Will the No Child Left Behind Act Promote Direct Instruction of Science ? [§] [*]

Richard R. Hake, Indiana University, Emeritus Professor of Physics
Woodland Hills, CA 91367, <rrhake@earthlink.net>

What we assess is what we value. We get what we assess, and if we don't assess it, we won't get it.
Lauren Resnick [quoted by Grant Wiggins (1990)]

The No Child Left Behind (NCLB) Act requires testing in science achievement starting in 2007. Will such testing tend to propagate California's Direct Science Instruction (DSI) [Hake (2004a)] throughout the entire nation? After discussing the evidence for the superiority of "interactive engagement" or "guided inquiry" methods over DSI in conceptually difficult areas of science, I indicate seven reasons why NCLB might promote DSI, and one reason – possible *effective* intervention by the National Research Council – why it might not.

I. Introduction

In a recent *Education Week* article "As Test Date Looms, Educators Renewing Emphasis on Science," Sean Cavenaugh (2005) cogently discusses the 2007 onset of science achievement testing under the No Child Left Behind (NCLB) Act of 2001 [See at < http://www.ed.gov/policy/elsec/leg/esea02/index.html > and < http://www.ed.gov/nclb/landing.jhtml?src=pb >.] Cavenaugh writes [my *italics*]:

> "Over the past three years, much of that science subject matter has been pushed aside, many state and local officials acknowledge. States and schools during that time have been consumed with the federal law's demand that they improve annual test scores in reading and mathematics in grades 3-8—or face such penalties as offering students a choice of schools to attend, or restructuring their own schools. But beginning in the

[§] The reference is: Hake, R.R. 2005. "Will the No Child Left Behind Act Promote Direct Instruction of Science?" *Bull. Am. Phys. Soc.* **50** (1): 851; APS March Meeting, Los Angles, CA. 21-25 March; online as reference 36 at < http://www.physics.indiana.edu/~hake >. I welcome comments and suggestions directed to <rrhake@earthlink.net>.

[*] Partially supported by NSF Grant DUE/MDR-9253965.

> 2007-08 school year, states will also have to test students in science in each of three grade spans. That deadline, some state and local officials say, is already giving science lessons more prominence in elementary and middle grades, after several years in which teachers were cutting back in that discipline because of more pressing demands. Under the federal law, states are required to have academic-content standards in place for science by the 2005-06 school year, and they must test students in science at least once in each of the 3-5, 6-9, and 10-12 grade spans in 2007-08. . . . *Although the upcoming wave of science testing does not carry the threat of federal penalties, others suggest that schools and districts would be motivated to perform well regardless because their scores will be made public.* "

Is there a danger that the NCLB's upcoming testing for science achievement will propagate California's Direct Science Instruction (DSI) [Hake (2004a)] throughout the entire nation [Hake (2004b, 2005a)]? After discussing the evidence for the superiority of "interactive engagement" or "guided inquiry" methods over DSI in conceptually difficult areas of science, I indicate seven reasons why NCLB might promote DSI, and one reason – possible *effective* intervention by the National Research Council – why it might not.

II. Evidence for the Superiority of Interactive Engagement (IE) over Traditional (T) Direct Instruction in High School and Undergraduate Mechanics Courses

A. Operational Definitions* [Hake (1998a)]:

IE courses: promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors;

T courses: employ *direct instruction* making little or no use of IE methods and relying primarily on passive-student lectures, recipe labs, and algorithmic problem exams;

Average Nomalized Gain <g>:

<g> = (average actual gain) / (maximum possible average gain) . . . . (1a)

<g> = (<%post> - <%pre>) / (100% - <%pre>) , ... (1b)

where the angle brackets indicate the class averages. The average normalized gain <g> has been *independently* utilized by Hovland et al. (1949), Gery (1972), and Hake (1998a,b). The latter usage was derived from a strictly *empirical* observation that a consistent analysis over diverse student populations with widely varying initial knowledge states, as gauged by <%pre>, could be obtained by taking <g> as a rough measure of the relative effectiveness of a course in promoting conceptual understanding. That inference was bolstered by the fact that the correlation of <g> with <pre> for the 62 survey courses was a very low +0.02.

---

* In my opinion, the failure to *operationally* define [even despite the anti-positivist vigilantes (Phillips 2000)] key descriptors such as, e.g., "direct instruction," "discovery learning," "inquiry," "constructivism," "traditional," "progressive," and "scientifically based, " contributes to the confusion that permeates much of the current debate over educational methods. I discussed this problem at some length in Hake (2004a).

