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v79/i5>.

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.

†Hrepic, Z., D.A Zollman, & N.S. Rebello. 2007. “Comparing students’ and experts’ understanding of the content of a lecture,” J. *Sci. Educ. Technol* **16**: 213–224; online as a 98 kB pdf at <http://bit.ly/rcEax7>. Among their conclusions was:

> “This study has shown that a variety of misunderstandings can occur during the lecture even if the conditions for learning are far better than in a typical classroom. From this perspective, the answer to the question, ‘Do they just sit there?’ is – no, they don’t! Instead they go through intensive cognitive processes which, unfortunately, may be very different from the instructor’s intentions and expectations. We see evidence that in a traditional lecture setting, with one-way instruction, these cognitive processes have to be extremely carefully guided and monitored. If possible, *feedback from students should be sought along the way*. . . . [[my *italics*]]. . . . either directly or through interactive technologies (e.g. classroom response systems). In this respect we suggest, at a minimum, inclusion of interactive lecture methods such as Peer Instruction (Mazur, 1997) to facilitate learning through students’ engagement even in the largest auditoriums.”

Huffman, D. & P. Heller. 1995. “What Does the Force Concept Inventory Actually Measure?” *Phys. Teach.* **33**(3): 138-143; online to subscribers at <http://tpt.aapt.org/resource/1/phteah/v33/i3>.

IEAC. 1996. Institution of Engineers, Australia, Canbera, Review of Engineering Education, “Changing the Culture: Engineering Education into the Future, Review Report,” Institution of Engineers, Australia, Canberra. I’ve not found an online version, but “Engineers for the Future: addressing the supply and quality of Australian engineering graduates for the 21st century” [King (2008)] reviews the recommendations of that report.

JCSEE. 1994. Joint Committee on Standards for Educational Evaluation, *The Program Evaluation Standards*, 2nd ed., Sage. A glossary of evaluation terms from this publication was formerly online at <http://ec.wmich.edu/glossary/prog-glossary.htf> but has evidently been removed in recognition of the 3rd edition of the Evaluation Standards by Yarbrough et al. (2011).

Junker, B. 1999. “Some statistical models and computational methods that may be useful for cognitively-relevant assessment,” paper prepared for the NRC Committee on the Foundations of Assessment; online at <http://bit.ly/qkJ6w9> - Scroll down to “View the draft paper in PDF format click on “PDF format” to obtain a 565 kB pdf at <http://bit.ly/oSbUPJ>; obtained from Pellegrino et al. (2001, page 329) (Other interesting papers by Brian Junker are online at <http://bit.ly/oEeQlZ>.

Kelly, A.E., R.A. Lesh, & J.Y. Baek. 2008. "Handbook of Design Research Methods in Education: Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching," Routledge. Publisher's information at <http://bit.ly/dkLabI>; Amazon.com information at <http://amzn.to/gtRpbU>.

†Keiner, L.E. & T.E. Burns, 2010. “Interactive Engagement: How Much Is Enough?” *Phys. Teach.* **48**(2): 108- 111; online to subscribers at <http://tpt.aapt.org/resource/1/phteah/v48/i2>.

King, R. 2008. “Engineers for the Future: addressing the supply and quality of Australian engineering graduates for the 21st century,” Australian Council of Engineering Deans (ACED); online at <http://bit.ly/mTNbRo>.

Kim, J. & C. Mueller. 1978. *Factor Analysis: Statistical Methods and Practical Issues*. Sage Publications. Amazon.com information at <http://amzn.to/raop4I>, note the searchable “Look Inside” feature.

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 as a 176 kB pdf at <http://bit.ly/duJVG4> .

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 as a 388 kB pdf at <http://tinyurl.com/2kho83>. For a discussion of the widespread misinterpretation of this paper see Klahr & Li (2005) and Hake (2005a, 2008d).

Klahr, D. & J. Li. 2005. “Cognitive Research and Elementary Science Instruction: From the Laboratory, to the Classroom, and Back,” *Journal of Science Education and Technology* **14**(2): 217-238; online as a 536 kB pdf at <http://bit.ly/apA7es>.

Kleiman, M. 2010. “Dukenfield’s Law of Incentive Management,” *The Atlantic*, 13 August 2010; online at <http://bit.ly/bsRokM>.

Knight, J.K., & W.B. Wood. 2005. “Teaching more by lecturing less,” *Cell Biology Education* **4**: 298-310; online at <http://bit.ly/nADMd4>.

Lasry, N., E. Mazur, & J. Watkins. 2008. “Peer instruction: From Harvard to the two-year college,” *Am. J. Phys.* **76**(11): 1066-1069; online <http://bit.ly/oSzUtw>.

Lasry, N., S. Rosenfield, H. Dedic, A. Dahan, & O. Reshef. 2011. “The puzzling reliability of the Force Concept Inventory,” *Am. J. Phys.* **79**(9): 909-912; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias>.

Labov, J.B. 2003. “Education at the National Academies,” *Cell Biology Education* **2**(3); online at <http://bit.ly/oWcV4O>.

Labov, J.B., S.R. Singer, M.D. George, H.A. Schweingruber, & M.L. Hilton. 2009. "Effective Practices in Undergraduate STEM Education Part 1: Examining the Evidence," CBE Life Sci Educ 8(3): 157-161; online at <http://bit.ly/cRc0JC>.

This is a discussion of the "Workshop on Linking Evidence and Promising Practices in STEM Undergraduate Education" [National Academies (2008)]. A Google search on 19 July 2011 suggests that "Part 2,"if written, has not been published. See also the NRC (2011a) book “Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics Education: Summary of Two Workshops.”

Lagemann, E.C. 2000. *An Elusive Science: The Troubling History of Educational Research*. University of Chicago Press, publisher’s in information at <http://bit.ly/onL7Rj>. Amazon.com information at <http://amzn.to/o7AAFU>, note the searchable “Look Inside” feature.

Lawson, A.E. 1978. “The development and validation of a classroom test of formal reasoning,” *J. Res. Sci. Teach.* **15**(1): 11–24. The first page is online at <http://bit.ly/nEKXUj>. A multiple-choice version is online in the Appendix of Coletta & Phillips (2005_.

Lawson, A.E. 1992. “What do tests of 'formal' reasoning actually measure?” *Journal of Research in Science Teaching* **29**(9): 965–83; an abstract is online at <http://bit.ly/9HEZ3e>.

Lawson, A.E. 1995. *Science Teaching and the Development of Thinking*. Wadsworth. Amazon.com information at <http://amzn.to/aqWhQ2>, note the “Look Inside” feature.

Appendix F contains the “Classroom Test of Scientific Reasoning,” a 12-item test that, according to Lawson: (a) is a test of “ability to apply aspects of scientific and mathematical reasoning to analyze a situation, make a prediction, or solve a problem”; (b) total scores indicate the following levels of thinking: of 0-4 - empirical-inductive; 5-8 - transitional; 9-12 hypothetical-deductive. A 24-item multiple-choice version of this test is in the Appendix of Coletta & Phillips (2005).

Lawson, A.E. 2003. “Using the RTOP To Evaluate Reformed Science and Mathematics Instruction,” in *Improving Undergraduate Instruction in Science, Technology, Engineering, and Mathematics: Report of a Workshop* (NRC (2003b).

