International Journal for the Scholarship of Teaching and Learning

Volume 2 | Number 1 Article 21

1-2008

Can Distance and Classroom Learning Be Increased?

Richard Hake
Indiana University, Emeritus, rrhake@earthlink.net

Recommended Citation
Hake, Richard (2008) "Can Distance and Classroom Learning Be Increased?," *International Journal for the Scholarship of Teaching and Learning*: Vol. 2: No. 1, Article 21.
Available at: https://doi.org/10.20429/ijsotl.2008.020121

Can Distance and Classroom Learning Be Increased?

Abstract
Professor Scott Overmyer of Baker College, in a discussion list post, raised four points bearing on a question of interest to those involved in the Scholarship of Teaching and Learning (SoTL): *Can Distance and Classroom Learning Be Increased*? My answer: “YES” - judging from the fact that pre/post testing in courses in Newtonian mechanics has demonstrated an approximately two-standard-deviation superiority in average normalized gains for classroom “interactive engagement” methods over “traditional” classroom methods. Similarly, pre/post testing *might* demonstrate a substantive superiority over traditional classroom teaching for *both* classroom and distance education that recognize recent advances in cognitive science and emphasize learning rather than teaching. But such demonstration probably cannot be achieved if scholars of teaching and learning continue to rely on low-resolution gauges of students' learning.

Keywords
Distance learning, SoTL, Classroom learning

Can Distance and Classroom Learning Be Increased?

Richard Hake Indiana
University, Emeritus Woodland
Hills, California , USA
rrhake@earthlink.net

Abstract

Professor Scott Overmyer of Baker College, in a discussion list post, raised four points bearing on a question of interest to those involved in the Scholarship of Teaching and Learning (SoTL): *Can Distance and Classroom Learning Be Increased*? My answer: "YES" - judging from the fact that pre/post testing in courses in Newtonian mechanics has demonstrated an approximately two-standard-deviation superiority in average normalized gains <g> for classroom "interactive engagement" methods over "traditional" classroom methods. Similarly, pre/post testing *might* demonstrate a substantive superiority over traditional classroom teaching for *both* classroom and distance education that recognize recent advances in cognitive science and emphasize learning rather than teaching. But such demonstration probably cannot be achieved if scholars of teaching and learning continue to rely on low-resolution gauges of students' learning.

Introduction

According to a recent Survey of Distance Education (DE) [PRG (2007)], the revenues of DE programs in higher education grew by a mean of 15.5% in 2006, with more than a third of the enrollment increase coming from increased enrollment from students that pursue traditional classes. Despite the growing importance of DE, it has attracted relatively little attention from those involved in the Scholarship of Teaching and Learning – exceptions are Buchanan (2001), Garner et al. (2005), and Hostetter & Busch (2006).

In reaction to a discussion-list post, provocatively titled "Distance Learning: Is There Any?" [Hake (2007a)], online learning specialist Britt Watwood [private communication – he has given me permission to divulge his name] countered with a fair question: "Face-To-Face Learning: Is There Any?," and professor Scott Overmyer (2007), in his IFETS (International Forum of Educational Technology & Society) post of 12 August 2007, raised four points bearing on a broader and deeper question that should be of general interest to those involved in the Scholarship of Teaching and Learning: *Can Distance and Classroom Learning Be Increased*? I respond to Overmyer's points below:

Responses to Overmyer's Four Points

1. "Definitive pre- and post-testing of what?"

What's being pre/post tested is students' performance on valid and consistently reliable diagnostic tests (developed by disciplinary experts) of conceptual understanding. For discussion of the development, administration, and interpretation of such tests for undergraduate astronomy, biology, chemistry, economics, engineering, geoscience, math, and physics see the reviews Hake (2005, 2007b, 2008a) and the discussion-list posts Hake (2004; 2007d,e; 2008b).

https://doi.org/10.20429/ijsotl.2008.020121

In my opinion such *direct* gain measurements of higher-order student learning are far superior to the *indirect* (and therefore in my view problematic) gauges have been utilized by education researchers: e.g., end-of-course exams and course grades; Student Evaluations of Teaching (SET's); Reformed Teaching Observation Protocol (RTOP) [MacIsaac (2008)]; National Survey Of Student Engagement [NSSE (2008)]; Faculty Survey of Student Engagement (FSSE) plus CLAss Survey of Student Engagement (CLASSE) [Rhem (2007)]; Student Assessment of Learning Gains (SALG) [Seymour et al. (2005)]; and Knowledge Surveys [Nuhfer & Knipp (2003)].

2. "This . . .[dearth of pre/post testing in DE is]. . . . **, of course, in sharp contrast with the 'extensive' pre- and post-testing of face-to-face education (which I assert) is also not done."**
Although Overmyer is (unfortunately) correct in stating that *extensive* pre/post testing is not done in face-to-face (FTF) education - see e.g., "The Physics Education Reform Effort: A Possible Model for Higher Education" [Hake (2005)], there appears to be more pre/post testing in FTF education than in distance education (DE).