B. Quantitative Evidence

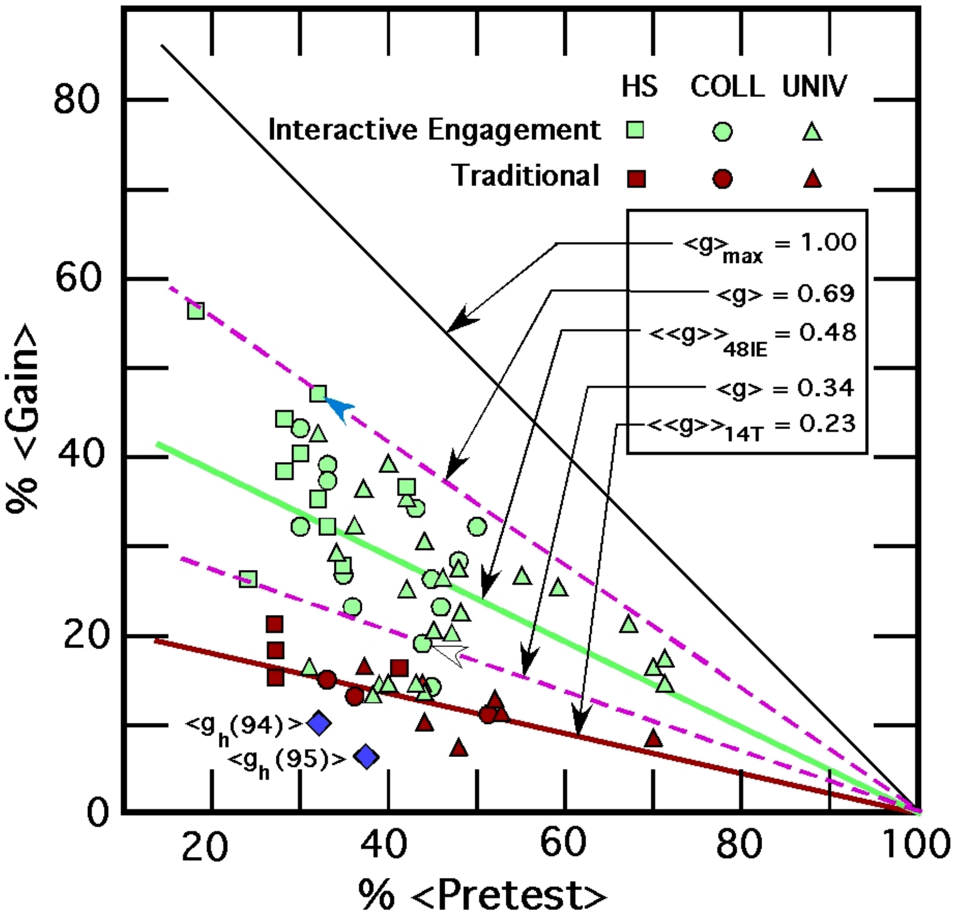

Fig. 1. Data of Hake (1998a,b) as plotted in Fig. 1 of Hake (2002a), with the addition of "*hypothesized* <g>'s" (blue diamonds) for high-school physics classes [Hake (2000a)].

For the data of Hake (1998a,b):

a. 62 courses (N = 6542) were surveyed;

b. 48 IE courses enrolling N = 4458 students yielded : <<g>> (48 IE) = 0.48 ± 0.14sd;

c. 14 T courses enrolling N = 2084 students yielded : <<g>> (14 T) = 0.23 ± 0.04sd;

d. <<g>> (IE) is over twice that of <<g>> (T) and differs by over 6 sd's of <<g>>14T and almost two sd's of <<g>>48IE, reminiscent of differences seen in comparing instruction delivered to students in large groups with one-on-one instruction (Bloom 1984);

e. tests employed were the valid and consistently reliable *Mechanics Diagnostic* [Halloun & Hestenes (1985a,b)] or *Force Concept Inventory* [Hestenes et al. (1992)], both research based;

f. Cohen's (1988) $d = [\langle\langle g\rangle\rangle_{48IE} - \langle\langle g\rangle\rangle_{14T}] / [(sd_{48IE}{}^2 + sd_{14T}{}^2)/2]^{0.5} = 2.43$ [Hake (2002a)]; much larger than any recorded in the meta-meta-anlaysis of social science research by Lipsey & Wilson (1993). Eight reasons for this relatively large d value are given in Hake (2002b).

C. Corroboration

Normalized gain differences between T and IE courses that are consistent the data shown above [Hake (1998a,b))] have been reported by many other physics education research groups, as referenced in Hake (2002a,b).

D. Is the Above Pre/post Testing Evidence "Scientifically Based"?

Such pre/post testing does not meet the U.S. Dept. of Education's (USDE's) much contested [for a review see Hake (2004c)], "gold standard" of randomized control trials, but would nevertheless probably pass muster at the USDE's "What Works Clearing House" < http://www.w-w-c.org/ > as "quasi-experimental studies [Shadish et al. (2002)] of especially strong design" [see < http://www.w-w-c.org/reviewprocess/standards.html >]. The strength of the design derives partly from (a) the use of reasonably well-matched control groups offered by the "traditional" courses, and (b) corroboration of the results by many different research groups (see above). Despite rampant pre/post paranoia [Hake (2004e,f; 2005d)], pre/post assessments of student learning are being more and more utilized in fields such as astronomy, economics, biology, chemistry, computer science, physics, and engineering [see Hake (2002a,b; 2004g,h)].

III. Evidence for the Superiority of Interactive Engagement (IE) or "Guided Inquiry " Instruction over Direct Science Instruction (DSI) in Conceptually Difficult Areas of K-12 Science Education : references from Hake (2004a), e.g.:

A. Meta-meta-analysis of Lipsey & Wilson (1993)

The eleven K-12 science-education studies listed in Table 1 of Lipsey & Wilson (where the test group is characterized by reform methods) yield a total N = 888 students and average effect size $\langle d\rangle = 0.36$ [Cohen (1988)]. Most of these studies include grades 4 or 6 to 12 with the effect size control group being traditional direct instruction and the measurement unit being "achievement" or "learning" (presumably as measured by tests). Cohen's (1988, p. 24) rule of thumb – based on typical results in social science research – that d = 0.2, 0.5, 0.8 imply respectively "small," "medium," and "large"effects. But Cohen cautions that the adjectives "are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation."

B. Review by Lawrence Lowery (2003)

C. Review by Heidi Doss-Hammel (2004)

D. Review by Lopez & Schultz (2001)

E. References in AAAS (1993, 2004)

F. References in NRC (1996; 1997a,b; 1999, 2000, 2001, 2003)

G. References in Bransford et al. (2000)

H. References in Anderson (2002)

I. References in Jorgenson & Vanosdall (2002)

J. Review (in progress) of Levy & Century (2005)

It is important to note that none of the above research included in A – I above includes the strawman of unguided "discovery learning," purportedly shown by Klahr & Nigam (2004) to be inferior to DSI.

IV. But the Research (Sections II & III) Showing the Superiority of Interactive Engagement (IE) and Guided Inquiry (GI) Methods to Direct Science Instruction Is Ignored

A. In California

Direct Science Instruction (DSI) predominates in CA because of the DSI-orientation of the CA State Board of Education and Curriculum Commission as discussed in Hake (2004a).