Libarkin, J. 2008. “Concept Inventories in Higher Education Science,” paper commissioned by the National Academies, online as a 336 kB pdf at <http://bit.ly/p4ejnL>. See also National Academies (2008).

“Table 1 gives a "Comprehensive list of published concept inventories in science.”

Table 2. Gives a "List of unpublished science concept inventories under development."

Lissitz, R.W. & K. Samuelsen. 2007. "A Suggested Change in Terminology and Emphasis Regarding Validity and Education," *Educational Researcher* **36**(8): 437-448; online to subscribers at <http://edr.sagepub.com/content/vol36/issue8/>, along with (a) comments on the article by Embretson, Gorin, Mislevy, Moss, and Sireci, (b) response to the comments by Lissitz & Samuelson.

ABSTRACT: "This article raises a number of questions about the current unified theory of test validity that has construct validity at its center. The authors suggest a different way of conceptualizing the problem of establishing validity by considering whether the focus of the investigation of a test is internal to the test itself or focuses on constructs and relationships that are external to the test. They also consider whether the perspective on the test examination is theoretical or practical. The resulting taxonomy, encompassing both investigative focus and perspective, serves to organize a reconceptualization of the field of validity studies. The authors argue that this approach, together with changes in the rest of the terminology regarding validity, leads to a more understandable and usable model."

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.

Maloney, D.P., T.L. O'Kuma, Curtis J. Hieggelke, and Alan Van Heuvelen. 2001. "Surveying students' conceptual knowledge of electricity and magnetism," Phys. Educ. Res., *Am. J. Phys. Suppl.* **69**, S12–S23; online to subscribers at <http://bit.ly/qI9Y8z>.

Mallinckrodt, J. 2001. "Re: Using the FCI," PhysLrnR post of 3 Oct 2001 19:08:11–0700; online on the PhysLrnR archives at <http://bit.ly/omSE9u>.

Machiavelli, N. 1513. *The Prince*. For A 2011 edition published by Simon & Brown see Amazon.com at <http://amzn.to/rp35T8>.

†Maki, P.L. 2004. *Assessing for Learning: Building a Sustainable Commitment Across the Institution*. Stylus, publisher's information at <http://bit.ly/lq9tQt>. Amazon.com information at <http://amzn.to/k63qTr>. Note the searchable "Look Inside" feature.

†Marusic, M. & J. Slisko. 2011. "Influence of Three Different Methods of Teaching Physics on the Gain in Students' Development of Reasoning," *International Journal of Science Education;* an abstract is online at <http://bit.ly/phQzSp>.

Mazur, E. 1997. *Peer instruction: a user's manual*. Prentice Hall; information online at <http://bit.ly/n73k9g>.

Mazur, E. 2009. “Confessions of a Converted Lecturer” talk at the University of Maryland on 11 November 2009. That talk is now on UTube at <http://bit.ly/dBYsXh>, and the abstract, slides, and references - sometimes obscured in the UTube talk - are at <http://bit.ly/9qzDIq> as a 4 MB pdf.
As of 10 September 19:33-0700 Eric's talk had been viewed 43,323 times. In contrast, serious articles in the education literature (or even articles such as this one) are often read only by the author and a few cloistered specialists, creating tsunamis in educational practice equivalent to those produced by a pebble dropped into the middle of the Pacific Ocean.

McDermott L.C. & E.F. Redish. 1999. “RL-PER1: Resource letter on physics education Research,” *Am. J. Phys.* **67**(9): 755-767 ; online at <http://bit.ly/baDTtQ> and <http://bit.ly/jWvlqS>.

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://bit.ly/evNlqb>. See also the Addendum “Normalized Learning Gain: A Key Measure Of Student Learning” at the same online location.

Menand, L. 2011. “Live And Learn: Why we have college,” *New Yorker*, 6 June; online at <http://nyr.kr/p5IHsp>.

Mervis, J. 2011. "A Better Way to Teach?" *ScienceNOW*. 12 May; online at <http://bit.ly/mOxumC>.

Mestre, J.P . 2005. “Facts and myths about pedagogies of engagement in science learning,” *Peer Review* **7**(2,) Winter Issue): 24-27; online at <http://bit.ly/pweg2K>.

Mestre, J.P. 2008. "Learning Goals in Undergraduate STEM Education and Evidence for Achieving Them," paper commissioned by the National Academies Board on Science Education <http://bit.ly/hDio0Z>; online as a 53 kB pdf at <http://bit.ly/fvrU7X>.

†Michael, J. & H, I. Modell. 2003. *Active Learning in Secondary and College Science Classrooms: A Working Model for Helping the Learner To Learn*. Routledge. Amazon.com information at <http://amzn.to/ok03wD>. See also Michael (2006).

Michael, J. 2006. “Where’s the evidence that active learning works?” *Advances in Physiology Education* **30**: 159-167, online at <http://bit.ly/9x4l7g>, a masterful review by a medical education researcher/developer. Michael wrote:
“One of the most striking findings [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, 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.”
See also Michael & Modell (2003).

Michell, J. 2005. “The logic of measurement: A realist overview,” *Measurement* **38**(4): 285-294; abstract online at <http://bit.ly/9RXSHD>.

Michell, J. 2008. "Is Psychometrics Pathological Science?" *Measurement: Interdisciplinary Research & Perspective* **6**(1 & 2), Issue 1 & 2: 7-24; abstract online at <http://bit.ly/cjZCgI>. See also Michell (2005).

Millar, R. & J. Osborne. 2009. "Research and Practice: A Complex Relationship? Chapter 3, pages 41-62, of Shelley et al. (2009). To see some of this go to Amazon’s “Look Inside” feature at <http://amzn.to/97OVJx>, and search for “Millar” (without the quotes).

Miller, K., N. Lasry, O. Reshef, J. Dowd, I. Araujo, and E. Mazur. 2010. “Losing it: The Influence of Losses on Individuals’ Normalized Gains,” PERC 2010, online as a 259 kB pdf at <http://bit.ly/nCULpL>.

Miller et al. (2010) reference “Reassessing Hake’s Gain” [Dellwo (2009)] as “available on request,” evidently unaware of the published “Course assessment using multi-stage pre/post testing and the components of normalized change” [Dellwo (2010)].

†Millis, B. ed. 2010. *Cooperative Learning in Higher Education: Across the Disciplines, Across the Academy (New Pedagogies and Practices for Teaching in Higher Education*), foreword by James Rhem. Stylus. Amazon.com information at <http://amzn.to/b1Ysa1>. Note the searchable “Look Inside” feature.

Mislevy, R. 2006: (a) “On approaches to assessing change,” and (b) “Clarification”; both online at <http://bit.ly/98WXLP>. In the latter Mislevy wrote:

“I should clarify the point I made in my piece on approaches for assessing change. Most notably, it is not about the substance of Prof. Hake’s work, which I agree is important - important in taking prior knowledge into account, important in the comparisons he makes, important in implications for instruction, and important in using *a* method (namely <g>) that mitigates certain problems with ceiling effects with test scores. My point is, perhaps, more in the nature of a ‘sociology of science’ comment concerning a methodological choice. The substantive aspect of the work would be equally valuable, and provide the same essential results, with an alternative (and in the relevant aspects analogous) method for dealing with those effects. . . . .[[I’d be interested in knowing how Mislevy thinks IRT could be used to compare the effectiveness of many courses with widely varying average pretest scores.]]. . . . The reason that psychometricians today would be prone to use an alternative such as IRT rather than <g> is that the alternative accomplishes the same purpose but is grounded in the methodological web of a generative program of research. Specifically, it is founded on probability-based inference, is invariant (with the appropriate qualifications) across testing situations, and is extensible to broader models for the nature of knowledge, learning, and change.”

