Design-Based Research:

A Primer For Physics-Education Researchers *◊

Richard R. Hake

Indiana University (Emeritus), 24245 Hatteras Street, Woodland Hills, CA 91367

Some prominent education researchers now work in Pasteur's interdisciplinary, use-inspired, basic-research quadrant doing what they call "Design-Based Research" (DBR). After quoting descriptions of DBR by a few of its advocates, I discuss the insularity that has hidden DBR's from PER's (Physics Education Researchers) and PER's from DBR's. I then attempt to make the case that: (a) some PER is also DBR; (b) randomized control trials (RCT's) - not generally a part of DBR - are not the "gold standard" of educational research, as hailed by the U.S. Dept. of Education; (c) DBR might develop into a force sufficient to accelerate even the ponderous educational system: (d) the pre/post test movement, generally ignored by the education community, could be a major component of that reforming force; and (e) non-classical analyses of tests heretofore used primarily for pre/post testing might assist the understanding of "transfer."

I. WHAT IS DESIGN-BASED RESEARCH?

The above question is answered at the "Design-Based Research Collective's" [see DBRC (2003)] website < http://www.designbasedresearch.org/dbr.html > as follows:

> Research in educational settings has historically been driven by two broad goals: understanding how people learn, particularly within school settings; and designing ways to better ensure that learning will happen in these settings. Pursuing these goals in parallel poses significant challenges. However, such work can yield significant rewards, as learning settings can be rapidly refined in response to ongoing research. In recent years, a new paradigm has emerged for engaging in

---

*****The reference is: Hake, R.R. 2004. "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 as a 310 kB pdf by clicking on < http://www.physics.indiana.edu/~hake/DBR-AJP-6.pdf >. 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. I welcome comments and suggestions directed to <rrhake@earthlink.net>.

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

> theoretical research in realistic learning settings. Design experimentation is an inter-disciplinary approach that acknowledges the fundamentally applied nature of educational research.
> Within this approach, researchers working in partnership with educators seek to refine theories of learning by designing, studying, and refining rich, theory-based innovations in realistic classroom environments. Design experimentation reflects a range of practices and methodologies that are drawn from a variety of disciplines. However, *the broad array of methods, claims, theoretical stances, and intellectual traditions makes it extremely difficult to articulate exactly what design experimentation is and how it can advance as a coherent field of study*. [My *italics*.]

Nevertheless, Anthony (Emmon) Kelly, guest editor of the *Educational Researcher* theme issue on "The Role of Design in Educational Research" (ER 2003) took a crack at articulating the meaning of "design experimentation." In his lead article "Research as Design," Kelly (2003) wrote:

> Inspired by the seminal work of Ann Brown [(1992), Brown & Campione (1996)] Allan Collins (1992, 1999), Roy Pea (1999), and Jan Hawkins [Hawkins & Collins (unpublished)], a growing number of researchers have begun to adopt the metaphors and methods of the design and engineering fields. This special issue highlights the work of some of these active researchers and provides a number of commentaries on it . . . . [The] emerging research dialect, which is described in this special issue, attempts to support arguments constructed around the results of active innovation and intervention in classrooms. The operative grammar, which draws upon models from design and engineering, is generative and transformative. It is directed primarily at understanding learning and teaching processes when the researcher is active as an educator . . . . Design research in education would fall under Stokes' (1997) use-inspired basic research category. . .[Pasteur's quadrant]. . . In Toulmin's sense, this research is *clinical* (Toulmin, 2001) . . . . ["where the shared task of the experimenter, the designer, and the practitioner is to move the educational system from malfunction to function and not, primarily, the assured contribution to a body of basic knowledge propositions" (Zaritsky et al. , 2003)]. . . . Further, its proponents are willing to attempt to address, simultaneously and iteratively, the scientific processes of discovery, exploration, confirmation, and dissemination. In its goals and in its context of use, this emerging design research methodology attempts to be both scientific and educational.

Mathematicians Hugh Burkhardt and Alan Schoenfeld (2003) share the enthusiasm of the "Design-Based Research Collective" and Kelly for DBR. They write:

> Introduced in 1992, design experiments represent a significant attempt to conduct research in (experimental) practice, and to contribute to both research and practice (Brown, 1992; Collins, 1992; Kelly, 2003; Schoenfeld, in press). Instructional interventions are designed with explicit theoretical grounding. Data gathered before, during, and after the intervention serve purposes of theory testing. At the same time, they point to strengths and weaknesses of the intervention, informing its revision. Iterative cycles result in improvements in theory and in refinements of the intervention. Design experiments represent a much-needed melding of research and practice.

II. THE INSULARITY OF EDUCATIONAL RESEARCH

As indicated by Eamonn Kelly, the DBR movement as represented in ER (2003) was undoubtedly inspired by the 1992 work of Allan Collins and Ann Brown. But were they really the first to perform "design experiments" as asserted above by Burkhardt & Schoenfeld ? The answer depends on whether or not (a) one believes that legitimate education research is *only* performed by education specialists, psychologists, cognitive scientists, and mathematicians – as many in those fields seem to believe, and (b) pre-1992 educational research overlooked by Burkhardt & Schoenfeld falls under the DBR rubric. In the next Sect. III, I attempt to make the case that the physics education research (PER) is legitimate and that some of its studies of the 1970's and 1980's were essentially DBR.