For example, in "Design-Based Research in Physics Education Research: A Review" [Hake (2008a)], I referenced 23 peer reviewed articles by physics education researchers who have all found pre-to-posttest average normalized conceptual gains for interactive engagement pedagogy to be much larger than those for traditional instruction.

As for pre/post testing in DE, in response to my post "Distance Learning: Is There Any?" [Hake (2007a)] I received an email from Richard Walters of UC-Davis [he has given me permission to give his name and references to his publications] to the effect that pre/post testing had been utilized to compare the effectiveness of an introductory computing course as given by distance and classroom instruction [Walters (1996), Walters & Reed (1997)]. Walters & Reed reported no significant difference in the pre-to-posttest gains for delivery by distance and classroom education, a manifestation of the "No Significant Difference Phenomenon" [Russell (2001)].

Except for Walters (1996) & Walters & Reed (1997), direct measurement of learning gain by pre/post testing has, as far as I know (please correct me if I'm wrong), not been employed to gauge the effectiveness of DE, despite endorsement of pre/post testing by, e.g.:

a. Margaret Driscoll (2001) of IBM;

b. Susan Millar (2001), director of the "Learning through Evaluation, Adaptation, and Dissemination" (LEAD) Center at the Univ. of Wisconsin - Madison, through her reference to the "Field-tested Learning Assessment Guide" [FLAG (2008)];

c. Lloyd Bond (2005), senior Carnegie Scholar, who wrote [my insert at ". . . . [insert]. . . . . . .":

> If one wished to know what knowledge or skill Johnny has acquired over the course of a semester, it would seem a straightforward matter to assess what Johnny knew at the beginning of the semester and reassess him with the same or equivalent instrument at the end of the semester. It may come as a surprise to many that measurement specialists have long advised against this eminently sensible idea. Psychometricians don't like "change" or "difference" scores in statistical analyses because, among other things, they tend to have lower reliability than the original measures themselves. Their objection to change scores is embodied in the very title of a famous paper by Cronbach and Furby (1970) "How we should measure 'change,' – or should we?". . . . [but see "Should We Measure Change? Yes!" (Hake, 2007b)]. . . .

Instead of measuring pre-to-post test gains so as to definitively gauge student *learning* in a course, distance and classroom education researchers, including those involved in SoTL, generally utilize *low-resolution* measures of students learning, such as student evaluations

of teaching, student self-assessments, and teacher-made tests and course grades. Regarding the latter, McKeachie (1987) has pointed out that end-of-course examinations and grades typically measure lower-level educational objectives such as memory of facts and definitions rather than higher-order outcomes such as critical thinking and non-algorithmic problem solving.

For example, in traditional physics introductory courses, student grades depend primarily on students' lower-order abilities in rote-memorization, recipe following, and algorithmic problem-solving, as witness the fact that many students receive A's and B's, while at the same time achieving very low normalized gains on diagnostic tests of the conceptual understanding that they should have been led to construct during the course.

Among DE references which, as far as I know, generally fail to mention pre/post testing of learning are these seven:

(1) Distance education resources recommended by EdResMeth subscribers in a thread "Online Learning Evaluation Articles," accessible at the October 2007 EdResMeth archives at <http://listserv.uconn.edu/edresmeth-l.html> - viz., *What Works in Distance Education*: *Guidelines* [O'Neil (2005)] and "Assessing the Quality of Online Instruction: Integrating Instructional Quality and Web Usability Assessments" [Ciavarelli (2004)].

(2) Two interesting DE articles referenced at the *Open and Distance Learning Association of Australia* (ODLAA) site <http://odlaa.une.edu.au/> (this link may be erratic): (a) "Rethinking space and time: The role of Internet technology in a large lecture course" [Harley et al. (2004)]; and (b) *Online Nation: Five Years of Growth in Online Learning* [Allen & Seaman (2007)].

(3) *Distance Learning Clearinghouse* of the University of Wisconsin - Extension <http://www.uwex.edu/disted/journals.cfm>. Over 35 distance-learning journals and newsletters listed at <http://www.uwex.edu/disted/journals.cfm>. See especially the 16 references at "Evaluation and Assessment" <http://www.uwex.edu/disted/evaluation.cfm>. I can find no mention of pre/post testing in any of those, except indirectly by Susan Millar (2001) in her reference to the "Field-tested Learning Assessment Guide" [FLAG (2008)].

(4) *eLearn Magazine* <http://www.elearnmag.org/> with archives at <http://www.elearnmag.org/subpage.cfm?section=archives> (a publication of the *Association for Computing Machinery* <http://www.acm.org/>).

(5*) Learning Circuits* <http://www.learningcircuits.org/mission.html>, sponsored by *American Society of Training and Development* <http://www.astd.org/astd> [No apparent mention of pre/post testing except for that by Margaret Driscoll (2001)].

(6) *Sloan Consortium* <http://www.sloan-c.org/> - see e.g., Sloan (2005a,b); Allen & Seaman (2007), Allen et al. (2007).