See also Mislevy (2008).

Mislevy, R. 2008. “How Cognitive Science Challenges the Educational Measurement Tradition,” *Measurement* **6**: 124; online as a 250 kB pdf at <http://bit.ly/b6213i>. Therein Mislevy wrote:

“The procedures through which measurement specialists investigate validity, establish reliability, and ensure fairness are enmeshed with the language and worldview of trait and behavioral psychology. The resulting narrative space articulates poorly with an emerging integration of individual, situative, and social perspectives on cognition—a ‘sociocognitive’ perspective, in the terminology of Atkinson, et al. (2007). Cognitive science challenges the educational measurement tradition to bridge this widening chasm. Doing so entails a broader conceptions of the nature of proficiency and ways it is evidenced. . . . . An assessment argument cast in sociocognitive terms might use the same probability mode as a trait argument, but the situated meanings of the elements, the interpretation of observations, and nature of the target inferences would be different, with regard to (1) alternative interpretations of performance, (2) generalizations beyond the observational context, and (3) the role of information about the examinee/situation relationships. More explicit development of these ideas is beginning to appear in learning domains. *Redish (2003), for example, describes key ideas of a sociocognitive perspective to physics educators, and shows them how to leverage the ideas to improve teaching and assessment*.” [My *italics*.]

But Redish (2003) in his Fig. 5.2 shows an early and now superseded graph of <g> vs %pretest score similar to Fig. 1 of Hake (1998a), with no mention of Mislevy’s (2006) uninterest in <g>. See also “Is Psychometrics Pathological Science?” [Michell (2008), Hake (2010c)].

Morris, G.A., L. Branum-Martin, N. Harshman, S.D. Baker, E. Mazur, S. Dutta, T. Mzoughi, and V. McCauley. 2006. “Testing the test: Item response curves and test quality,” *Am. J. Phys.* **74**(5): 449-453; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v74/i5>.

National Academies. 2008. "Evidence on Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education: Workshop on Linking Evidence and Promising Practices in STEM Undergraduate Education" online at <http://bit.ly/fAhNpA>: Meeting 1 of 30 June, online at <http://bit.ly/ciNwjQ>; Meeting 2 of 13-14 October containing commissioned papers online at <http://bit.ly/ceg1Bx>. See also the commentary on these workshops by Labov et al. (2009), and the NRC (2011a) book *Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics Education: Summary of Two Workshops.*

NAE. 1995. National Academy of Engineering, *Forces Shaping the U.S. Academic Engineering Research Enterprise,* Committee on Academic Engineering Research; online at <http://bit.ly/rgijt5>.

NAE. 2004. National Academy of Engineering, *The Engineer of 2020: Visions of Engineering in the New Century*. The National Academies Press; online at <http://bit.ly/o0mQAu>.

NAE. 2005a. National Academy of Engineering “Educating the engineer of 2020: Adapting engineering education to the new century,” The National Academies Press; online at <http://bit.ly/oeO0ZI>. The description reads:

NAE. 2005b. National Academy of Engineering, *Engineering Research and America’s Future* Committee to Assess the Capacity of the U.S. Engineering Research Enterprise; online at <http://bit.ly/q1xD1q>.

NEERC. 2006. National Engineering Education Research Colloquies, "Special Report." The introduction is at *Journal of Engineering Education* **95**(4): 257–258, online as a 41 kB pdf at <http://bit.ly/n5RCgc>. The report itself is at *Journal of Engineering Education* **95**(4): 259–261; online as a 66 kB pdf at <http://bit.ly/n5JVh3>.

NSF. 1995. "Restructuring Engineering Education: A Focus on Change," online as a 57 kB pdf at <http://1.usa.gov/oV60Q4>.

†NSF. 2010. *Science and Engineering Indicators: 2010*; online at <http://www.nsf.gov/statistics/seind10/start.htm>.

NCSU. 2011. "Assessment Instrument Information Page," Physics Education R & D Group, North Carolina State University; online at <http://bit.ly/9gfUpY>.

NCEE. 2007. National Center On Education And The Economy, *Tough Choices for Tough Times*; the executive summary is online as a 1.9 MB pdf at <http://bit.ly/nUgKOt>.

†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://bit.ly/mY05RR>.

†Nelson, C. 2009. "Dysfunctional Illusions of Rigor: Lessons from the Scholarship of Teaching and Learning," Chapter 10 in Nilson & Miller (2009); also in the *Tomorrow's Professor* messages 1058 and 1059 of 12 and 16 November 2010, archived at <http://bit.ly/d09Y8r> - type the message numbers into the slot at the top of the page.

†Nilson, L.B. and J.E. Miller, eds. 2009. *To Improve the Academy: Resources for Faculty, Instructional, and Organizational Development, Volume 28*, Jossey-Bass. Barnes & Noble information at <http://bit.ly/9soRDk>, note the searchable "See inside" feature.

NRC. 1997. *Science Teaching Reconsidered: A Handbook*, National Research Council, Committee on Undergraduate Science Education, National Academy Press; online at <http://www.nap.edu/catalog/5287.html>.

NRC. 1999. *Transforming Undergraduate Education in Science, Mathematics, Engineering, and Technology,* National Research Council, Committee on Undergraduate Science Education, National Academy Press; online at <http://www.nap.edu/catalog/6453.html>.

NRC. 2003a. *Evaluating and Improving Undergraduate Teaching in Science and Technology, Engineering, and Mathematics*, ed. by M.A. Fox & N. Hackerman, National Research Council, Committee on Undergraduate Science Education, National Academy Press; online at <http://www.nap.edu/catalog/10024.html>.

NRC. 2003b. *Improving Undergraduate Instruction in Science, Technology, Engineering, andMathematics: Report of a Workshop*," McCray, R.A., R.L. DeHaan, J.A. Schuck, eds. Committee on Undergraduate STEM Instruction, National Academy Press; online at <http://www.nap.edu/catalog/10711.html>.

NRC. 2007a. Committee on Science, Engineering, and Public Policy, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Future*, National Academies Press; online at <http://bit.ly/a3g4P3>. The description reads:

"In a world where advanced knowledge is widespread and low-cost labor is readily available, U.S. advantages in the marketplace and in science and technology have begun to erode. A comprehensive and coordinated federal effort is urgently needed to bolster *U.S. competitiveness* . . . .[[my *italics]*]. . . . and pre-eminence in these areas. . . . .[[For an issue even more important than "U.S. competitiveness" see "The Threat to Life on Planet Earth Is a More Important Issue Than David Brooks' Skills Slowdown" (Hake, 2009b)]]. . . . . This congressionally requested report by a pre-eminent committee makes four recommendations along with 20 implementation actions that federal policy-makers should take to create high-quality jobs and focus new science and technology efforts on meeting the nation's needs, especially in the area of clean, affordable energy:

1) Increase America's talent pool by vastly improving K-12 mathematics and science education;
2) Sustain and strengthen the nation's commitment to long-term basic research;
3) Develop, recruit, and retain top students, scientists, and engineers from both the U.S. and abroad; and
4) Ensure that the United States is the premier place in the world for innovation.