It appears that outside the physics community (and even *inside* some parts of the physics community) the work of PER's is virtually unknown. Several years ago I stumbled upon Allan Collins' (1999) valuable article "The Changing Infrastructure of Education Research." Since PER's have been doing classroom research for about three decades (McDermott & Redish, 1999), I was puzzled by the italicized sentence below in Collins' claim that:

> There has always been a great divide between education research and practice. Most practitioners regard education research as irrelevant to their day-to-day concerns, and so they pay little attention to what researchers recommend. This partly derives from the origins of education research in the field of psychology. The methods employed, based on laboratory studies, have led to a body of findings that has problematic relation to questions of practice. Learning in the real world occurs in complex social situations, and laboratory methods of studying learning so fundamentally alter the conditions of learning that is not clear what to conclude from any such study. *Recently researchers have begun to study teaching and learning in the context of real-world learning environments*. [My *italics*.] Their work is propelled by the desire to bridge the gap between research and practice and by changing the emphases in cognitive and sociocultural research (Greeno et al., 1996). These attempts to address the problems of practice are leading to changes in the infrastructure of education research. . . . .

Then, more recently, I picked up the *Educational Researcher* of January/February 2004 to find cognitive scientist David Olson (2004) essentially maintaining that the search for "what works" in education is folly. In the same issue, psychologist Robert Slavin (2004) wrote:

> One key assumption in Olson's response is that there have been thousands of experiments evaluating educational practices and that they haven't amounted to much. There may be thousands of brief laboratory studies, but experimental studies of replicable treatments that take place over a semester or more are extremely rare . . . [as judged by surveying the *American Educational Research Journal* over the period 2000–2003].

Slavin is probably unaware of the many replicable, non-laboratory, multi-semester, experimental PER studies, if the word "experiment" is taken in mean the same as in traditional hard-core physics research – "an experiment can be thought of as an act of observation designed to yield a particular type of empirical knowledge" (Ziman, 2000, p. 93), or to mean the same as the word "quasi-experiment." The latter term is used by Cook & Campbell (1979) to designate a study that does *not* involve randomized control group trials (RCT's). Likewise, Olsen seems unaware that PER's have been able to shown what works and what does not work for many areas of physics instruction, most notable Newtonian mechanics; see section III below.

The statements of Kelly, Burkhardt & Schoenfeld, Collins, Olson, and Slavin, and - more generally - the educational literature of education specialists, mathematics, DBR [including ER (2003)], cognitive science, and psychology suggest that most practitioners in those fields have little acquaintance with PER. With the exception of Donovan & Pellegrino (2003), even the NRC's expert science-education committees [see e.g., Labov (2003), McCray et al. (2003)] regularly ignore the PER-induced pre/post testing movement (see sections V and VI below). Conversely, PER's could benefit from more extensive knowledge of non-PER work. Such insularity is due in part to the current difficulty of accessing, assessing, and communicating education studies that may be of interest on a particular topic in the [according to Mosteller et al. (2004)] "approximately 1,100 education journals [that] collectively publish more than 20,000 education research articles each year."

Aside from interdisciplinary reviews such as this one, among possible ways to reduce the access, assess, and communication problems are:

(a) establishment of more *free* electronic peer-reviewed electronic journals such as those listed at the AERA's SIG-COR (2004);

(b) support of the National Science Digital Library (NSDL, 2004);

(c) formation and use of a *free* preprint library similar to the Los Alamos (now Cornell) preprint archive at < http:www.arxiv.org >, as suggested by Gene Glass [see Robinson (2004)];

(d) use of "structured abstracts" (Mosteller et al., 2004) for all education-research articles;

(e) promotion of review articles on education research in the interdisciplinary press such as those by Stokstad (2001), Powell (2003), and Handelsman et al. (2004);

(f) participation in the interdisciplinary "Scholarship of Teaching and Learning" (SOTL) movement (Boyer, 1990; Boyer Commission, 1998; Carnegie Academy, 2004)

(g) more effective use of internet discussion lists and cross-posting thereon so as to tunnel through disciplinary barriers (Hake 2000a, 2003b);

(h) bypassing pricey print journals in favor of simply placing unpublished articles on the web, as currently done by publishing-industry critic Gene Glass at < http://glass.ed.asu.edu/gene/fulltext.html > [see Robinson (2004)]. Scriven (2004b) has suggested that such "webtracts" might be informally peer reviewed on web Work-In-Progress "WIP-sites" such as his own [Scriven (2004a)].