B. In the U.S.

There are at least seven reasons why DSI threatens to predominate nationally under the aegis of the No Child Left Behind Act:

1. Most IE and GI 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 philosophers Dennis Phillips [Shavelson, Phillips, Towne, & Feuer (2003)] and Michael Scrivin (2004); 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 RTC 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).

2. The heavily publicized [Adelson (2004), Begley (2004a,b), Cavenaugh (2004a,b), Tweed (2004a,b), USDE (2004)] **research of Klahr and Nigam (KN) (2004) is widely misinterpreted as demonstrating the general superiority of direct instruction.**

Many DI enthusiasts [e.g., Mathematically Correct (2005), Bishop (2004)] misleadingly imply that KN’s research demonstrates the superiority of their favored drill and practice methods over any other type of science instruction. Commenting on Klahr & Nigam, Bishop (2004) opines: “Deliberate, direct instruction is more effective yet again. Surprise, surprise. Replicate a replicable experiment and you get the same results. It's the scientific approach.”

But the “guided inquiry” and “interactive engagement” methods of Sections II and III above have little to do with KN’s strawman of *Extreme* Discovery Learning (EDL), in which there is almost no teacher guidance. KN showed, not surprisingly, that EDL is inferior to "direct instruction" for increasing third and fourth grade children's effective use of the control of variables strategy, a so-called "process skill." It might be interesting for Klahr & Nigam to extend their study to more guided forms of "discovery learning" and to children's acquisition of "operative knowledge" [Arons (1983)].

Consistent with the above, Adelson (2004) wrote: “Psychologist Rich Shavelson, professor of education and (by courtesy) psychology at Stanford University, notes that totally unguided discovery of the type used in [KN’s] study is rarely used in the classroom.” Still, he says, “This study uses a strong research design. I'd like to see a replication with [the more typical] guided discovery. Plus, the extent to which results would travel to classrooms with varying teacher quality, opportunity to learn, et cetera, has yet to be found out.”

Klahr himself is not the DSI radical painted by DI zealots such as Bishop (2004) and Mathematically Correct (2005)]. Cavenaugh (2004a) wrote: “While David Klahr . . . . believes that complex science lessons often require a more direct type of instruction, he also cautions against too rigid an adherence to either method by teachers or administrators. ‘It depends on what's being taught,’ Mr. Klahr, a psychology professor at Carnegie Mellon, said in an interview.”

I agree with David Klahr's caution that the appropriate method of science instruction depends on what's being taught. Teachers, to be effective, need to use different approaches (e.g., didactic lectures, coaching, collaborative discussions, and Socratic dialogue) to fit the classroom occasions and diverse natures of their students. Each method has its strengths and weaknesses for each type of student, but in the hands of a *skilled teacher* each can be made to compliment the other methods so as to advance *every* student's learning. A skilled teacher might *lecture* on material that can be rote memorized, *coach* skills such as typing or playing a musical instrument, and use *Socratic dialogue* or *collaborative discussions* (or some other "interactive engagement" method) to induce students to construct their conceptual understanding of difficult counter-intuitive material such as Newton's Laws.

3. **It's easier to test for rote memorized material implanted by DSI than for conceptual understanding of science and its methods induced by interactive engagement methods**.

Among the lessons of the physics education reform effort [Hake (2002a)] is that full scale qualitative and quantitative research projects *by disciplinary experts* are required to develop valid and consistently reliable tests of conceptual understanding [see e.g., the exemplary research of Halloun & Hestenes (1985a,b) in developing the *Mechanics Diagnostic* test, precursor to the widely used *Force Concept Inventory* [Hestenes et al. (1992)]. That amateurs cannot develop such tests has been painfully demonstrated by the California Standardized Testing and Reporting (STAR) Program [for a discussion see Woolf (2005a,b,c) & Hake (2005c)].

4. The U.S. Department of Education (USDE) is evidently bereft of advisors from physical sciences.

Advisors to the USDE consist primarily of psychologists, psychometricians, statisticians, economists, sociologists, administrators, medical specialists, policy analyists, and education specialists, most with a proclivity towards “Random Control Trials” (RCT’s). As far as I am aware, no physical scientists are members of the:

a. Technical Advisory Group < http://www.w-w-c.org/whatwedo/factsheet.pdf > (68kB) for the “What Works Clearinghouse” < http://www.w-w-c.org/ >.

b. Advisory board < http://www.excelgov.org/displayContent.asp?Keyword=prppcAdvisory > for the "Coalition for Evidence-Based Policy" (CEBP) < http://www.excelgov.org/displayContent.asp?Keyword=prppcHomePage >.

The CEBP’s board of advisors include luminaries such as political economist David Ellwood (Harvard); statistician Robert Boruch (Univ. of Pennsylvania); former FDA commissioner David Kessler (Univ. of California – San Francisco); past American Psychological Association president Martin Seligman (University of Pennsylvania); psychologist Robert Slavin (Johns Hopkins); economics Nobelist Robert Solow (MIT); and progressive-education critic Diane Ravitch (2000). The qualifications of all the above for overseeing the assessment of science/math instruction are, in my view, problematic.

5. Douglas Carnine is a member of the Technical Advisory Group for the “What Works Clearinghouse.”

Carnine is perhaps the Nation’s most prominent advocate of Direct Instruction [see e.g., Carnine (2000)]. Carnine played a leading role in undermining effective math instruction in California [see, e.g., Schoenfeld (2003)], and, I suspect, is now poised to attempt the same on a national scale for the 3 R's and for science instruction.

The Fordham Foundation's Chester Finn introduces Carnine's (2000) paper “Why Education Experts Resist Effective Practices (And What It Would Take to Make Education More Like Medicine),” by eulogizing: “After describing assorted hijinks in math and reading instruction, Doug devotes considerable space to examining what educators did with the results of ‘Project Follow Through,’ one of the largest education experiments ever undertaken. This study compared constructivist education models with those based on direct instruction. One might have expected that, when the results showed that direct instruction models produced better outcomes, these models would have been embraced by the profession. Instead, many education experts discouraged their use."