Some actions will involve changing existing laws, while others will require financial support that would come from reallocating existing budgets or increasing them. *Rising Above the Gathering Storm* will be of great interest to federal and state government agencies, educators and schools, public decision makers, research sponsors, regulatory analysts, and scholars."

See also the update *Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5* [NRC (201o) and other similar dire warnings in *Is America Falling Off the Flat Earth?* NRC (2007b).

NRC. 2007b. *Is America Falling Off the Flat Earth?* online at <http://bit.ly/qwdyDu>. The description reads:

"The aviation and telecommunication revolutions have conspired to make distance increasingly irrelevant. An important consequence of this is that US citizens, accustomed to competing with their neighbors for jobs, now must compete with candidates from all around the world. These candidates are numerous, highly motivated, increasingly well educated, and willing to work for a fraction of the compensation traditionally expected by US workers. If the United States is to offset the latter disadvantage and provide its citizens with the opportunity for high-quality jobs, it will require the nation to excel at innovation -- that is, to be first to market new products and services based on new knowledge and the ability to apply that knowledge. *This capacity to discover, create and market* will continue to be heavily dependent on the nation's prowess in science and technology. Indicators of trends in these fields are, at best, highly disconcerting. While many factors warrant urgent attention, the two most critical are these: (1) America must repair its failing K-12 educational system, particularly in mathematics and science, in part by providing more teachers qualified to teach those subjects, and (2) the federal government must markedly increase its investment in basic research, that is, in the creation of new knowledge. Only by providing leading-edge human capital and knowledge capital can America continue to maintain a high standard of living--including providing national security--for its citizens."

NRC. 2010. *Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5*, online at <http://bit.ly/dtQhbS>. The description reads:

"In the face of so many daunting near-term challenges, U.S. government and industry are letting the crucial strategic issues of U.S. competitiveness slip below the surface. . . . . .[[For an issue even more important than "U.S. competitiveness" see "The Threat to Life on Planet Earth Is a More Important Issue Than David Brooks' Skills Slowdown" (Hake, 2009b)]]. . . .. Five years ago, the National Academies prepared *Rising Above the Gathering Storm*, a book that cautioned: "Without a renewed effort to bolster the foundations of our competitiveness, we can expect to lose our privileged position." Since that time we find ourselves in a country where much has changed--and a great deal has not changed.

So where *does* America stand relative to its position of five years ago when the *Gathering Storm* book was prepared? The unanimous view of the authors is that our nation's outlook has worsened. The present volume, *Rising Above the Gathering Storm, Revisited*, explores the tipping point America now faces. Addressing America's competitiveness challenge will require many years if not decades; however, the requisite federal funding of much of that effort is about to terminate.

Rising Above the Gathering Storm, Revisited provides a snapshot of the work of the government and the private sector in the past five years, analyzing how the original recommendations have or have not been acted upon, what consequences this may have on future competitiveness, and priorities going forward. In addition, readers will find a series of thought- and discussion-provoking factoids--many of them alarming--about the state of science and innovation in America.

Rising Above the Gathering Storm, Revisited is a wake-up call. To reverse the foreboding outlook will require a sustained commitment by both individual citizens and government officials--at all levels. This book, together with the original *Gathering Storm* volume, provides the roadmap to meet that goal. While this book is essential for policy makers, anyone concerned with the future of innovation, competitiveness, and the standard of living in the United States will find this book an ideal tool for engaging their government representatives, peers, and community about this momentous issue."

NRC. 2011a. *Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics Education: Summary of Two Workshops*, online at <http://bit.ly/nCMLk7>. Natalie Nielsen, Rapporteur. Planning Committee on Evidence on Selected Innovations in Undergraduate STEM Education: Susan Singer (Chair), Melvin George, Kenneth Heller, David Mogk, & William B. Wood, Board on Science Education, Division of Behavioral and Social Sciences and Education. The National Academies Press; online at <http://bit.ly/nCMLk7>. The description reads:

> "Numerous teaching, learning, assessment, and institutional innovations in undergraduate science, technology, engineering, and mathematics (STEM) education have emerged in the past decade. Because virtually all of these innovations have been developed independently of one another, their goals and purposes vary widely. Some focus on making science accessible and meaningful to the vast majority of students who will not pursue STEM majors or careers; others aim to increase the diversity of students who enroll and succeed in STEM courses and programs; still other efforts focus on reforming the overall curriculum in specific disciplines. In addition to this variation in focus, these innovations have been implemented at scales that range from individual classrooms to entire departments or institutions.
>
> By 2008, partly because of this wide variability, it was apparent that little was known about the feasibility of replicating individual innovations or about their potential for broader impact beyond the specific contexts in which they were created. The research base on innovations in undergraduate STEM education was expanding rapidly, but the process of synthesizing that knowledge base had not yet begun. If future investments were to be informed by the past, then the field clearly needed a retrospective look at the ways in which earlier innovations had influenced undergraduate STEM education.
>
> To address this need, the National Research Council (NRC) convened two public workshops to examine the impact and effectiveness of selected STEM undergraduate education innovations. This volume summarizes the workshops, which addressed such topics as the link between learning goals and evidence; promising practices at the individual faculty and institutional levels; classroom-based promising practices; and professional development for graduate students, new faculty, and veteran faculty. The workshops concluded with a broader examination of the barriers and opportunities associated with systemic change."

NRC. 2011b. "*Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics*, online at <http://bit.ly/opEXhn>, "Committee on Highly Successful Science Programs for K-12 Science Education" of the Division of Behavioral and Social Sciences and Education. The description reads:

> "Science, mathematics, engineering, and technology (STEM) are fundamental aspects of everyone's lives as citizens, consumers, parents, and workers. Providing all students with access to high-quality education in STEM is important to their futures and that of the U.S. What can schools do to meet this goal for their students?
>
> *Successful K-12 STEM Education* tackles this question, focusing on the science and mathematics parts of STEM and on criteria for identifying effective STEM schools and practices. *Successful K-12 STEM Education* gives an overview of the landscape of K-12 STEM education by considering different school models, highlighting research on effective STEM education practices, and identifying some conditions that promote and limit school- and student-level success in STEM. It can serve as a guide for those involved in K-12 education at all levels: policy makers; decision makers at the school and district levels; local, state, and federal government agencies; curriculum developers; educators; and parent and education advocacy groups."

See also NRC (2011c).