(7) Charles Partridge's (2007) post "The Effectiveness of Distance Education vs Classroom Instruction: A Review of the Evidence," which contains 7 references relevant to DE.

Regarding "(7)" above, Partridge considers three meta-analyses of classroom vs distance education: Bernard et al. (2004), Lou et al. (2006), and Sitzmann et al. (2006). According to Bernard et al. (p. 390)]:

> Achievement outcomes were objective measures - standardized tests, researcher-made or teacher-made tests, or a combination of these - that assessed the extent to which students had achieved the instructional (i.e., learning) objectives of a course. While most measured the acquisition of content knowledge, tests of comprehension and application of knowledge were also included.

Since Lou et al. and Sitzman et al. used methodology similar to that of Bernard et al., the above quote probably also applies to those two studies. Thus among the uncertainties in "achievement outcomes" of those three meta-analyses are:

(a) the average validity and reliability of the "objective measures" is unknown,

(b) "content knowledge" may mean lower-order rote-memorized knowledge, and

(c) average end-of-course outcomes do not measure average *course-induced learning* unless the students' average pre-course knowledge and understanding was negligible.

Thus, in my opinion, these achievement outcomes offer much less resolution of *course-induced learning* than do students' pre-to-posttest gain on valid and consistently reliable diagnostic tests of conceptual understanding developed by disciplinary experts.

And Partridge's conclusion:

> The main finding in this literature is that there is little difference in the educational . . . . outcomes of students of well-designed distance education programs vs. classroom learning. . . . .

may simply be due to insufficient resolution in the measurement of learning outcomes. Low resolution detectors cannot distinguish lower- from higher-order learning, and it is not unreasonable to think that lower-order learning is about the same in distance and classroom education.

3. "Perhaps a better question is: 'Learning via Institutions Of Higher Education: Is There Any?' "

Since many institutions of higher education offer distance education courses, an improved question might be the one suggested to me by Britt Watwood "Face-to-Face Learning: Is there any?" or, since few would dispute that learning usually does occur in one-on-one, face-to-face tutoring [Bloom (1984)], an even better question might be:

"CLASSROOM LEARNING: IS THERE ANY?" . . . . (1)

As discussed in my reviews Hake (2005, 2007b, 2008a), physics education researchers have shown that for traditional (T) passive-student lecture courses in Newtonian mechanics, the average pre-to-posttest gains <g> on standardized tests of conceptual understanding are appallingly low. But "interactive engagement" (IE) courses achieve about a two-standard deviation superiority in <g> over T courses [cf., Bloom (1984)]. Here:

a. The average normalized gain <g> is the average *actual* gain [<%post> - <%pre>] divided by the *maximum possible average gain* [100% - <%pre>], where the angle brackets indicate the class averages. For a detailed discussion of the rationale and halfcentury-old history of the normalized gain see "Should We Measure Change? Yes! [Hake (2007b)].

b. IE courses are *operationally* defined [even despite the anti-positivist vigilantes (Phillips, 2000)] as those designed at least in part to promote conceptual understanding through *continual* interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield *immediate* feedback through discussion with peers and/or instructors.

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

d. The conceptual formative evaluation tests of Newtonian mechanics were either the *Force Concept Inventory* (FCI) [Hestenes et al. (1992)] or its precursor the *Mechanics Diagnostic* (MD) [Halloun & Hestenes (1985a,b)]; in both cases developed by disciplinary experts through arduous qualitative and quantitative research, and widely recognized as valid and consistently reliable. Since these are multiple-choice tests (MCT's), they may be given to thousands of students in hundreds of courses under varying conditions in such a manner that meta-analyses can be performed, thus establishing general causal relationships in a convincing manner.

But can MCT's measure conceptual understanding and higher-order learning? Wilson & Bertenthal (2005) think so, writing (p. 94):

> Performance assessment is an approach that offers great potential for assessing complex thinking and learning abilities, but multiple-choice items also have their strengths. For example, although many people recognize that multiple-choice items are an efficient and effective way of determining how well students have acquired basic content knowledge, many do not recognize that they can also be used to measure complex cognitive processes. For example, the *Force Concept Inventory* . . . [Hestenes et al. (1992)] . . . is an assessment that uses multiple-choice items to tap into higher-level cognitive processes.

So my answer to the question "Classroom Learning: Is There Any?" is:

PROBABLY NOT MUCH FOR CONCEPTUALLY DIFFICULT SUBJECTS
TAUGHT IN THE TRADITIONAL PASSIVE-STUDENT LECTURE MODE. . . . . . .(2)

But definitive pre/post test measurement of students' higher-order learning of Newtonian mechanics has demonstrated an approximately two-standard deviation improvement by classroom "interactive engagement" courses over traditional classroom instruction.