III. SOME PHYSICS EDUCATION RESEARCH IS DESIGN-BASED RESEARCH

In their resource letter on physics education research (PER), McDermott & Redish (1999) list about 160 empirical studies, extending over almost three decades, that (a) focus on the learning of physics by students, (b) represent systematic research, and (c) give procedures in sufficient detail that they can be reproduced. In addition some of these studies were:

A. legitimate in the sense of being "scientific" as defined by Shavelson & Towne (2000), since they generally:

1. pose significant questions that can be investigated empirically,
2. link research to relevant theory,
3. use methods that permit direct investigation of the questions,
4. provide a coherent and explicit chain of reasoning,
5. attempt to yield findings that replicate and generalize across studies, and
6. disclose research data and methods to enable and encourage professional scrutiny and critique;

B. both legitimately "scientific" and examples of DBR, especially those carried out by the leading PER groups in the U.S., even though none of those groups, as far as I know, ever characterized its own brand of PER as falling under the DBR banner held aloft by Kelly (2003).

However, paraphrasing Percy Bridgman (1927) "the true meaning of PER is to be found by observing what a group *does* with it, not what a group *says* about it."Consider, for example, PER groups at Arizona State University < http://modeling.asu.edu/ >, Dickinson College < http://physics.dickinson.edu/~abp_web/abp_homepage.html >, Kansas State University < http://web.phys.ksu.edu/ >, the University of Maryland < http://www.physics.umd.edu/rgroups/ripe/perg/ >, and the University of Washington < http://www.phys.washington.edu/groups/peg/ >. Judging from what they *do* with PER (see their websites), and not necessarily what they *say* about it [Hestenes (1992, 1998) of ASU; Laws (1997) of Dickinson; Zollman (1996) of Kansas State; Redish (1994, 1999, 2003, 2004) and Hammer et al. (2004) of Maryland; and McDermott (1991, 2001) of Washington]; I think its fair to say that:

(1) these researchers seem "willing to attempt to address, simultaneously and iteratively, the scientific processes of discovery, exploration, confirmation, and dissemination" with a research methodology that "attempts to be both scientific and educational"; and

(2) their research: (a) involves active innovation and intervention in classrooms, (b) draws upon models from design and engineering, (c) is generative and transformative, (d) is directed primarily at understanding learning and teaching processes when the researcher is active as an educator, (e) could be placed in Stokes's use-inspired basic research category, and (f) attempts to move the educational system from malfunction to function.

I submit, therefore, *that some PER qualifies as design-based research as characterized by Kelly*.

Further support for this assertion can be found by scanning (a) the Millikan and Oersted award addresses of some of the PER leaders other than those indicated above [e.g., Arons (1973); Karplus (1981); Fuller (1993); Reif (1995); Van Heuvelen (2001); and Goldberg (2003)], (b) a listing of PER programs in the U.S. [ISPEG (2004)], (c) a listing of PER papers on the web [UMPERG (2004)], and (d) consideration of Lesson #5 of the physics education reform effort [Hake (2002a)] :

> *The development of effective educational methods within each discipline requires a redesign process of continuous long-term classroom use, feedback, assessment, research analysis, and revision*. Wilson and Daviss (1994) suggest that such a "redesign process," used so successfully to advance technology in aviation, railroads, automobiles, and computers can be adapted to K-12 education reform through "System Redesign Schools." Redesign processes in the reform of introductory undergraduate physics education have been undertaken and described by McDermott (1991, 2001) and by Hake (1992, 2004a).

Thus the role of design was recognized by physics Nobelist Kenneth Wilson and education journalist Bennett Daviss in *Redesigning Education,* but their exemplary DBR-like program was apparently never implemented, presumably because of a lack of funding. Wilson & Barsky (1998) later wrote (my *italics*):

> We see the need for a launch of a research and development initiative in education, paralleling existing national research initiatives related to AIDS or global climate change . . .Today we have to think of education as demanding in multiple dimensions: as a science, *as a design challenge*, and as a performing art while still being an imperative for life in a democracy. Handed down traditions are no longer enough.

The U.S. educational system's monumental inertia – witness the stagnation of K-12 education and the inaction on the potentially fruitful Wilson/Daviss plan - was considered in a volume of *Daedalus* (1998) that contains essays by researchers in education and by historians of more rapidly developing institutions such as power systems, communications, health care, and agriculture; and that set out to answer a challenge posed by Wilson:

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

Aside from Wilson & Barsky's (1998) vision of a new applied research discipline called "change science" that *Daedalus* (1998) might still serve to initiate, that potentially seminal issue of *Daedalus* did not, as far as I know, provoke any operative ideas for overcoming the inertia of the educational system. But more recently Slavin (2002) has argued that randomized control trials (RCT's - see Sect. IV) will lead to successful change in the education system – never mind (at least for science education), the problem of inadequately prepared teachers and the rudimentary state of exploratory research [see e.g., Lipsey & Wilson (1993, Table 1, Sect. 3.5.1], RCT's may not be the gold bullet that will transform K-12 education as maintained by Slavin, but the *Daedalus* (1998) discussions of the relatively rapid development of non-educational systems and the DBR literature [e.g., the articles in ER (2003) and references therein], suggest that *DBR might develop into a force sufficient to accelerate the ponderous educational system*. As an example, I think that the pre/post testing movement [see Sect. V and VI below], stimulated to some extent by DBR in physics education, has the potential to drastically improve undergraduate science instruction and thereby upgrade K-12 science education.