NRC. 2011c. "A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas," Committee on Conceptual Framework for the New K-12 Science Education Standards; National Research Council, Helen Quinn, chair; online at <http://bit.ly/o1STSC>. The description reads:

> "Science, engineering, and technology permeate nearly every facet of modern life and hold the key to meeting many of humanity's most pressing challenges, both present and future. To address the critical issues of *U.S. competitiveness* and to better prepare the workforce, *Framework for K-12 Science Education* proposes a new approach to K-12 science education that will capture students' interest and provide them with the necessary foundational knowledge in the field."
>
> *Framework for K-12 Science Education* outlines a broad set of expectations for students in science and engineering in grades K-12. These expectations will inform the development of new standards for K-12 science education and, subsequently, revisions to curriculum, instruction, assessment, and professional development for educators. This book identifies three dimensions that convey the disciplinary core ideas and practices around which science and engineering education in these grades should be built. These three dimensions are: cross-cutting concepts that unify the study of science and engineering through their common application across these fields; scientific and engineering practices; and core ideas in four disciplinary areas: physical sciences, life sciences, earth and space sciences, and engineering, technology, and the applications of science. The overarching goal is for all high school graduates to have sufficient knowledge of science and engineering to engage in public discussions on science-related issues; be careful consumers of scientific and technological information; and have the skills to enter the careers of their choice.
>
> *Framework for K-12 Science Education* is the first step in a process that will inform state-level decisions and provide a research-grounded basis for improving science teaching and learning across the country. The book will guide standards developers, curriculum designers, assessment developers, teacher educators, state and district science administrators, teachers, and educators who work in informal science environments."

See also NRC (2011b).

NRC. 2011d. Board on Science Education (BOSE), Project: "Status, Contributions, and Future Directions of Discipline Based Education Research," information online at <http://bit.ly/qfj0AU>: "The project is sponsored by the National Science Foundation. The approximate start date for the project is September 15, 2009. A report will be issued at the end of the project in approximately 30 months" (about March 2012). Committee Membership: <http://bit.ly/ogzQ5T>: biologist Susan R. Singer - (Chair), physicist Robert Beichner, chemist Stacey Lowery Bretz, chemist Melanie Cooper, chemist Sean Decatur, education specialist James Fairweather, physicist Kenneth Heller, geoscientist Kim A. Kastens, education specialist Michael E. Martinez, geologist David Mogk, chemist Marcy Osgood, astronomer Timothy Slater, engineer Karl A. Smith, biologist William B. Wood, & psychologist Laura R. Novick. The description reads:

> "Numerous teaching, learning, assessment, and institutional innovations in undergraduate science, technology, engineering, and mathematics (STEM) education have emerged in the past decade. Because virtually all of these innovations have been developed independently of one another, their goals and purposes vary widely. Some focus on making science accessible and meaningful to the vast majority of students who will not pursue STEM majors or careers; others aim to increase the diversity of students who enroll and succeed in STEM courses and programs; still other efforts focus on reforming the overall curriculum in specific disciplines. In addition to this variation in focus, these innovations have been implemented at scales that range from individual classrooms to entire departments or institutions.
>
> By 2008, partly because of this wide variability, it was apparent that little was known about the feasibility of replicating individual innovations or about their potential for broader impact beyond the specific contexts in which they were created. The research base on innovations in undergraduate STEM education was expanding rapidly, but the process of synthesizing that knowledge base had not yet begun. If future investments were to be informed by the past, then the field clearly needed a retrospective look at the ways in which earlier innovations had influenced undergraduate STEM education.
>
> To address this need, the National Research Council (NRC) convened two public workshops to examine the impact and effectiveness of selected STEM undergraduate education innovations. This volume summarizes the workshops, which addressed such topics as the link between learning goals and evidence; promising practices at the individual faculty and institutional levels; classroom-based promising practices; and professional development for graduate students, new faculty, and veteran faculty. The workshops concluded with a broader examination of the barriers and opportunities associated with systemic change."

NSF. 1995. *Restructuring Engineering Education: A Focus on Change*, online at <http://1.usa.gov/oAae9Q>.

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 as a 1.8 MB pdf at <http://1.usa.gov/9X4Nh9>. This report is one of the few that emphasizes the crucial role of higher education in determining the quality of K-12 education.

NSF. 2005. *Making the Case for Engineering: Study and Recommendations*, online at <http://bit.ly/p6NQHo>. The first paragraph of the Executive Summary is:

> "The reasons for making the case for engineering are clear and compelling. First, engineering drives our economy; it builds the foundation for a prosperous and secure future. Second, the engine of engineering innovation does not run solely on the value of its contributions to society; it requires a committed and active public and private investment. It also requires a strong educational foundation and a world-class workforce. Finally, the public awareness and understanding of engineering – its process and its impact – are not well understood by the public. This final point hinders our nation's ability to adequately attract a diverse, world-class engineering workforce, and to enact the bold programs necessary to ensure our continued leadership in engineering innovation."

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://bit.ly/bmhkvo>. If your institution doesn't have a subscription, then 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://bit.ly/pj10Wr>.

Nuhfer, E. 2010. "Re: Control Group Studies," online on the OPEN! POD archives at <http://bit.ly/pv8JE9>.

Nuhfer, E. 2011. "First Results From The Science Literacy Concept Inventory: The Reasoning We Don't Produce Through Gen-Ed," paper delivered at the Geological Society of America (GSA) meeting in Minneapolis, 9–12 October; online at <http://bit.ly/nsu1MJ>.

Olsen, D.R. 2004. "The Triumph of Hope Over Experience in the Search for 'What Works': A Response to Slavin," *Educational Researcher* **33**(1): 27-38 (2004); online to subscribers at <http://bit.ly/ksTkNA>.

Osborne, J. 2007. "In praise of armchair science education," contained within E-NARST News 50(2) at <http://www.narst.org/news/e-narstnews_July2007.pdf> (3.2 MB). The talk itself is online at <http://www.kcl.ac.uk/content/1/c6/01/29/36/joconference.pdf> (112 kB).

†Pasachoff, J.M., R.M. Ros, & N. Pasachoff, eds. 2008. *Innovation in Astronomy Education*. Cambridge University Press. publisher's information at <http://bit.ly/nRlbdv>. Amazon.com information at <http://amzn.to/pNwYug>, note the searchable "Look Inside" feature. See especially Chapter 7 "Clickers: a new teaching tool of exceptional promise" by Douglas Duncan, and Chapter 28 "Revitalizing astronomy teaching through research on student understanding" by Timothy Slater.

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>. The description reads:

"Education is a hot topic. From the stage of presidential debates to tonight's dinner table, it is an issue that most Americans are deeply concerned about. While there are many strategies for improving the educational process, we need a way to find out what works and what doesn't work as well. Educational assessment seeks to determine just how well students are learning and is an integral part of our quest for improved education.

The nation is pinning greater expectations on educational assessment than ever before. We look to these assessment tools when documenting whether students and institutions are truly meeting education goals. But we must stop and ask a crucial question: What kind of assessment is most effective?

At a time when traditional testing is subject to increasing criticism, research suggests that new, exciting approaches to assessment may be on the horizon. Advances in the sciences of how people learn and how to measure such learning offer the hope of developing new kinds of assessments-assessments that help students succeed in school by making as clear as possible the nature of their accomplishments and the progress of their learning.

Knowing What Students Know essentially explains how expanding knowledge in the scientific fields of human learning and educational measurement can form the foundations of an improved approach to assessment. These advances suggest ways that the targets of assessment - what students know and how well they know it - as well as the methods used to make inferences about student learning can be made more valid and instructionally useful. Principles for designing and using these new kinds of assessments are presented, and examples are used to illustrate the principles. Implications for policy, practice, and research are also explored.