Similarly pre/post testing *might* demonstrate a substantive improvement in student learning over traditional classroom instruction for *both* DE courses and classroom courses that take advantage of advances in cognitive science [e.g., Bransford et al. (2000), Donovan & Bradsford (2005)], and emphasize *learning* rather than *teaching* - see e.g., "From Teaching to Learning: A New Paradigm for Undergraduate Education" [Barr & Tagg (1995)]. *But such demonstration probably cannot be achieved if education researchers continue to rely on low resolution gauges of students learning.*

For example, the research of Lou et al. (2006) points to the superiority of asynchronous distance education which is optimized in the following ways (my *italics*):

a. Systematically designed interactive multimedia are used to provide more effective *student-content interaction*.

b. Collaborative discussion among students is structured using asynchronous communication media with some opportunity for peer face-to-face meetings for more effective *student-student interaction*.

c. *Student-instructor interaction* is encouraged through planned activities such as instructor participation in discussion board forums, question and answer chat sessions, and opportunity for face-to-face meetings with the instructor.

d. Students are provided with advanced information about DE courses so that they are better prepared and more ready for the DE courses.

The above features of "optimized asynchronous distance education" (OADE - pronounced "ode"), except for "d," are consistent with:

(1) The recommendations of Michael Moore (1989) for effective DE. Moore wrote (my *italics*):

> I suggested that, as a minimum, distance educators need to agree on the distinctions between three types of interaction, which I labeled *learner-content interaction, learner-instructor interaction, and learner-learner interaction.*
> . . . .
> Educators need to organize programs to ensure maximum effectiveness of each type of interaction, and ensure they provide the type of interaction that is most suitable for the various teaching tasks of different subject areas, and for learners at different stages of development.
> . . . .
> In short, *it is vitally important that distance educators in all media do more to plan for all three kinds of interaction,* and use the expertise of educators and communication specialists in both traditional media-printed, broadcast, or recorded-and newer teleconference media.

(2) The "Interactive Engagement" (IE) pedagogy has been found to be much more effective than traditional (T) passive student methods in promoting students' higher order learning in classroom physics instruction. IE methods are defined, as indicated above, as those designed at least in part to promote conceptual understanding through *continual* interactive engagement of students in heads-on (always) and hands-on (usually) activities. . . ["*learner-content interaction*"]. . . . which yield *immediate* feedback through discussion with peers . . . ["*learner-learner interaction*"]. . . and/or instructors ["*learner-instructor interaction*"]. . . . .

Thus it is possible that OADE *might* be found to be much more effective than traditional classroom education.

But Lou et al., using only *low resolution gauges* of student learning, found a weighted mean effect size for 120 OADE undergraduate courses (compared with traditional classroom education) of only $d_h = +0.058$ [Table 2, p. 158; "d_h" is Cohen's (1988) "d" converted to Hedges's "d_h" - see Eq. (2), p. 149]. This is to be compared with:

(1) Cohen's (1988, p. 24) rule of thumb, based on typical results in social science research, that d = 0.2, 0.5, 0.8 implies respectively "small," "medium," and "large" effects. [However, 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."]

(2) d = +2.43 obtained by comparing average pre/post test average normalized gains <g>'s [Hake (2002a)] for the survey by Hake (1998a,b). [Eight reasons for this unusually large d have been given in Hake (2008a).]

As I see it, the challenge to Distance Education is to demonstrate considerable *improvement* over traditional classroom instruction, rather than the essentially insignificant mean effect sizes observed by Lou et al. for OADE - another "No Significant Difference Phenomenon" - i.e., ineffectiveness comparable to traditional classroom education.

4. "We generally measure learning in distance education as the level of satisfaction of both students and employers with the outcomes of the educational experience, and internally, have extensive measurement of course, program, and institutional outcomes to ensure program quality. Maybe I'm naive, but doesn't everyone?"

In my view, measurement of the levels of satisfaction of students, as e.g., by student evaluations of teaching (SET's), is not a valid way to gauge the cognitive (as opposed to the affective) impact of courses, as discussed in e.g., "Re: Problems with Student Evaluations: Is Assessment the Remedy?" [Hake (2002b)]. And whether or not employer satisfaction can serve as a measure of student learning is, I think, problematic.

Overmyer's Baker College <http://www.baker.edu/>, may have "extensive measurements of course, program, and institutional outcomes to ensure program quality" but, I doubt that program quality is generally *ensured* by "extensive measurements" in higher education.

If it were, how would one account for the criticisms of higher education by some of its leaders - reviewed in Section IV, "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?" in "Can Scientific Research Enhance the Art of Teaching?" [Hake (2007c)] - such as:

a. Derek Bok, former and now interim president of Harvard University in *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More* [Bok (2005)];

b. James Duderstadt, President Emeritus and University Professor of Science and Engineering at the University of Michigan, in *A University for the 21st Century* [Duderstadt (2000)].