What is the (generally unrecognized) connection of one with the other? Just this: currently, prospective K-12 teachers derive little conceptual understanding from traditional undergraduate introductory science courses and then tend to teach as they were taught, with similar negative results. As emphasized by Goodlad (1990) to deaf ears:

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

IV. RANDOMIZED CONTROL TRIALS AND PRE/POST TESTING

The "Coalition for Evidence-Based Policy," CEBP (2004), under the aegis of the U.S. Department of Education's "Institute for Education Sciences" [headed by Grover Whitehust (2003)] has produced *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide* (IES 2004). The CEBP's board of advisors < http://www.excelgov.org/displayContent.asp?Keyword=prppcAdvisory > 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 education's parapatetic policy analyst Diane Ravitch. Unfortunately, no physical scientists, mathematicians, philosophers, or K-12 teachers are members of the CEBP. The CEBP's *Guide* is addressed to K-12 education, but its recommendations could influence funding for educational research at the postsecondary level – of primary interest to many PER's. According to the *Guide*:

> Well-designed and implemented randomized controlled trials are considered the "gold standard" for evaluating an intervention's effectiveness, in fields such as medicine, welfare and employment policy, and psychology . . . . randomized controlled trials are studies that randomly assign individuals to an intervention group or to a control group, in order to measure the effects of the intervention . . . .There is persuasive evidence that the randomized controlled trial, when properly designed and implemented, is superior to other study designs in measuring an interventions's true effect . . . . "Pre-post" study designs often produce erroneous results . . . . A "pre-post" study examines whether participants in an intervention improve or regress during the course of the intervention, and then attributes any such improvement or regression to the intervention. The problem with this type of study is that, *without reference to a control group,* it cannot answer whether the participants' improvement or decline would have occurred anyway, even without the intervention. This often leads to erroneous conclusions about the effectiveness of the intervention. [My *italics*.]

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 (2000) - 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." CEBP seems to imply that most questions regarding the effectiveness of an intervention can be answered by RCT's. This may be the case *after* a teaching method has been researched and engineered to its full potential and is ready for full scale deployment. But crucial questions in the early stages of an intervention require exploratory research methods that do *not* necessarily involve RCT's and for which RCT's might actually be counterproductive.

A case in point: some RCT enthusiasts might suggest that RCT's would serve to adjudicate the K-8 California science education wars [see e.g., Hake (2004d] between "direct" vs "hands-on" instruction. But since many K-8 teachers are scientifically illiterate (thanks in part to the failure of society to reward teachers commensurate with their vital societal contribution, and the failure of universities to properly educated them), RCT's might well favor "direct instruction" (DI). The reason is that DI requires little conceptual understanding of science on the part of teachers, while "hands-on" lessons guided by scientifically unprepared teachers can be even worse than DI insofar as advancing students' understanding of science is concerned. Thus RTC's could "prove" the superiority of DI and thereby stifle K-8 science education reform in California.

Judging from the modest effect sizes listed Section 3.5.1 in Table 1 of Lipsey & Wilson (1993), K-8 science education is sorely in need of non-RCT *exploratory* research of the caliber of recent PER research at the high-school and undergraduate level. That research is consistent with the recommendations of Shavelson & Towne (2000) - ST. In the words of Eisenhardt & Towne (2003), "[ST] argued for a postpositivist approach to scientifically based research in education, including a range of research designs (experimental, case study, ethnographic, survey) and mixed methods (qualitative and quantitative) *depending on the research questions under investigation*.

Furthermore:

A. The RTC gold standard is considered problematic by a wide array of scholars and organizations, not just by those in schools of education as implied by Thomas Cook (2001, 2002). Taking issue with the RTC gold standard are philosophers Dennis Phillips [Shavelson, Phillips, Towne, & Feuer (2003)] and Michael Scrivin (2004a); 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 (not all of whose members are in schools of education) oppose the RTC gold standard: American Evaluation Association (AEA) < http://www.eval.org/doestatement.htm >, the American Education Research Association (AERA) < http://www.eval.org/doeaera.htm >, and the National Education Association < http://www.eval.org/doe.nearesponse.pdf > (88 kB).

B. Physicists have made progress in *both* traditional (Ziman, 1992) and educational research (McDermott & Redish, 1999), not by following rigid "scientific" research procedures, but by:

1. Asking the right questions. Werner Heisenberg (1999) put it this way: "In the course of coming into contact with empirical material, physicists have gradually learned how to pose a question properly. Now proper questioning often means that one is more than half the way towards solving the problem."

2. "Doing their damndest with their minds, no holds barred" (Bridgman, 1947) in an attempt to build a "community map" (Redish, 1999; Ziman, 2000; Hake, 2002a – "Can Education Research be Scientific Research?").