With the promise of a productive research-based approach to assessment of student learning, *Knowing What Students Know* will be important to education administrators, assessment designers, teachers and teacher educators, and education advocates."

Pellegrino, J. W. 2006. "Rethinking and redesigning curriculum, instruction and assessment: What contemporary research and theory suggests," a paper commissioned by the National Center on Education and the Economy for the *New* Commission on the Skills of the American Workforce [NCEE (2007)]; online as a 377 kB pdf at <http://bit.ly/n6NdfT>.

†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://bit.ly/fj2P1E>. Amazon.com information at <http://amzn.to/qPGjZc>. The late Paul Meehl <http://en.wikipedia.org/wiki/Paul_Meehl> wrote:

"Should be required reading for all Ph.D. candidates in social science. It is a mind clearing analysis of the highest order, prophylactic and curative of the numerous methodological and substantive ills that afflict us. It is especially needed today when the 'positivist-bashers' are using the Vienna Circle's mistakes and Kuhn's exaggerations for obscurantist purposes."

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

†Planinic, M., L. Ivanjek, & A. Susac. 2010. "Rasch model based analysis of the Force Concept Inventory," *Physical Review Special Topics - Physics Education Research* **6**, 010103; online as a 504 kB pdf at <http://bit.ly/pHw9SP>.

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://bit.ly/eALEaO>.

Prince, M. 2004. "Does Active Learning Work? A Review of the Research," *Journal of Engineering Education* **93**(3): 223-231; online as a 752 kB pdf at <http://bit.ly/rkiBjq>. See also Prince & Felder (2006, 2007)), and Prince et al. (2007).

Prince, M., & R. Felder. 2006. "Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases," *Journal of Engineering Education* **95**(2):123-138; online as a 148 kB pdf at <http://bit.ly/qheyVH>.

Prince, M., & R. Felder. 2007. "The Many Faces of Inductive Teaching and Learning," *Journal of College Science Teaching* **36**(5): 14-20, March/April. online as a 324 kB pdf at <http://bit.ly/om3rcg> (324 kB).

Prince, M., R. Felder, & R. Brent. 2007. "Does Faculty Research Improve Undergraduate Teaching? An Analysis of Existing and Potential Synergies." *Journal of Engineering Education* **96**(4): 283-294; online as a 116 kB pdf at <http://bit.ly/mVl5mf>.

RAE. 2006. Royal Academy of Engineering, United Kingdom, "Educating Engineers for the 21st Century: The Industry View, " online as a 2.2 MB pdf at <http://bit.ly/qxUJFo>.

Ramlo, S. 2008a. "Re: measure of gain, " online on the PhysLrnR archives at <http://bit.ly/oS8joC>. Post of 11 Jun 2008 21:30:33-0400. See also Ramlo (2008b).

Ramlo, S. 2008b. "Re: measure of gain," online on the PhysLrnR archives at <http://bit.ly/nzTjTQ>. Post of 16 Jun 2008 23:26:54-0400.

Ramlo S. 2008c. “Validity and reliability of the force and motion conceptual evaluation,” *Am. J. Phys.* **76**(9): 882-886; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias>.
ABSTRACT: “The assessment of learning has become a key component in program evaluation, grant proposals, and education research. Assessment requires valid and reliable instruments. The Force and Motion Conceptual Evaluation FMCE is one of several multiple-choice tests used to evaluate the learning of force and motion concepts. Although many physics education researchers accept its validity and reliability, validity and reliability estimates based on typical statistical analyses of data have not been established. This study used FMCE post-test results for approximately 150 students in a first-semester college physics course to estimate reliability and content validity. The results indicate that the FMCE is a valuable instrument for measuring student learning.”

Redish, E.F., J.M. Saul, & R.N. Steinberg. 1998. “Student expectations in introductory physics,” A*m. J. Phys*. **66**(3): 212–24; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v66/i3>.

Redish, E.F. 1999. “Millikan lecture 1998: Building a science of teaching physics,” *Am. J. Phys*. **67**(7): 562-573; online at <http://bit.ly/pePmuK>.

Redish, E.F. 2003. *Teaching Physics With the Physics Suite*, John Wiley. This book is online at <http://bit.ly/gdE3Tu>.
Note the crucial correction of Fig. 5.2 and its caption on page 100 of the online version.

Redish, E.F . 2004. “A Theoretical Framework for Physics Education Research: Modeling student thinking,” in E.F. Redish and M. Vicentini (Eds.) *Proceedings of the International School of Physics, "Enrico Fermi" Course CLVI*, Amsterdam: IOS Press; online as a 741 kB pdf at <http://bit.ly/mPNeGE>.

Redish, E.F. and K.A. Smith. 2008. “Looking beyond content: Skill development for engineers,” *Journal of Engineering Education* **97**(3): 295–307; online as a 2.4 MB pdf at <http://bit.ly/od8LEb>.

†Redish, E.F. & D. Hammer. 2009. "Reinventing college physics for biologists: Explicating an epistemological curriculum," *Am. J. Phys.* **77**(7): 629-642; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v77/i7>.

Reed-Rhoads, T. & P.K. Imbrie. 2008. “Concept Inventories in Engineering Education,” paper commissioned by the National Academies, online as a 123 kB pdf at <http://bit.ly/dW0Fpa>. See also National Academies (2008).

Reed-Rhoads, T., P. K. Imbrie, K. Allen, J. Froyd, J. Martin, R. Miller, P. Steif, A. Stone, and R. Terry. 2007. “Tools to facilitate better teaching and learning: Concept inventories,” Panel at ASEE/IEEE Frontiers in Education Conference. Milwaukee, WI; online as a 172 kB pdf at <http://bit.ly/oKGG2y>.

Richardson, J. 2004. “Concept Inventories: Tools For Uncovering STEM Students’ Misconceptions,” in the section “Assessment and Education Research” of AAAS (2004); online as a 82 kB pdf at <http://bit.ly/obr6n6>.

Rigden, J. 1984. “The Introductory Course,” editorial, *Am. J. Phys*. **52**(4): 304; unaccountably missing from <http://ajp.aapt.org/resource/1/ajpias/v52/i4>.

Rimer, S. 2009. “At M.I.T., Large Lectures Are Going the Way of the Blackboard,” *New York Times*, 12 January; online at <http://nyti.ms/awEWKY>.

Rogers, E.M. 2003. *Diffusion of Innovations*, 5th edition. Free Press. Amazon.com information at <http://amzn.to/qk0WZP>, note the "Look Inside" feature.

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. 1998, “Psychometric Models of Student Conceptions in Science: Reconciling Qualitative Studies and Distractor-Driven Assessment Instruments,” J. *Res. Sci. Teach.* **35**(3): 265–96; an abstract is online at <http://bit.ly/orO1f7>.

†Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., and Gould, R. R. 2010. “The Astronomy and Space Science Concept Inventory: Development and Validation of Assessment Instruments Aligned with the K-12 National Science Standards,” *Astron. Educ. Rev*. **8**, 010111; an abstract is online at <http://bit.ly/oyMeH1>

Sadler, P. 2008a. “Re: measure of gain,” PhysLnrR post of 15 Jun 200820:49:11-0400; online at <http://bit.ly/pPLSwh>.