Summary and Implications for SoTL

This essay originated in discussion-list exchanges initiated by a provocatively titled post "Distance Learning: Is There Any?" [Hake (2007a)], which through responses from Britt Watwood and Scott Overmyer developed into a more fundamental question of interest to SoTL: "Can Distance and Classroom Learning Be Increased?" My answer is "Yes" judging from the fact that physics education researchers (PER's) have discovered that for introductory physics courses covering Newtonian mechanics "interactive engagement" pedagogy can result in normalized learning gains in conceptual understanding that are about two-standard deviations greater than those measured for courses employing traditional passive-student lectures, recipe labs, and algorithmic-problem exams. I see no reason to believe that similar marked enhancement of higher-order student learning in disciplines other than physics could not occur for *both* classroom and distance-education courses that cover conceptually difficult material, especially if assisted by the SoTL movement [heretofore virtually oblivious of PER (see, e.g., Hake, 2007g)].

The lessons of the physics education reform effort [Hake (2002a, 2007f)] suggest that

undergraduate learning gains might be substantively increased by SoTL researchers who develop:

> (a) interactive engagement (IE) methods that benefit from cognitive science's progress in understanding human learning [Bransford et al. (2000), Donovan & Bransford (2005)]. For a discussion see Section 5, "Why Are Interactive Engagement Courses More Effective Than Traditional Passive-Student Courses?" in Hake (2007b);
>
> (b) valid and consistently reliable *high-resolution* multiple-choice tests (MCT's) that assess the need for and results of the above IE methods by *directly* measuring students' course-induced higher-order learning [Wilson & Bertenthal (2005)], in lieu of the more commonly employed *low-resolution* gauges of student learning, such as student evaluations of teaching, student self-assessments, and teacher-made tests and course grades.

Although "b" requires arduous qualitative and quantitative research of the type undertaken by Halloun & Hestenes (1998a,b), such MCT's have the advantage that they may be given to thousands of students in hundreds of courses under varying conditions such that meta-analyses can be performed, thus establishing general causal relationships in a convincing manner [Shadish et al. (2002), Shavelson & Towne (2002), Schneider et al. (2007)].

Unfortunately, aside from the disciplines of astronomy, biology, chemistry, engineering, geoscience, and math there has been little effort in academia to emulate the educational reform effort in physics. I hope that this essay will help to rectify this neglect.

Acknowledgements
I thank (a) Britt Watwood of Virginia Commonwealth University and Scott Overmyer of Baker College for their discussion-list responses that prompted this article; (b) IJ-SoTL editor Alan Altany for valuable suggestions; (c) the NSF for partial support from Grant DUE/MDR-9253965; and (d) Charles Partridge for calling my attention to the meta-analyses of Bernard et al. (2004), Sitzmann et al. (2006), and Lou et al. (2006)

References

Tiny URL's courtesy <http://tinyurl.com/create.php>.
All URL's were accessed on 4 January 2008.

Allen, I.E. & J. Seaman. 2007. *Online Nation: Five Years of Growth in Online Learning*, online at <http://www.sloan-c.org/publications/survey/pdf/online_nation.pdf> (668 kB).

Allen, I.E., J. Seaman, & R. Garret. 2007. *Blending In: The Extent and Promise of Blended Education in the United States*, online at <http://www.sloan-c.org/publications/survey/pdf/Blending_In.pdf> (440 KB).

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

Bernard, R.M., P.C. Abrami, Y. Lou, E. Borokhovski, A. Wade, L. Wozney, P.A. Wallet, M. Fiset, & B.Huang. 2004. How Does Distance Education Compare With Classroom Instruction? A Meta-Analysis of the Empirical Literature, *Review of Educational Research 74*: 379-439; online at <http://rer.sagepub.com/content/vol74/issue3/> (this URL may be erratic): free to subscribers, $25 to non-subscribers, free abstract at the same URL. According to Lou et al. (2006), Bernard et al. "synthesized comparative DE studies on student achievement, attitudes, and retention at all levels of education, K-12, undergraduate, graduate, adult, military, and corporate training, from 1985 to 2002."

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 at <http://edr.sagepub.com/content/vol13/issue6/>; free to subscribers, $25 to nonsubscribers. Bloom wrote:

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

Bok, D. 2005. *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More.* Princeton University Press - information including the preface and Chapter 1 is online at <http://press.princeton.edu/titles/8125.html>.

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

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

Buchanan, E. 2001. The scholarship of teaching and learning in undergraduate distance education, in *Academic Exchange Quarterly 5*(2), 138-143; online at <http://tinyurl.com/2w9xvf>.

Ciavarelli, A. 2004. Assessing the Quality of Online Instruction: Integrating Instructional Quality and Web Usability Assessments, in *Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators*, ed. by J.W. Wall and G.R. Walz. Pro-Ed. Austin, TX. Publisher's information at <http://tinyurl.com/ys8jyg>. Online as a 364 kB pdf at <http://tinyurl.com/yvnhyx>.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaum, 2nd ed. Amazon.com information at <http://tinyurl.com/2ts86w>, note the "Search Inside" feature.

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

Donovan, S.M. & J.D. Bransford, eds. 2005. *How Students Learn History, Mathematics, and Science in the Classroom*. Nat. Acad. Press; online at

<http://books.nap.edu/catalog/10126.html>.