But according to Lagemann (2000), the results of Project Follow Through were inconclusive. She writes: "Some experiments ended inconclusively. One of these was the 'planned variation' strategy used in implementing Project Follow Through in the late 1960s . . . . as education researcher David K. Cohen . . . [1970]. . . concluded, all the Follow Through experiment really demonstrated was that power in education was so decentralized that the controls necessary for experimentation were virtually impossible to maintain."

6. Psychologist Grover Whitehurst, director of the U.S. Education Department's Institute of Education Sciences
< http://www.ed.gov/about/offices/list/ies/index.html?exp=0 >**), evidently believes that "In science education, there is almost nothing of proven efficacy."**

The above quote is from Sharon Begley's (2004b) article "To Improve Education, We Need Clinical Trials To Show What Works." Whitehurst (as well of most of those at the US Dept. of Education) is evidently either unaware or dismissive of the evidence cited in Sections II & III above.

7. Campbell's Law [Campbell (1975)]:
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.

Nichols & Berliner (2005) cite the voluminous evidence of the malfunctioning of the NCLB as consistent with Campbell's Law and also (incidentally) consistent with Robert Sternberg's (2004) "Dozen Reasons Why the No Child Left Behind Act is Failing Our Schools" [see also Hake (2004d)].

V. Is There Any Hope that NCLB Will *Not* Promote DSI Throughout the Nation?

According to Cavenaugh (2004b), the NRC's Center for Education, directed by Martin Orland, is overseeing three studies aimed at exploring how students learn most effectively in science, and how it is best taught and tested. The studies were stimulated by the NCLB's planned testing of science achievement to commence in 2007.

1. One study is to provide states with "practical advice" about how to design tests that will ask students to demonstrate a broad range of skills in science, without encouraging states or districts to scale back their curricula in the subject. Meryl Bertenthal, a senior program officer in the NRC's Center for Education is the study's director.

2. A second project is focused on increasing understanding of how students learn science, with a particular emphasis on kindergarten through 8th grade. The committee working on that project will examine existing research, identify areas in which new research is needed, and determine what that body of work suggests about how science subjects should be taught.

3. The third study will look at the role that science laboratories should play in the high school classroom. That question has drawn renewed interest among teachers, administrators, and researchers recently, partly because of speculation that the No Child Left Behind Act's testing requirements may compel some districts to scale back classroom experimentation in favor of more direct forms of instruction.

I hope that the NRC's expert science-education committees can switch the USDE's Direct Instruction juggernaut onto the guided inquiry (NOT discovery) track advised by the NRC (1996, 1997a, 1997b, 1999, 2000, 2001, 2003) itself.

Prospects for such redirection seem at least possible after NRC's publication of Donovan & Pellegrino (2003). The latter appreciate the most convincing evidence (Section II) for the superiority of guided inquiry methods in promoting students' conceptual understanding: pre/post-testing research *by disciplinary experts* using valid and consistently reliable research-based tests [see e.g. Halloun & Hestenes 1985a,b)] and reasonably well-matched control groups. Previous NRC's expert science education committees [e.g., McCray et al. (2003), Labov (2003)] appeared to be oblivious of such evidence (Hake 2003a,b).

VI. Why Bother With K-12 Science Instruction?

According to Adelson (2004), Klahr notes that his research comes at a critical juncture because, in Klahr's words [my *italics*], "The United States produces proportionately fewer scientists than many other 'competitor' countries, so better science teaching is certainly a national priority if we are to maintain our scientific leadership. Early mastery of the basics of the scientist's toolkit can help kids to understand and appreciate science. *More generally, a critical understanding of the difference between good and bad science is essential to informed adult decisions* in the marketplace *and in the voting booth*."

I agree completely with Klahr's last sentence. In the "The General Population's Ignorance of Science Related Societal Issues: A Challenge for the University" [Hake (2000b)], I cited voluminous evidence for the appalling science illiteracy of the general population, even including graduating seniors at Harvard and MIT.

But why should one be concerned about the science illiteracy of the general population? Because life-threatening science-related societal problems [e.g., terrorism; overpopulation (doubles about every 35 years); threat of weapons of mass destruction; human welfare (starvation, homelessness, unemployment, drugs, epidemics, AIDS, etc.); pollution of air, water, land, food; global warming; ozone depletion; deforestation; loss of biodiversity; etc., etc., etc.] cannot be resolved when a scientifically illiterate society elects scientifically illiterate leaders.

References

NOTE: “beckered” references [see e.g., Begley (2004a,b)] are those that I have copied into open and relatively permanent discussion-list archives, following the lead of mathematician Jerry Becker. I think such circulation constitutes "fair use" of copyrighted material under section 107 of the US Copyright Law.

AAAS. 1993. *Benchmarks For Science Literacy*. Oxford University Press; online at < http://www.project2061.org/tools/benchol/bolintro.htm >. See especially Chapter 15 "The Research Base" at < http://www.project2061.org/tools/benchol/ch15/ch15.htm >.

AAAS. 2004. *Project 2061*, online at < http://www.project2061.org/default_flash.htm > . Especially:
(a) "Research on Teaching and Learning" < http://www.project2061.org/research/learning.htm >,
(b) "Curriculum Materials" < http://www.project2061.org/research/curriculum.htm >,
(c) "Testing and Assessment" < http://www.project2061.org/research/assessment.htm >,
(d) "Learning Goals" < http://www.project2061.org/research/goals.htm > .

Adelson, R. 2004. “Instruction versus exploration in science learning: Recent psychological research calls ‘discovery learning’ into question,” *Monitor On Psychology* **35**(6): 34; online at < http://www.apa.org/monitor/jun04/instruct.html >.

Anderson, R.D. 2002. "Reforming Science Teaching: What Research says about Inquiry," *Journal of Science Teacher Education*, **13**(1): 1-12; based on a commissioned paper prepared for the Center for Science, Mathematics and Engineering Education at the National Research Council; online to subscribers at < http://www.kluweronline.com/issn/1046-560X/ >. Discusses various meanings of "inquiry."