Sadler, P. 2008b. “Re: measure of gain,” PhysLnrR post of 17 Jun 2008 10:12:11-0400 online at <http://bit.ly/ppeMHo>.

Scriven, M. 2004. “Re: pre-post testing in assessment,” AERA-D post of 15 Sept 2004 19:27:14-0400; online at <http://bit.ly/n9PPaH>.

Scriven, M. 2008. “A Summative Evaluation of RCT Methodology: & An Alternative Approach to Causal Research,” *Journal of MultiDisciplinary Evaluation* **5**(9): 11-24; online at <http://bit.ly/93VcWD>. Commenting on Scriven’s article, Ed Nuhfer (2010) in a POD post “Re: Control Group Studies” wrote (paraphrasing) : “. . . . it appears to me that Scriven has deduced on his own “The Method of Multiple Working Hypotheses” articulated for natural sciences by T.C. Chamberlin (1890) 120 years earlier, but Scriven has simply articulated the same for social science.” And Nuhfer (2006a,b) in parts 1 and 2 of “A Fractal Thinker Looks at Measuring Change” bases his defense of the conclusions of Hake (1998a,b) on Chamberlin’s Method of Multiple Working Hypotheses.

†Shadish, W.R., T.D. Cook, & D.T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Pages 1-32 and 456 -504 are online at <http://bit.ly/n22TvF>. Houghton Mifflin. Amazon.com information at <http://amzn.to/pZPDkK>. 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. 2008. "Formative Assessment," Guest Editor's Introduction, special issue of *Applied Measurement in Education*; online at <http://bit.ly/nn2Rcz>. Also free online at the same URL are five articles on formative assessment that appear in *Applied Measurement in Education* **21**(4) of 4 October; online to subscribers at <http://bit.ly/phtxnG>.

Shelley, M.C., L.D. Yore, & B. Hand, eds. 2009. *Quality Research in Literacy and Science Education: International Perspectives and Gold Standards*. Springer, publisher's information at <http://bit.ly/b58vbP>. Amazon.com information at <http://amzn.to/97OVJx>, note the searchable "Look Inside" feature. Barnes & Noble information at <http://bit.ly/p40bKu>. An expurgated (teaser) version is online as a Google "book preview" at <http://bit.ly/qK8T9P>.

Singer, S.R. 2008. "Linking Evidence and Learning Goals" paper commissioned by the National Academies Board on Science Education <http://bit.ly/hDio0Z>; online as a 37 kB pdf at <http://bit.ly/pgANnw>. Singer wrote:

> "Physics has a theoretical framework for physics education research that offers approaches to closing gaps in evidence (Redish, 2004).. . . . Given epistemological differences among STEM disciplines, establishing similar agendas in the other disciplines could prove helpful in more broadly addressing gaps in evidence. . . . . *Classroom activities that actively engage students*: Often referred to as pedagogies of engagement, active learning includes a range of activities where faulty members replace at least a portion of lecture with activities that invite student participation. Enhanced student learning has been shown with assessments including pre- and post-tests and homework (Knight & Wood 2005). Overall, evidence supporting active learning is strong."

†Slater, T. & J. Adams. 2002. "Learner-Centered Astronomy Teaching, Strategies for ASTRO 101." Benjamin Cummings. Amazon.com information at <http://amzn.to/p8KpYI>.

Slavin, R.E. 1992. *Research Methods in Education.* Allyn and Bacon, 2nd ed. Amazon.com information at <http://amzn.to/ql7doK>.

Smith, K.A. 2007. "Preparing Students for an Interdependent World: Role of Cooperation and Social Interdependence Theory," Regional Conference on Engineering Education – Johor 2007, online as a 528 kB pdf at <http://bit.ly/pwKcSx>.

†Smith, K.A. 2011. "Resources: Books, Articles, Presentations and Workshops>,"online at <http://bit.ly/pjEDyK>.

†Springer, L., Stanne, M. E., & Donovan, S. S. 1999. "Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis," Review of Educational Research, 69(1), 21- 51; the first page is online at <http://bit.ly/n077Nd>.

†Stewart, J. & G. Stewart. 2010. "Correcting the Normalized Gain for Guessing," Phys.Teach. 48(3):194-196; online to subscribers at <http://tpt.aapt.org/resource/1/phteah/v48/i3>.

Steif, P. S. & J. A. Dantzler. 2005. "A statics concept inventory: Development and psychometric analysis," *Journal of Engineering Education* **94**(4): 363–71; online as a 172 kB pdf at <http://bit.ly/noaDXT>.

Steif, P.S. & M.A. Hansen. 2007. "New practices for administering and analyzing the results of concept inventories," *Journal of Engineering Education* **96**(3): 205–12; online as a 139 kB pdf at <http://bit.ly/pHy8Bc>.

Stokstad, E. 2001. "Reintroducing the Intro Course," *Science* 293: 1608-1610, 31 August; online at <http://bit.ly/a018BM>. 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."

Stone, A. K. Allen, T. Reed-Rhoads, T.J. Murphy, R.L. Shehab, and C. Saha. 2003. "The Statistics Concept Inventory: A Pilot Study," 33rd ASEE/IEEE Frontiers in Education Conference; online as a 53 kB pdf at <http://bit.ly/nm7zxj>. See also Stone (2006).

Stone, A. 2006. "A Psychometric Analysis of the Statistics Concept Inventory," Doctoral Dissertation, Mathematics, University of Oklahoma, Norman, OK; online as a 2.6 MB pdf at <http://bit.ly/phIcjZ>.

Streveler, R.A., T.A. Litzinger, R.L. Miller, and P.S. Steif. 2008. "Learning Conceptual Knowledge in the Engineering Sciences: Overview and Future Research Directions," *Journal of Engineering Education* **97**(3): 279-294; online as a 5.8 MB pdf at <http://bit.ly/nX0ypy>.

Suskie, L. 2004a. *Assessing Student Learning: A Common Sense Guide*. Anker Publishing. Anker information is at <http://bit.ly/mA5nVq>. Contrast this book by the virulently anti-pre/post Suskie, with one by the pro-pre/post Peggy Maki (2011). See also Suskie (2004b).

Suskie, L. 2004b. "Re: pre-post testing in assessment,"ASSESS post of 19 Aug 2004 08:19:53- 0400; online at <http://bit.ly/9yIOyf>. For counters see "Re: pre-post testing in assessment," Hake (2004) and Scriven (2004). See also Suskie (2004a).

†Swartz, C. E. 1999. "Editorial: Demise of a Shibboleth," *The Physics Teacher* **37**: 330, online to subscribers at <http://bit.ly/rfuuCy>. Swartz – normally an acerbic critic of physics education research - 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*. . . . .[[my i*talics*]]. . . . 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*."

Tagg, J. 2003. *The Learning Paradigm College*, with a forward by Peter Ewall. Jossey-Bass, publisher's information at <http://bit.ly/lkABm5>. Amazon.com information at <http://amzn.to/jKGcXO>. Note the searchable "Look Inside" feature. See also Tagg's website "Reflections on Learning" at <http://daphne.palomar.edu/jtagg/>.

Tarde, G. 1903. *The Laws of Immitation*, trans. by Else Clews Parsons, Univ. of Chicago Press; online at <http://bit.ly/nTCaxL>, thanks to the Univ. of Illinois.