C. The CEBP's statement that "pre-post study designs often produce erroneous results . . . . the problem . . . . is that, *without reference to a control group*, [they] cannot answer whether the participants' improvement or decline would have occurred anyway," is irrelevant for most of the pre/post studies considered below in Sections V and VI. The reason is that control groups *have* been utilized - they are the introductory courses taught by the traditional method. The matching is due to the fact that (a) within any one institution the test [Interactive Engagement (IE)] and control [Traditional (T)] groups are drawn from the same generic introductory course taken by relatively homogeneous groups of students, and (b) IE-course teachers in all institutions are drawn from the same generic pool of introductory course physics teachers who, judging from uniformly poor average normalized gains <g> [see section VA below] they obtain in teaching traditional (T) courses, do not vary greatly in their ability to enhance student learning.

Then too, the canonical anti-pre/post arguments by the psychometric authorities Lord (1956, 1958) and Cronbach & Furby (1970) that gain scores are unreliable, have been called into question by e.g., Werner Wittmann (1997), former Cronbach student David Rogosa (1995), Rogosa & Willett (1983), Zimmerman & Williams (1982), and Collins and Horn (1991). All this more recent work should (but does not) serve as an antidote for the emotional pre/post paranoia that grips many educational researchers.

V. PRE/POST TESTING IN PHYSICS EDUCATION RESEARCH

The pre/post testing movement in PER was initiated by the landmark work of Ibrahim Halloun and David Hestenes (1985a,b). In "Lessons from the Physics Education Reform Effort" (Hake, 2002a) I wrote (could there be lessons here for other disciplines?):

> For over three decades, physics-education researchers repeatedly showed that *Traditional* (T) introductory physics courses with passive-student lectures, recipe labs, and algorithmic problem exams were of limited value in enhancing conceptual understanding of the subject (McDermottt & Redish, 1999). Unfortunately, this work was largely ignored by the physics and education communities until Halloun & Hestenes devised the *Mechanics Diagnostic* (MD) test of conceptual understanding of Newtonian mechanics. Among the virtues of the MD, and the subsequent *Force Concept Inventory* (FCI) [Hestenes et al. (1992), Halloun et al. (1995)] tests, are: (a) the multiple-choice format facilitates relatively easy administration of the tests to thousands of students, (b) the questions probe for conceptual understanding of basic concepts of Newtonian mechanics in a way that is understandable to the novice who has never taken a physics course (and thus can be given as an introductory-course pre-test), while at the same time rigorous enough for the initiate.

The MD test construction involved laborious *qualitative* analysis of extensive student interviews and the study of prior qualitative and quantitative work on misconceptions by, among others: Viennot (1979), Champaign et al. (1980), Trowbridge & McDermott (1980, 1981), Gunstone & White (1981), Champaign & Klopfer (1982), Clement (1982), Minstrell (1982), McCloskey (1983, 1989), and Maloney (1984). All this led to a "taxonomy of common sense concepts about motion" [see also Hestenes et al. (1992)] and finally construction of a balanced and valid test that has consistently proven to be highly reliable, as judged by relatively high Kuder-Richardson reliability coefficients KR-20 in the 0.8 to 0.9 range [see e.g. Halloun & Hestenes (1985a); Hake (1998b)].

Halloun & Hestenes then used the MD in *quantitative* classroom research involving massive pre- and post-course testing of students in both calculus and non-calculus-based introductory physics courses at Arizona State University. Their conclusions were:

> (1) . . . the student's initial qualitative, common-sense beliefs about motion and . . . (its) . . . causes have a large effect on performance in physics, but conventional instruction induces only a small change in those beliefs.
>
> (2) Considering the wide differences in the teaching styles of the four professors . . . (involved in the study) . . . the basic knowledge gain under conventional instruction is essentially independent of the professor.

These outcomes were consistent with the findings of many researchers in physics education (McDermott & Redish, 1999), which suggested that traditional passive-student introductory physics courses, even those delivered by the most talented and popular instructors, imparted little conceptual understanding of Newtonian mechanics. But the Halloun & Hestenes research went far beyond earlier work because it offered physics teachers and researchers a valid and consistently reliable test that could be employed to gauge the effectiveness of *traditional* mechanics instruction, and then to continually track the merit of the *non-traditional* methods with respect to (a) traditional methods, (b) one another, and (c) various modes of implementation [See Sect. VII below.]. Thus it could (and has) contributed to a steady albeit very slow iterative increase in the effectiveness (as gauged by student learning gains) of introductory mechanics instruction nation wide.

For example, consider the MD/FCI-induced changes in introductory physics courses at pace setters Harvard and MIT. Harvard's Mazur (1997, p. 4) wrote:

> When reading this. . . . .[Halloun & Hestenes (1985a,b; 1987); Hestenes (1987)]. . . my first reaction was " Not my students. . .!" Intrigued, I decided to test my own students' conceptual understanding, as well as that of physics majors at Harvard. . . . . the results of the test came as a shock: The students faired hardly better on the Halloun and Hestenes test [1985a] than on their midterm exam. Yet the Halloun and Hestenes test is *simple*, whereas the material covered by the examination (rotational dynamics, moments of inertia) if of far greater difficulty, or so I thought.