Driscoll, M. 2001. Building Better E-Assessments: Online testing offers untapped potential for instructional designers building e-courses, online at <http://www.learningcircuits.org/2001/jun2001/driscoll.html>. See also Driscoll (2002). For information on "Learning Circuits" see <http://www.learningcircuits.org/mission.html>. "Learning Circuits" is sponsored by *American Society of Training and Development* <http://www.astd.org/astd>.

Driscoll, M. 2002. *Web-Based Training: Designing e-Learning Experiences*. Pfeiffer, 2nd edition. Amazon.com information at <http://tinyurl.com/27zhdg>. Note the "Search Inside" feature.

Duderstadt, J.J. 2000. *A University for the 21st Century*. Univ. of Michigan Press; for a description see <http://tinyurl.com/9lhpl>.

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 at <http://math.unipa.it/~grim/21_project/21_charlotte_EpsteinPaperEdit.pdf> (48 kB).

FLAG. 2008. 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] for 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>. Some conceptual diagnostic tests are referenced for Physics & Astronomy, Chemistry, and Math at <http://www.flaguide.org/tools/tools_discipline.php>. For more complete references to conceptual diagnostic tests in science and math see Hake (2004; 2007d,e). Michael Zeilik discusses "Conceptual Diagnostic Tests" at <http://www.flaguide.org/cat/diagnostic/diagnostic7.php>. For physics diagnostic tests see also NCSU (2008).

Garner, C., W. Pepicello, & C. Swenson. 2005. Faculty Scholarship in a Nontraditional University: The University of Phoenix, in O'Meara & Rice (2005).

Hake, R.R. 1998a. Interactive-engagement vs traditional methods: A six thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys*. *66*(1): 64-74; online at <http://www.physics.indiana.edu/~sdi/ajpv3i.pdf> (84 kB).

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

Hake, R.R. 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. For an update see Hake (2007g).

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

Hake, R.R. 2004. Re: Measuring Content Knowledge, POD posts of 14 &15 Mar 2004, online at <http://listserv.nd.edu/cgi-bin/wa?A2=ind0403&L=pod&P=R13279&I=-3> and <http://listserv.nd.edu/cgi-bin/wa?A2=ind0403&L=pod&P=R13963&I=-3>. A more recent important reference, not included in the above is Epstein (2007).

Hake, R. R. 2005. The Physics Education Reform Effort: A Possible Model for Higher Education, online at <http://www.physics.indiana.edu/~hake/NTLF42.pdf> (100 kB). This is a slightly edited version of an article that was (a) published in the *National Teaching and Learning Forum 15(*1), December 2005, online to subscribers at <http://www.ntlf.com/FTPSite/issues/v15n1/physics.htm>, and (b) disseminated by the *Tomorrow's Professor* list <http://ctl.stanford.edu/Tomprof/postings.html> as Msg. 698 on 14 Feb 2006.

Hake, R.R. 2006. Possible Palliatives for the Paralyzing Pre/Post Paranoia that Plagues Some PEP's, *Journal of MultiDisciplinary Evaluation 6,* November, online at <http://survey.ate.wmich.edu/jmde/index.php/jmde_1/issue/view/22>.

Hake, R.R. 2007a. Distance Learning: Is There Any? online at <http://listserv.nd.edu/cgi-bin/wa?A2=ind0708&L=pod&O=D&P=10532>. Post of 9 August to POD and other discussion lists.

Hake, R.R. 2007b. Should We Measure Change? Yes! online at <http://www.physics.indiana.edu/~hake/MeasChangeS.pdf> (2.5 MB), and as ref. 43 at <http://www.physics.indiana.edu/~hake>. To appear as a chapter in *Evaluation of Teaching and Student Learning in Higher Education*, a Monograph of the American Evaluation Association <http://www.eval.org/>. For a severely truncated version of this article see Hake (2006).

Hake, R.R. 2007c. Can Scientific Research Enhance the Art of Teaching? Greensboro AAPT meeting, invited talk; online at <http://www.physics.indiana.edu/~hake/Sci&Art3.pdf> (1.2 MB).

Hake, R.R. 2007d. Pre/post Tests For College Chemistry, online at <http://listserv.nd.edu/cgi-bin/wa?A2=ind0709&L=pod&O=A&P=9031>. Post of 9 Sep 2007 18:59:09-0700 to POD and other discussion lists.

Hake, R.R. 2007e. Pre/post Tests For College Biology, online at <http://listserv.nd.edu/cgi-bin/wa?A2=ind0709&L=pod&O=A&P=8765>. Post of 8 Sep 2007 22:01:17-0700 to POD and other discussion lists.

Hake, R.R. 2007f. Six Lessons From the Physics Education Reform Effort, *Latin American Journal of Physics Education* 1(1), September; online at <http://journal.lapen.org.mx/sep07/HAKE%20Final.pdf> (124 kB).