TFFAI. 2005. Task Force On The Future Of American Innovation, "The Knowledge Economy: Is the United States Losing Its Competitive Edge? Benchmarks Of Our Innovation Future," online as a 1.6 MB pdf at <http://bit.ly/nxOCqW>. The Introduction reads, in part:

> "Research, education, the technical workforce, scientific discovery, innovation and economic growth are intertwined. To remain *competitive on the global stage*, ...[[my *italics*]]. . . .we must ensure that each remains vigorous and healthy. That requires sustained investments and informed policies.
>
> Federal support of science and engineering research in universities and national laboratories has been key to America's prosperity for more than half a century. A robust educational system to support and train the best U.S. scientists and engineers and to attract outstanding students from other nations is essential for producing a world-class workforce and enabling the R&D enterprise it underpins. But in recent years federal investments in the physical sciences, math and engineering have not kept pace with the demands of a knowledge economy, declining sharply as a percentage of the gross domestic product. This has placed future innovation and our *economic competitiveness* at risk. . . . .[[my *italics*]]. . . .
>
> To help policymakers and others assess U.S. *high-tech competitiveness* . . . . .[[my *italics*]]. . . .and the health of the American science and engineering enterprise, we have identified key benchmarks in six essential areas—education, the workforce, knowledge creation and new ideas, R&D investments, the high-tech economy, and specific high-tech sectors. We conclude that although the United States still leads the world in research and discovery, our advantage is eroding rapidly as other countries commit significant resources to enhance their own innovative capabilities."

Tobias, S. 1994. *They're Not Dumb, They're Different: Stalking the Second Tier*. Research Corporation. Amazon.com information at <http://amzn.to/oT47xx>.

†Tobias S. & R.R. Hake. 1988. "Professors as physics students: What can they teach us?" *Am. J. Phys.* **56**(9): 786-794, online as a 1.1 MB pdf at <http://bit.ly/eaqFzU>.

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://ajp.aapt.org/resource/1/ajpias/v68/i2>.

Tribus, M. 2001. "Quality in Education According to the Teachings of Deming and Feuerstein," online as a 78 kB pdf at <http://bit.ly/hwcbjn>. A Wikipedia entry on Myron Tribus is at <http://bit.ly/g5uGEk>.

Uretsky, J. 2009. “Re: Student engagement,” Phys-L post of 5 Dec 2009 22:33:27-0600, online on the OPEN! Phys-L archives at <http://bit.ly/qcyu2V>.

Wang, J. & L. Bao. 2009. “Analyzing force concept inventory with item response theory,” *Am. J. Phys.* **78** (10): 1064-1070; online to subscribers at <http://ajp.aapt.org/resource/1/ajpias/v78/i10>

Wallace, C.S. & J. M. Bailey. 2010. “Do Concept Inventories Actually Measure Anything?” *Astronomy Education Review* **9**(1); online as a 1.2 MB pdf at <http://bit.ly/oybsAJ>.

†Wehlburg., C.M., ed. 2010. *New Directions in Teaching and Learning, Vol. 2010, No. 123*; Special Issue: “Landmark Issues in Teaching and Learning: A Look Back at New Directions for Teaching and Learning.” Jossey-Bass, publisher's information at <http://bit.ly/gdecLz> - includes authors and titles of articles. Amazon.com information at <http://amzn.to/eclUr4>.

A review by James Rhem in *National Teaching and Learning Forum* **19**(6) is online to subscribers at <http://bit.ly/hVweNX> (If your institution doesn't subscribe to NTLF, it *should*).

†Wieman, C. & K. Perkins. 2005. “Transforming Physics Education,” *Phys. Today* **58**(11): 36-41; online as a 292 kB pdf at <http://bit.ly/9DRJ6l >. Carl Wieman was awarded the 2001 Nobel prize in physics.

Wieman, C. 2007. “Why Not Try a Scientific Approach to Science Education?” *Change Magazine*, September/October; online as a 804 kB pdf at <http://bit.ly/anTMfF>.

Wieman, C. , K. Perkins, & S. Gilbert. 2010. “Transforming Science Education at Large Research Universities: A Case Study in Progress,” *Change*, March-April; online at <http://bit.ly/oy2Vgs>. They wrote:

“There are countless reports stressing the economic and societal benefits to be gained from improved science, technology, engineering, and math (STEM) education for all students. But although there is extensive research on alternative teaching methods that increase student learning and are practical to implement [Wieman (2007), Redish (2003)], *the combined efforts of federal agencies, private foundations, and many internal institutional programs have achieved little overall change in STEM teaching at the large research universities.*. . . . [[my *italics*]]. . . . This remains a major problem for improving science education at any level, since these universities largely set the norms for how to teach science and what it means to learn science. We are currently testing a way to change the departmental culture for undergraduate science and math education at research universities in nine departments at the University of Colorado (CU) and the University of British Columbia (UBC), funded through the Science Education Initiatives . . . . . While it is still relatively early in this change process, there are significant indications of progress.”

Wilson, M.R. & Bertenthal, M.W., eds. 2005. *Systems for State Science Assessment*. National Academy Press; online at <http://bit.ly/f6WFeg>.

Wood, W.B., & J.M. Gentile. 2003. “Teaching in a research context,” *Science* **302**: 1510; 28 November; online as a 209 kB pdf at <http://bit.ly/oK46p7>. Wood and Gentile wrote:

> “Physics educators have led the way in developing and using objective tests to compare student learning gains in different types of courses, and chemists, biologists, and others are now developing similar instruments. 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.”

Woods, D.R. 2011b. *Motivating and Rewarding University Teachers to Improve Student Learning: A Guide for Faculty and Administrators*, ordering information at eurospanbookstore.com <http://bit.ly/geXRc5>, wherein the publication date is given as 15 August 2011.

Yarbrough, D.B., L.M. Shulha, R.K. Hopson, & F.A. Caruthers. 2011. *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users*, Third Edition, publisher’s information at <http://bit.ly/l98gg2>. Amazon.com information at <http://amzn.to/pt6rtR>.

Zar, J. 1994. “Candidate for a Pullet Surprise,” *Journal of Irreproducible Results*, January/February, p. 13. Reprinted “by popular demand” in the *Journal of Irreproducible Results* 45(5/6): 20, 2000; online at <http://www.jir.com/favorites.html>.

Zaritsky, R., A.E. Kelly, W. Flowers, E. Rogers, & P. Patrick. 2003. “Clinical Design Sciences: A View From Sister Design Efforts, ” Educational Researcher 32(1), 33-34; online at <http://www.aera.net/publications/?id=393>. See also “Why change, Been doin’ it this way for 4000 years!” Flowers (2000)].

Ziman, J. 1969. “Information, Communication, Knowledge,” *Nature* **224**: 318-324; abstract online at <http://bit.ly/cNPB1d>.

Ziman, J. 2000. *Real science: What it is, and what it means*. Cambridge, UK: Cambridge University Press, publisher's information at<http://bit.ly/c8N2U1>. Amazon.com information at <http://amzn.to/nsl2q2>, note the searchable “Look Inside” feature. See especially Sec. 9.3 “Codified knowledge,” pages 258-266.