At Mazur's "Overview of Test Data" < http://galileo.harvard.edu/galileo/lgm/pi/testdata.html >, note:

(a) The abrupt average normalized gain <g> increase from 0.25 in 1990 to 0.49 in 1991 when he replaced his passive-student lectures (that netted very positive student evaluations – many administrators erroneously regard them as valid measures of student learning!) with the interactive engagement "Peer Instruction."

(b) The gradual increase in <g> from 0.49 in 1991 to 0.74 in 1997 as various improvements [Crouch & Mazur (2001)] were made in the implementation of "Peer Instruction."

MIT's John Belcher (2003), describing MIT's transition from a traditional to an interactive-engagement type introductory physics class wrote:

> What is the motivation for this transition to such a different mode for teaching introductory physics? First, the traditional lecture/recitation format for teaching 8.01 and 8.02 has had a 40-50% attendance rate, even with spectacularly good lecturers (e.g., Professor Walter Lewin), and a 10% or higher failure rate. Second, there has been a range of educational innovations at universities other than MIT over the last few decades that demostrate that any pedagogy using "interactive engagement" methods results in higher learning gains as compared to the traditional lecture format (e.g., see Halloun and Hestenes 1985a, Hake 1998a, Crouch and Mazur 2001), usually accompanied by lower failure rates. Finally, the mainline introductory physics courses at MIT do not have a laboratory component.

The Harvard and MIT results are consistent with those from hundreds of other introductory physics courses employing either traditional or interactive engagement methods as evidenced by the meta-analysis discussed below.

A. Meta-Analysis of Pre/Post Learning Gains

In his cogent discussion of "Design Research for Sustained Innovation," Carl Bereiter (2002) writes (my *italics*):

> Rather more successful than . . .[attribute-treatment interactions (ATI's) to discover the optimal matching of persons to treatments (Cronbach 1957)] . . . . has been meta-analysis (Glass, McGaw, & Smith, 1981), in which a number of different studies that are judged to involve the same variable are brought together into a statistically powerful test of the effects of the variable. Educational research journals regularly carry meta-analyses on topics ranging from the effects of computer use to the effects of phonemic awareness training. Meta-analysis, however, takes quantitative research an additional step away from design relevance. In combining results from a large number of experiments in the use of educational games, for instance, all the differences among games and in ways of using them are averaged out, leaving nothing to aid the person who would like to design a more effective educational game.

But Bereiter's hypothetical failure of the meta-analysis of the effects of heterogenous computer games does not justify the conclusion that all meta-analyses "take quantitative research an additional step away from design relevance." For example, my own meta-analysis [Hake (1998a,b; 2002a,b)] of pre/post MD/FCI data for introductory Newtonian mechanics instruction, shown graphically in Fig. 1 [in accord with Gene Glass' dictum (Robinson, 2004) that "the result of a meta-analysis should never be an average; it should be a graph"] has proven to be of direct interest to course designers, and this even though the data were not (and could not have been) obtained from a RCT study.

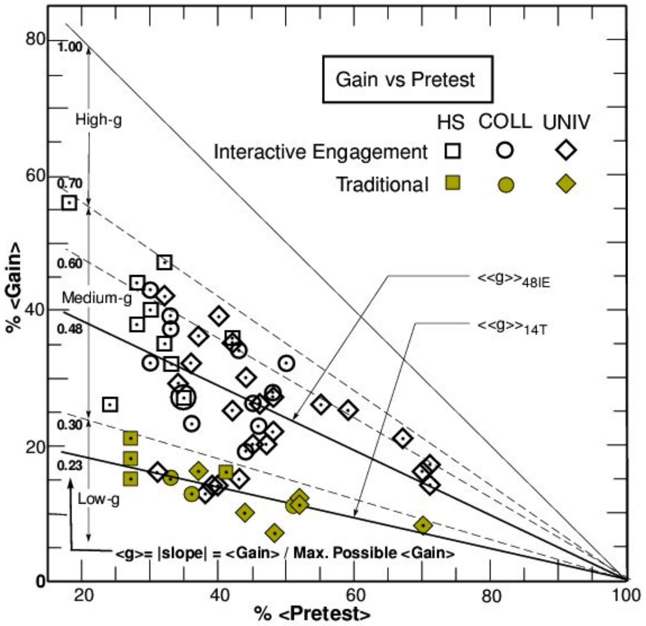

FIGURE 1. From Hake (1998a). %<Gain> vs %<Pretest> score on the conceptual Mechanics Diagnostic (MD) or Force Concept (FCI) tests for 62 courses enrolling a total N = 6542 students: 14 traditional (T) courses (N = 2084) which made little or no use of interactive engagement (IE) methods, and 48 IE courses (N = 4458) which made considerable use of IE methods. Slope lines for the average of the 14 T courses <<g>>$_{14T}$ = 0.23 ± 0.04 (std dev), and 48 IE courses <<g>>$_{48IE}$ = 0.48 ± 0.14 (std dev) are shown. The negative-slope straight lines are lines of constant normalized average <g> = <Gain> / Max. Possible <Gain> = (<%post > – <%pre >) / (100 – <%pre>). Thus, e.g., if a class averaged 40% on the pretest, and 60% on the posttest then the class-average normalized gain <g> = (60% - 40%)/(100% - 40%) = 20%/60% = 0.33.