Hake, R.R. 2007g. Re: Examples of discipline-specific SoTL projects, online at <http://listserv.nd.edu/cgi-bin/wa?A2=ind0710&L=pod&P=R25817&I=-3>, <http://listserv.nd.edu/cgi-bin/wa?A2=ind0710&L=pod&P=R26685&I=-3>, and <http://listserv.nd.edu/cgi-bin/wa?A2=ind0710&L=pod&P=R28377&I=-3. Posts of 23, 24. & 26 Oct 2007 to POD and other discussion lists.

Hake, R.R. 2008a. Design-Based Research in Physics Education Research: A Review, in A.E. Kelly, R.A. Lesh, & J.Y. Baek, eds. (in press), *Handbook of Innovative Design Research in Science, Technology, Engineering, & Mathematics (STEM) Education*. Taylor and Francis, online at <http://www.physics.indiana.edu/~hake/DBR-Physics3.pdf> (1.1 MB).

Hake, R.R. 2008b. Demonstrated Value of Formative Pre/post Testing, online at <http://tinyurl.com/ysm6se>. Post of 7 Jan 2008 to AERA-J and other lists.

Halloun, I. & D. Hestenes. 1985a. The initial knowledge state of college physics students. *Am. J. Phys. 53*(11): 1043-1055; online at <http://modeling.asu.edu/R&E/Research.html>. The print version contains the *Mechanics Diagnostic* test, precursor to the *Force Concept Inventory* [Hestenes et al. (1992)].

Halloun, I. & D. Hestenes. 1985b. Common sense concepts about motion, *Am. J. Phys*. *53*(11): 1056-1065; online at <http://modeling.asu.edu/R&E/Research.html>.

Harley, D., J. Henke, & M. Maher. 2004. Rethinking space and time: The role of Internet technology in a large lecture course, *Innovate 1*(1); online <http://cshe.berkeley.edu/publications/docs/harley_rethinking-space.pdf> (300 kB), and (for those who sign in) at <http://www.innovateonline.info/index.php?view=article&id=3>. Unfortunately, the authors failed to compare *student learning* in online vs classroom lecture modes as might have been accomplished with formative diagnostic chemistry concept tests.

Hestenes, D., M. Wells, & G. Swackhamer. 1992. Force Concept Inventory, *The Physics Teacher 30*(3): 141-158, March; online (except for the test itself) at <http://modeling.asu.edu/R&E/Research.html>. The 1995 revision by Halloun, Hake, Mosca, & Hestenes is online (password protected) at the same URL, and is currently available in 15 languages: Chinese, Czech, English, Finnish, German, Greek, Italian, Malaysian, Persian, Portuguese, Russian, Slovak, Spanish, Swedish, & Turkish. A French version should soon be available**.**

Hostetter, C. & M. Busch. 2006. Measuring Up Online: The Relationship between Social Presence and Student Learning Satisfaction, *Journal of Scholarship of Teaching and Learning, 6*(2), 1-2; online at <http://www.iupui.edu/~josotl/VOL_6/NO_2/V6N2HostetterFinal.pdf> (104 kB).

Lou, Y., R.M. Bernard, & P.C. Abrami. 2006. Media and pedagogy in undergraduate distance education: A theory-based meta-analysis of empirical literature. *Educational Technology Research and Development 54*: 141-176; freely online to subscribers at <http://www.springerlink.com/content/872241162022v0j1/fulltext.pdf> (224 kB); $32 to non subscribers. ETR&D is a bi-monthly publication of the *Association for Educational Communications & Technology* <http://www.aect.org/default.asp>.

MacIsaac, D.L. 2008. Reformed Teaching Observation Protocol (RTOP) website at <http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/index.htm>. Includes the instrument itself, sample videos, and relevant references.

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

Millar, S.B. 2001. Assessment Tools for Distance Education, *Focus* 6.3, March, online at <http://www.uwex.edu/disted/desien/2001/0103/focus.htm>.

Moore, M.G. 1989. Three types of interaction, *The American Journal of Distance Education 3*(2), 1-6; online at <http://www.ajde.com/Contents/vol3_2.htm#editorial>.

Morrison, J.L. editor. 1997. *Technology Tools for Today's Campuses*, online at

<http://horizon.unc.edu/projects/monograph/CD/>.

NCSU. 2008. *Assessment Instrument Information Page*, Physics Education R & D Group, North Carolina State University; online at <http://www.ncsu.edu/per/TestInfo.html>.

NSSE. 2008. *National Survey of Student Engagement*, homepage at <http://nsse.iub.edu/index.cfm>.

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

O'Neil, H.F. ed. 2005. *What Works in Distance Learning: Guidelines.* Information Age. Publisher's information at <http://www.infoagepub.com/products/books/1-59311-261-0.html>. A prior report to the Office of Naval Research with the same title and edited by O'Neil is online at <http://adlcommunity.net/file.php/10/WhatWorksinDL.pdf> (380 kB). Contents: Background and Purpose: Harold F. O'Neil, Eva L. Baker, and William Bewley; Multimedia Strategies: Richard E. Mayer; Instructional Strategies: Richard Clark; Learning Strategies: Myron H. Dembo and Linda Gubler Junge; Assessment Strategies: Eva L. Baker, Zenaida Aguirre-Muñoz, Jia Wang, and David Niemi [formative assessment strategies are discussed, but pre/post testing per se is not mentioned]; Motivation Strategies: Richard Clark; Self-Regulation Strategies: Harold F. O'Neil and Sanhui (Sabrina) Chuang; Management Strategies: Edward Kazlauskas; Transition and Implementation Issues: Harold F. O'Neil.