Arons, A.B. 1983. "Achieving Wider Scientific Literacy," *Daedalus*, Spring. Arons wrote: "Researchers in cognitive development describe two principle classes of knowledge: figurative (or declarative) and operative (or procedural). ‘Declarative knowledge’ consists of knowing ‘facts,’ for example, that the moon shines by reflected sunlight, that the earth and planets revolve around the sun . . . . ‘operative knowledge’, on the other hand, involves understanding the source of such declarative knowledge (How do we know the moon shines by reflected sunlight? Why do we believe the earth and planets revolve around the sun when appearances suggest that everything revolves around the earth? . . . .) and the capacity to use, apply, transform, or recognize the relevance of the declarative knowledge to new or unfamiliar situations.

Begley, S. 2004a. "The Best Ways to Make Schoolchildren Learn? We Just Don't Know." Wall Street Journal, 10 December, page B1; online to Wall Street Journal subscribers at < http://users1.wsj.com/lmda/do/checkLogin?a=t&url=http%3A%2F%2Fonline.wsj.com%2Farticle%2F0%2C%2CSB110263537231796249%2C00.html%3Femailf%3Dyes >. Beckered at < http://lists.asu.edu/cgi-bin/wa?A2=ind0501&L=aera-l&T=0&F=&S=&P=646 >. See also Anne Tweed's (2004b) response and the later article [Begley (2004b)].

Begley, S. 2004b. "To Improve Education, We Need Clinical Trials To Show What Works," Wall Street Journal, 17 December, page B1; online to Wall Street Journal subscribers at <http://users2.wsj.com/lmda/do/checkLogin?a=t&url=http%3A%2F%2Fonline.wsj.com%2Farticle%2F0%2C%2CSB110322729667502382%2C00.html%3Femailf%3Dyes >. Beckered at < http://lists.asu.edu/cgi-bin/wa?A2=ind0501&L=aera-l&T=0&F=&S=&P=789 >. See also Begley (2004a).

Bishop, W. 2004a. "Direct Instruction in Science" Math-Teach post of 26 Jul 2004 10:43: am ; online at < http://mathforum.org/kb/message.jspa?messageID=1492024&tstart=0 >.

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

Bransford, J.D., A.L. Brown, R.R. Cocking, eds. 2000. *How people learn: brain, mind, experience, and school.* Nat. Acad. Press; online at < http://books.nap.edu/books/0309070368/html/index.html >. This is an update of the earlier 1999 edition.

Burkhardt, H. & A.H. Schoenfeld. 2003. "Improving Educational Research: Toward a More Useful, More Influential, and Better-Funded Enterprise," *Educational Researcher* **32**(9): 3-14; online at < http://www.aera.net/publications/?id=401 >.

Campbell, D. T. 1975. “Assessing the impact of planned social change,” in G. Lyons, ed. , *Social research and public policies: The Dartmouth/OECD Conference,* Chapter 1, pp. 3-45. Dartmouth College Public Affairs Center, p. 35; online at Western Michigan University as Occasional Paper #8 at < http://www.wmich.edu/evalctr/pubs/ops/ >, or download directly at < http://www.wmich.edu/evalctr/pubs/ops/ops08.pdf > (196 kB).

Carnine, D. 2000. “Why Education Experts Resist Effective Practices (And What It Would Take to Make Education More Like Medicine),” online as a 52kB pdf at < http://www.edexcellence.net/foundation/global/found.cfm?author=72&keyword=&submit=Search >.

Cavanagh, S. 2004a. "NCLB Could Alter Science Teaching," *Education Week* **24**(11): 1, 12-13, November 10; online at < http://www.edweek.org/ew/articles/2004/11/10/11science.h24.html > . Beckered at < http://lists.asu.edu/cgi-bin/wa?A2=ind0411&L=aera-l&T=0&F=&S=&P=1219 >. See also Anne Tweed's (2004a) response and the later article [Cavanagh (2004b)].

Cavanagh, S. 2004b. "Research-Council Studies to Explore Teaching and Testing of Science," *Education Week* **24**(11): 12-13, November 10. < http://www.edweek.org/ew/articles/2004/11/10/11nrc.h24.html >. Beckered at < http://lists.asu.edu/cgi-bin/wa?A2=ind0411&L=aera-l&T=0&F=&S=&P=1328 >.

Cavanagh, S. 2005. “As Test Date Looms, Educators Renewing Emphasis on Science,” *Education Week* **24**(29): 1, 14-15, March 30; online at < http://www.edweek.org/ew/articles/2005/03/30/29science.h24.html >. Beckered at < http://lists.asu.edu/cgi-bin/wa?A2=ind0503&L=aera-l&T=0&O=D&P=1919 >.

Cobb, P., J. Confey, A. diSessa, R. Lehrer, L. Schauble. 2003. "Design Experiments in Educational Research," *Educational Researcher* **32**(1): 9-13; online at < http://www.aera.net/publications/?id=393 >.

Cohen, D.K. 1970. “Politics and Research: Evaluation of Social Action Programs in Education,” *Review of Educational Research* **40**: 231.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum, 2nd ed.

Doss-Hammel, H. 2004. "Overview of the Impact of Activity-Based Teaching Strategies on Learning Science," San Diego State University Center for Research in Mathematics & Science Education; online at < http://www.sci-ed-ga.org/standards/index.html > (several entries down from the top).

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 >.

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

Hake, R.R. 1998a. "Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *Am. J. Phys.* **66**: 64-74; online as ref. 24 at < http://www.physics.indiana.edu/~hake >, or download directly by clicking on < http://www.physics.indiana.edu/~sdi/ajpv3i.pdf > (84 kB).

Hake, R.R. 1998b. "Interactive-engagement methods in introductory mechanics courses," online as ref. 25 at < http://www.physics.indiana.edu/~hake >, or download directly by clicking on < http://www.physics.indiana.edu/~sdi/IEM-2b.pdf > (108 kB). Submitted on 6/19/98 to the Physics Education Research Supplement (PERS) to Am. J. Phys. but rejected on the grounds that its transparent data tables were "impenetrable" ! This paper is a crucial companion paper to Hake (1998a): average pre/post test scores, standard deviations, instructional methods, materials used, institutions, and instructors for each of the survey courses of Hake (1998a) are tabulated and referenced. In addition the paper includes: (a) case histories for the seven IE courses of Hake (1998a) 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.