Regarding the average normalized gain <g>, ever since the work of Hovland et al. (1949) it's been know by pre/post cognoscente (up until about 1998 probably less than 100 people worldwide) that <g> *is a much better indicator of the extent to which a treatment is effective than is either gain or posttest* (Hake, 1998a; Meltzer, 2002b). Justification for the use of <g> for the present data set resides in the fact that the correlation of <g> with <%pre> for the 62 survey courses is a very low +0.02. In contrast, the average posttest score <%post> and the average gain <G> are less suitable for comparing course effectiveness over diverse groups since their correlations with <%pre> are, respectively, +0.55 and –0.49.

Regrettably, the insular psychology-education-psychometric PEP community remains largely oblivious of normalized gain. Paraphrasing Lee Schulman, as quoted by the late Arnold Arons (1986): "it seems that in education, the wheel (more usually the flat tire) must be reinvented every few decades." Extrapolating the historical record, around 2030 yet another investigator will come up with the idea of <g>, and fruitlessly attempt to interest the pre/post paranoiac education community. Then around 2060 . . . .

Fig. 1 serves as an *existence proof* that a two-standard deviation difference between average pre-to-post "normalized gains" <g> on the FCI/MD between "interactive-engagement" (IE) and "traditional" courses *can* be obtained. I calculated a Cohen (1988) effect size "d" of 2.43 (Hake 2002a), much higher than any found by Lipsey & Wilson (1993) in their meta-meta-analysis of psychological, educational, and behavioral treatments. Seven reasons for the "d disparity" between my survey and other social-science research are given in Hake (2002a):

(1) *all* courses covered nearly the same material (here introductory Newtonian mechanics);
(2) the material is conceptually difficult and counterintuitive;
(3) the *same* test (either MD or FCI – see Sec. I) was administered to both IE and T classes;
(4) the tests employed are widely recognized for their validity and consistent reliability, have been carefully designed to measure understanding of the key concepts of the material, and are far superior to the plug-inregurgitation type tests so commonly used as measures of "achievement";
(5) the measurement unit gauges the normalized learning *gain* from start to finish of a course, *not* the "achievement" at the end of a course;
(6) the measurement unit <g> is not significantly correlated with students initial knowledge of the material being tested;
(7) the "treatments" are all patterned after those *published by education researchers in the discipline being tested.*

I should have included in the above list:

(8) possible preferential selection of outstanding IE courses.

In regard to "8" above, I stated in Hake (1998a):

> As in any scientific investigation, bias in the detector [due to the mode of data collection - voluntary contributions that tend to pre-select results which are biased in favor of outstanding courses] can be put to good advantage if appropriate research objectives are established. We do *not* attempt to assess the average effectiveness of introductory mechanics courses. Instead we seek to answer a question of considerable practical interest to physics teachers . . .[and to physics education researchers] . . . .: ***can*** *the classroom use of IE methods increase the effectiveness of introductory mechanics courses well beyond that attained by traditional methods*?"

Normalized gain differences between T and IE courses that are consistent with the work of Hake (1998a,b; 2002a,b) and Fig. 1 have been reported by Redish et al. (1997); Saul (1998); Francis et al. (1998); Redish & Steinberg (1999); Redish (1999); Beichner et al. (1999); Cummings et al. (1999); Novak et al. (1999); Beichner et al. (2000); Bernhard (2000); Crouch & Mazur (2001); Johnson (2001); Meltzer (2002a,b); Meltzer & Manivannan (2002); Savinainen & Scott (2002a,b); Steinberg and Donnelly (2002); Fagan et al. (2002); Van Domelen & Van Heuvelen (2002), and Belcher (2003).

Further *exploratory* (non-RCT !) research is required to increase the effectiveness of IE courses [none that I surveyed (Hake, 1998a,b) achieved an average normalized gain <g> greater than 0.69, only fair on an absolute scale]; ascertain the conditions under which IE courses can be most effective; and test IE courses in a wider variety of environments. In my opinion, new meta-analyses of mechanics-course results accruing (a) over the past decade, and (b) in the future (using a new and more secure test than the FCI or FMCE), are (or will be) badly needed.