O'Meara, K. and R.E. Rice. 2005. *Faculty Priorities Reconsidered: Rewarding Multiple Forms of Scholarship*. Jossey Bass. Publishers information at <http://www.josseybass.com/WileyCDA/WileyTitle/productCd-0787979201.html>. Amazon.com information at <http://tinyurl.com/yofefc> - note the "Search Inside" feature.

Overmyer, S. 2007. Re: ifets-discussion Digest - 11 Aug 2007 to 12 Aug 2007 (#2007-120), IFETS post of 12 Aug 2007 07:57:38-0700. Since Overmyer was responding to Hake (2007a), not "Digest #2007-120," the title should have been "Re: Distance Learning: Is There Any?" Unfortunately IFETS evidently runs on a "lite" version of L-Soft software with no automatically compiled archives.

Partridge, C. 2007. Online Learning Evaluation Articles: review. EdResMeth post of 26 Oct 2007 10:50:29-0400. His "The Effectiveness of Distance Education.doc" was sent as an attachment that did not show up in his post, so I have copied an ASCII version of it into the AREA-L archives where it appears as "The Effectiveness of Distance Education vs Classroom Instruction: A Review of the Evidence by Charles Partridge" at <http://tinyurl.com/3y68cr>.

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

PRG. 2007. Primary Research Group <http://www.primaryresearch.com/>, *The Survey of Distance Learning Programs in Higher Education, 2007-08 Edition.* For ordering information for this and other higher education reports see <http://tinyurl.com/23zep8>. For a press release see <http://www.primaryresearch.com/release-200707233.html>.

Rhem, J. 2007. CLASSE - The Missing Link? *National Teaching and Learning Forum 16*(4), May 2007, online to subscribers at <http://www.ntlf.com/FTPSite/issues/v16n4/classe.htm>. If your institution doesn't have a subscription, then it should.

Russell, T.L. 2001. *The No Significant Difference Phenomenon*. IDECC (International Distance Education Certification Center at

<https://www.arello.org/store/default.cfm?prodlistid=16-17> (this URL may be erratic): "This resource summarizes 355 different research studies that support the conclusion that 'no significant difference' exists between the effectiveness of classroom education and distance learning." See also Russell's website <http://www.nosignificantdifference.org/>, with its searchable data base of "No Difference," "Difference," and "Mixed results" comparisons of delivery methods.

Schneider, B. M. Carnoy, J. Kilpatrick, W.H. Schmidt, R.J. Shavelson. 2007. *Estimating Causal Effects: Using Experimental and Observational Designs.* American Educational Research Association, online at <http://www.aera.net/publications/Default.aspx?menu_id=46&id=3360>.

Seymour, E., D. Wiese, A. Hunter, & S. Daffinrud. 2005. Student Assessment of Learning Gains (SALG); online at <http://www.flaguide.org/cat/salg/salg7.php>.

Shadish, W.R., T.D. Cook, & D.T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin - information at <http://tinyurl.com/y3e7vw>. A goldmine of references to social-science research.

Shavelson, R.J. & L. Towne, eds. 2002. *Scientific Research in Education*, National Academy Press; online at <http://www.nap.edu/catalog/10236.html>.

Sitzmann, T., K. Kraiger, D. Stewart, & R. Wisher. 2006. The comparative effectiveness of web-based and classroom instruction: A meta-analysis, *Personnel Psychology 59*: 623-664; online at <http://tinyurl.com/33jh6p>: free to subscribers, $29 to non-subscribers, an abstract is available at the same URL. Judging from articles available at <http://adlcommunity.net/mod/resource/view.php?id=202>, Sitzmann et al. researched undergraduate, graduate, and corporate training programs.

Sloan Consortium. 2005a. *Growing by Degrees* online at <http://www.sloan-c.org/publications/books/pdf/best_of_sloan-c.pdf> (1.4 MB).

Sloan Consortium. 2005b. *A Synthesis Of Sloan-C Effective Practices, August 2005*, online at <http://www.sloan-c.org/publications/books/v9n3_moore.pdf> (244 kB). There's no mention of pre/post testing in Section 10. "How can schools assess learning effectiveness?"

Walters, R.F. & N.E. Reed. 1997. Outcome Analysis of Distance Learning: A Comparison Between Conventional and Independent Study Instruction, online at <http://horizon.unc.edu/projects/monograph/CD/>, scroll down and click on "Walters"; contained in Morrison (1997).

Walters, R.F. 1996. Distance learning: a controlled performance outcome analysis, *IEEE International Conference on Multi Media Engineering Education*, Melbourne, Australia, pp. 11-16, abstract online at <http://tinyurl.com/2zxotd>.

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