Hake, R.R. 2000a. "The Need for improved physics education of teachers: FCI pretest scores of graduates of high school physics courses," Physics Education Research Conference 2000: Teacher Education, Univ. of Guelph, August 2-2; online at < http://www.sci.ccny.cuny.edu/~rstein/perc2000.htm >, or ref. 12 at < http://www.physics.indiana.edu/~hake/ >, or download directly by clicking on < http://www.physics.indiana.edu/~hake/PERC2000-HSTeach-5.pdf > (928 kB).

Hake, R.R. 2000b. "The General Population's Ignorance of Science Related Societal Issues: A Challenge for the University," *AAPT Announcer* **30**(2): 105; online as ref. 11 at < http://www.physics.indiana.edu/~hake > or download directly by clicking on < http://www.physics.indiana.edu/~hake/GuelphSocietyG.pdf > (2.1MB). Based on an earlier libretto with the leitmotiv: "The road to U.S. science literacy begins with effective university science courses for pre-college teachers." The opera dramatizes the fact that the failure of universities throughout the universe to properly educate pre-college teachers is responsible for our failure to observe any signs of extraterrestrial intelligence.

Hake, R.R. 2002a. "Lessons from the physics education reform effort," *Ecology and Society* **5**(2): 28; online at < http://www.ecologyandsociety.org/vol5/iss2/art28/ >. *Ecology and Society* (formerly *Conservation Ecology*) is a free online "peer-reviewed journal of integrative science and fundamental policy research" with about 11,000 subscribers in about 108 countries.

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 ref. 29 at< http://www.physics.indiana.edu/~hake/ > or download directly by clicking on < http://www.physics.indiana.edu/~hake/Hake-SriLanka-Assessb.pdf > (84 kB).

Hake, R.R. 2003a. "NRC's CUSE: Stranded on Assessless Island?" post of 3 Aug 2003 12:52:16-0700 to ASSESS, Biopi-L, Chemed-L, EvalTalk, PhysLrnR, and POD; online at < http://listserv.nd.edu/cgi-bin/wa?A2=ind0308&L=pod&F=&S=&P=391 >. That post was later sent to AERA-D, STLHE-L, Phys-L, and Biolab.

Hake, R.R. 2003b. "Spare Me That Passive-Student Lecture," post of 1 Oct 2003 17:07:48-0700 to AERA-D, ASSESS, Biopi-L, Biloab, FYA, Chemed-L, EvalTalk, Phys-L, PhysLrnR, Math-Teach, and POD; online at < http://listserv.nd.edu/cgi-bin/wa?A2=ind0310&L=pod&O=D&P=947 >.

Hake, R.R. 2004a. "Direct Science Instruction Suffers a Setback in California - Or Does It?" *AAPT Announcer* **34**(2): 177; online as reference 33 at < http://www.physics.indiana.edu/~hake >, or download directly by clicking on < http://www.physics.indiana.edu/~hake/DirInstSetback-041104f.pdf > (420 KB) [about 160 references and 180 hot-linked URL's]. A pdf version of the slides shown at the meeting is also available at ref. 33 or can be downloaded directly by clicking on < http://www.physics.indiana.edu/~hake/AAPT-Slides.pdf > (132 kB). The history of the rise and *apparent* setback of DSI in California offers yet another example that evidence and logic are of little value in reforming education [Hake (2005b).

Hake, R.R. 2004b. “Will NCLB Promote Direct Instruction of Science?” online at < http://lists.nau.edu/cgi-bin/wa?A2=ind0411&L=phys-l&O=D&P=16592 >. Post of 11 Nov 2004 21:03:44-0800 to AP-Physics, Phys-L, PhysLrnR, Physhare. Later sent to AERA-C, AERA-G, AERA-H, AERA-J, AERA-K, AERA-L, ASSESS, EvalTalk, & Math-Learn.

Hake, R.R. 2004c. “Design-Based Research: A Primer for Physics Education Researchers,” submitted to the *American Journal of Physics* on 10 June 2004; online as reference 34 at < http://www.physics.indiana.edu/~hake >, or download directly by clicking on < http://www.physics.indiana.edu/~hake/DBR-AJP-6.pdf > (310kB). An expanded version of Section 3, "Some Physics Education Research is Design-Based Research," will appear in "Design Experiments in Education," A.E. Kelly & R.A. Lesh, eds.

Hake, R.R. 2004d. “Robert Sternberg's Critique of NCLB,” online at <http://lists.asu.edu/cgi-bin/wa?A2=ind0411&L=aera-l&T=0&F=&S=&P=1106 >. Post of 8 Nov 2004 21:33:59-0800 to AERA-G, AERA-L, AERA-K, ASSESS, EvalTalk, Math-Learn, & PhysLrnR. Later sent to AERA-C, AERA-H, & AERA-J.

Hake, R.R. 2004e. “Why Pre/Post Testing Leaves Some Unimpressed,” online at < http://lists.asu.edu/cgi-bin/wa?A2=ind0407&L=aera-d&P=R3758&I=-3 >. Post of 27 Jul 2004 14:57:12-0700 to to AERA-D, EvalTalk, EdStat-L, PhysLrnR.

Hake, R.R. 2004f. "Re: pre- post testing in assessment" online at < http://lists.asu.edu/cgi-bin/wa?A2=ind0408&L=aera-d&T=0&F=&S=&P=3354 >. Post of 19 Aug 2004 13:56:07-0700 to AERA-D, AERA-J, ASSESS, EdStat, EvalTalk, PhysLnrR, & POD.

Hake, R.R. 2004g. “Re: Measuring Content Knowledge,” online at < http://lists.asu.edu/cgi-bin/wa?A2=ind0403&L=aera-d&T=0&O=D&P=5436 >. Post of 14 Mar 2004 16:29:47-0800 to ASSESS, Biopi-L, Chemed-L, EvalTalk, Physhare, Phys-L, PhysLnrR, POD, and STLHE-L; later sent to AERA-D with a few corrections where it appears at < http://lists.asu.edu/cgi-bin/wa?A2=ind0403&L=aera-d&P=R3625&I=-3 >.

Hake, R.R. 2004h. “Re: Measuring Content Knowledge,” online at < http://listserv.nd.edu/cgi-bin/wa?A2=ind0403&L=pod&O=A&P=17167 >. Post of 15 Mar 2004 14:29:59-0800 to ASSESS, EvalTalk, Phys-L, PhysLrnR, & POD.

Hake, R.R. 2005a. “Will the NCLB Tend to Propagate California's Direct Science Instruction Throughout the Entire Nation?” online at < http://lists.asu.edu/cgi-bin/wa?A2=ind0501&L=aera-l&P=R1893&I=-3 >. Post of 14 Jan 2005 17:58:55-0800 to AERA-C, AERA-D, AERA-G, AERA-H, AERA-J, AERA-K, AERA-L, AP-Physics, ASSESS, Biopi-L, Chemed-L, EvalTalk, Phys-L, Physhare, POD, & STLHE-L.

Hake, R.R. 2005b. “Will Evidence and Logic Reform Education? (was California standards test in physics),” online at < http://lists.nau.edu/cgi-bin/wa?A2=ind0502&L=phys-l&P=R1541 >. Post of 3 Feb 2005 11:29:26-0800 to AERA-D, AERA-H, AERA-K, AERA-L, AP-Physics, ASSESS, Chemed-L, EvalTalk, Math-Learn, Phys-L, Physhare, POD.

Hake, R.R. 2005c. “Re: WriteCDE (California Standards Test, Physics 2003/04)” online at < http://lists.nau.edu/cgi-bin/wa?A2=ind0502&L=phys-l&P=R15655 >. Post of 18 Feb 2005 11:10:03-0800 to Phys-L, Phys-LrnR, Physhare, and Math-Learn.

Hake, R.R. 2005d. "Carnegie Scholar Backs Pre/post Testing," online at < http://listserv.nd.edu/cgi-bin/wa?A2=ind0503&L=pod&O=D&P=23157 >. Post of 25 March to AERA-C, AERA-D, AERA-J, AERA-K, ASSESS, EdStat, EvalTalk, PhysLrnR, POD, & STLHE-L

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

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

Halloun, I., R.R. Hake, E.P Mosca, D. Hestenes. 1995. Force Concept Inventory (Revised, 1995); online (password protected) at < http://modeling.asu.edu/R&E/Research.html >. (Available in English, Spanish, German, Malaysian, Chinese, Finnish, French, Turkish, Swedish, and Russian.)

Hestenes, D., M. Wells, & G. Swackhamer, 1992. “Force Concept Inventory.” *Phys. Teach*. **30**: 141-158; online (except for the test itself) at < http://modeling.asu.edu/R&E/Research.html >. For the 1995 revision see Halloun et al. (1995).

Hovland, C.I. A.A. Lumsdaine, and F.D. Sheffield. 1949. *Experiments on mass communication.* Princeton University Press, pp: 284-289; reprinted as “A baseline for measurement of percentage change,” in *The Language of Social Research*, Lazarsfeld and Rosenberg, eds. Free Press, 1955, pp. 77-82.

Jorgenson, O. & R. Vanosdall. 2002. "The Death Of Science? What We Risk in Our Rush Toward Standardized Testing and the Three R's," *Phi Delta Kappan* **83**(8): 601-605; online at < http://www.pdkintl.org/kappan/k0204jor.htm >.

Klahr, D. & M. Nigam. 2004. “The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning,” *Psychological Science* **15**: 661-667; online, as a pdf at < http://www.psy.cmu.edu/faculty/klahr/papers.html >, or download directly by clicking on < http://www.psy.cmu.edu/faculty/klahr/KlahrNigam.2-col.pdf > (1 MB).

Labov, J.B. 2003. "Education at the National Academies," *Cell Biology Education* **2**(3); online at < http://cellbioed.org/articles/vol2no3/article.cfm?articleID=63 >.

Lagemann, E.C. 2000. *An Elusive Science: The troubling history of education research.* Univ. of Chicago Press.

Levy, A.J. & J.R. Century. 2005. “Has Inquiry Made a Difference? A Synthesis of Research on the Impact of Inquiry Science on Student Outcomes,” 3-year NSF-funded project, Educational Development Center, online at < http://cse.edc.org/work/research/inquirysynth/default.asp > (in progress).

Lipsey, M.W. & D.B. Wilson. 1993. “The Efficacy of Psychological, Educational, and Behavioral Treatment: Confirmation From Meta-Analysis,” *American Psychologist* **48**(12): 1181-1209; online at < http://www.vanderbilt.edu/cerm/MWL_web_bib.htm#methods%20and%20issues >; or directly at [< http://www.vanderbilt.edu/cerm/Amer_Psych_paper.pdf > (3.4 MB);

Lowery, L. F. 2003. “Research on Hands-On Science Programs,” online at < http://www.fossworks.com/pdfs/HandsOnScienceResearch.pdf > (572 kB) .

Lopez, R.E. & T. Schultz. 2001. “Two Revolutions in K-8 Science Education,” *Physics Today* **54**(9): 44-49; online at < http://www.aip.org/pt/vol-54/iss-9/p44.html >.

Mathematically Correct. 2005. Online at < http://www.mathematicallycorrect.com/ >. At their “Science Corner” < http://mathematicallycorrect.com/science.htm > it is stated that ““Carnegie Mellon Researchers Say Direct Instruction, Rather Than "Discovery Learning" Is Best Way To Teach Process Skills In Science.”

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

Nichols, S.L & D.C. Berliner. 2005. “The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing,” Arizona State Univ. Education Policy Studies Laboratory, online at < http:www.asu.edu/educ/epsl/EPRU/documents/EPSL-0503-101-EPRU.pdf > (1.7 MB).

NRC. 1996. *National Science Education Standards*, National Academy Press; online in HTML at < http://books.nap.edu/catalog/4962.html >.

NRC. 1997a. *Introducing the National Science Education Standards, Booklet*, National Academy Press; online in HTML at < http://books.nap.edu/catalog/5704.html >.

NRC. 1997b. *Science Teaching Reconsidered: A Handbook*, National Academy Press; online at < http://books.nap.edu/catalog/5287.html >. This is oriented towards postsecondary education, but its message that "guided inquiry' and "interactive engagement" methods are generally more effective than direct instruction applies to K-12. See especially Chapter 2, "How Teachers Teach: Specific Methods."

NRC. 1999. *Improving Student Learning: A Strategic Plan for Education Research and Its Utilization* , National Academy Press; online in HTML at < http://books.nap.edu/catalog/6488.html >.

NRC. 2000. *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*, National Academy Press; online in HTML at < http://books.nap.edu/catalog/9596.html >. See especially Bruce Alberts's (2000) forward: "A Scientists Perspective on Inquiry" for a good operational definition of "inquiry."

NRC. 2001. *Classroom Assessment and the "National Science Education Standards,"* National Academy Press; online in HTML at < http://books.nap.edu/catalog/9847.html >.

NRC. 2003. *What Is the Influence of the National Science Education Standards? Reviewing the Evidence*, *A Workshop Summary*, National Academy Press; online at < http://books.nap.edu/catalog/10618.html >.

Phillips, D.C. 2000. *Expanded social scientist's bestiary: a guide to fabled threats to, and defenses of, naturalistic social science.* Rowman & Littlefield.

Ravitch, D. 2000. *Left Back: A Century of Failed School Reforms* (Simon & Shuster). Some of Ravitch's articles are available at the Thomas B. Fordham Foundation website < http://www.edexcellence.net/foundation/global/index.cfm > ensconced under < http://www.edexcellence.net/foundation/global/found.cfm?keyword=Ravitch&x=20&y=5 >.

Schoenfeld, A.H. 2003. "Math Wars," ["almost final draft of 5 August 2003"] to appear in *2004 Politics of Education Yearbook*, edited by B.C. Johnson and W.L. Boyd; online as a 76 kB pdf at < http://gse.berkeley.edu/faculty/AHSchoenfeld/AHSchoenfeld.html >, along with some other worthwhile papers, or access directly by clicking on < http://www-gse.berkeley.edu/faculty/aschoenfeld/Math_Wars.pdf > (76 kB).

Scriven, M. 2004. "Causation," unpublished; see, e.g. "Scriven on ramdomized control groups," online at < http://interversity.org/lists/arn-l/archives/Nov2003/msg00309.html >.

Shadish, W.R., T.D. Cook, & D.T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin. A goldmine of references to the social-science literature of experimentation.

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., D.C. Phillips, L. Towne, and M.J. Feuer. 2003. "On the Science of Education Design Studies," *Educational Researcher* **32**(1): 25-28; online at < http://www.aera.net/publications/?id=393 >.

Sternberg, R.J. 2004. "Good Intentions, Bad Results: A Dozen Reasons Why the No Child Left Behind Act is Failing Our Schools," *Education Week* **24**(9): 42, 56, 27 October; online at < http://www.edweek.org/ew/articles/2004/10/27/09sternberg.h24.html >. Beckered at < http://lists.asu.edu/cgi-bin/wa?A2=ind0411&L=aera-l&T=0&F=&S=&P=964 >.

Tweed, A. 2004a. "Direct Instruction: Is It the Most Effective Science Teaching Strategy?" *NSTA Reports*, 15 December; online at < http://www.nsta.org/main/news/stories/education_story.php?news_story_ID=50045 >. Beckered at < http://lists.asu.edu/cgi-bin/wa?A2=ind0501&L=aera-l&T=0&F=&S=&P=2058 >. Response to Cavanagh (2004a). Tweed is the president of the National Science Teachers Association. (NSTA).

Tweed, A. 2004b. Letter to the Editor of the Wall Street Journal. 16 December; online at NSTA Express, 20 December 2004, < http://science.nsta.org/nstaexpress/nstaexpress_2004_12_20_tweed.htm >. Beckered at < http://lists.asu.edu/cgi-bin/wa?A2=ind0501&L=aera-l&T=0&F=&S=&P=1081 >**.** Response to Begley (2004a). Tweed is the president of the National Science Teachers Association ((NSTA).

USDE. 2004. U.S. Dept. of Education - National Education Summit Press Release: "National Summit Goal: Taking What Works into the Classroom," online at < http://www.ed.gov/news/pressreleases/2004/05/05042004a.html >.

Wiggins, G. 1990. “The Truth May Make You Free, But the Test May Keep You Imprisoned: Toward Assessment Worthy of the Liberal Arts,” *AAHE Assessment Forum*: 17-31; online at the Mathematical Association of America (MAA) project “Supporting Assessment in Undergraduate Mathematics” (SAUM) at < http://www.maa.org/saum/ > / “Getting Started With Assessment” or download directly at < http://www.maa.org/saum/articles/wiggins_appendix.html >.

Woolf, L. 2005a. "California standards test in physics," Phys-L post of 5 Jan 2005 21:59:55-0800; online at < http://lists.nau.edu/cgi-bin/wa?A2=ind0501&L=phys-l&F=&S=&P=4046 >.

Woolf, L. 2005b. "Re: California standards test in physics," Phys-L post of 8 Jan 2005 15:12:57-0800; online at < http://lists.nau.edu/cgi-bin/wa?A2=ind0501&L=phys-l&F=&S=&P=10488 >.

Woolf, L. 2005c. “Incorrect ‘correct’ answers to test questions from the CA science standards test,” online as the second entry (as of 31 March 2005) at < http://www.sci-ed-ga.org/standards/index.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 >.