PER groups have also gone beyond the original survey in showing, for example, that there may be significant differences in the effectiveness of various IE methods (Saul, 1998; Redish, 1999). There has also been some investigation of contributions to <g> from "hidden variables" such as averages over a class of gender, math proficiency, spatial visualization ability, completion of high-school physics courses, scientific reasoning skills, physics aptitude, personality type, motivation, socio-economic level, ethnicity, IQ, SAT, and GPA. One approach to this question is to investigate the relationship of *individual* student learning gains g with such variables for *single courses* (Hake 2002c, Meltzer 2002a).

Thus in physics education research, just as in traditional physics research, it is possible to perform quantitative experiments that can be reproduced (or refuted) and extended by other investigators, and thus contribute to the construction of a continually more refined and extensive "community map."

VI. PRE/POST TESTING IN DISCIPLINES OTHER THAN PHYSICS

In many cases, some stimulated by pre/post testing in physics education, diagnostic tests of content knowledge [or (better) "operative" knowledge (Arons, 1983)] in various non-physics areas have been constructed by those interested in the development of pre/post tests to measure learning gains in science courses (for the references other than the recent Sundberg (2003), see Hake (2004c):

ASTRONOMY: Adams et al. (2000); Zeilik et al. (1997, 1998, 1999); Zeilik (2002);

ECONOMICS: Paden & Moyer (1969); Saunders (1991); Kennedy & Siegfried (1997); Chizmar & Ostrosky (1998); Allgood and Walstad (1999);

BIOLOGY: Roy (2001, 2003); Anderson et al. (2002); Klymkowsky et al. (2003); Sundberg & Moncada (1994); Sundberg (2002, 2003); Wood (2003);

CHEMISTRY: Milford (1996); Bowen & Bunch (1997); Robinson & Nurrenbern (2001); Gonzalez et al. (2003), Birk et al. (2003); ASU (2004);

COMPUTER SCIENCE [Almstead (2003)]; and

ENGINEERING [Evans & Hestenes (2001); Foundation Coalition (2003); Wage & Buck (2004)].

VII. FROM PRE/POST TESTING TO INVESTIGATION OF TRANSFER

Most of the analysis of the FCI, MD, and other physics diagnostic tests [for a listing see NCSU (2004) and FLAG (2004)] have been done within the framework of "Classical Test Theory" in which only the number of correct answers is considered in the scoring. However more sophisticated analyses are being developed [e.g., by Bao & Redish (2001) for the FCI, and by Thornton (1995) for the Force Motion Concept Inventory (Thornton & Sokoloff, 1998)]. These analyses can indicate incorrect student models that students form during instruction in a single course or in a series of courses successively redesigned in attempts to improve their effectiveness; suggest possible pedagogical improvements; and provide data for the investigation of "transfer," i.e., the transfer of learning or capability from one area to another [Bransford et al. (2000, Chap, 3 "Learning and Transfer"), Barnett & Ceci (2002), Lobato (2003), Hammer (2004), Rebello et al. (2004)].

Research on "transfer" is not easy. Barnett & Ceci’s (2002) abstract reads, in part:

> Despite a century's worth of research, spanning over 5,000 articles, chapters, and books, the claims and counterclaims surrounding the question of whether far transfer occurs are no nearer resolution today than at the turn of the previous century. We argue the reason for this confusion is a failure to specify various dimensions along which transfer can occur, resulting in comparisons of 'apples and oranges'. . . . the past 100 years of research shows that evidence for transfer under some conditions is substantial but critical conditions for many key questions are as yet untested.

Could sophisticated analyses of conceptual tests such as the FCI and FMCE offer a productive path to the understanding of transfer? Jane Lobato (2003) writes (my *italics*):

> *Reflecting upon several cycles of design led to a more nuanced and differentiated view of levels of transfer.* Ellis and Lobato (2002) discussed how a further revision of their design approach to slope . . .(evidently in the mathematical sense). . . resulted in evidence of even more sophisticated levels of transfer. Identifying levels of increasing sophistication in non-normative or incorrect displays of transfer is related to Minstrell's (2001) articulation of facets of students' understanding of physics. In Minstrell's approach. . .[see < http://www.talariainc.com/facet > and < http://tutor.psych.washington.edu/ >] . ., one can identify a particular facet as indicative of more complex and sophisticated understanding than another facet, even when both facets represent incorrect or non-normative reasoning. One can similarly identify levels of actor-oriented transfer, which is powerful for design studies because moving up levels of sophistication may be linked with successive iterations in the design cycle.

It would seem that “reflecting upon several cycles of design” might be augmented by non-classical analyses of tests heretofore used primarily for pre/post testing. In fact, Rebello et al. (2004) have already analyzed student responses to interview questions on FCI problems in an attempt to gain insight on transfer.

References and Footnotes

AERA-SIG-COR. 2004. AERA Special Interest Group, Communication of Research < http://aera-cr.ed.asu.edu/ >, listing of free scholarly peer-reviewed electronic journals in education at < http://aera-cr.ed.asu.edu/links.html >.

Arons, A.B. 1973. “Toward wider public understanding of science,” *Am. J. Phys*. **41**(6): 769-782. (Oersted medal address.) See also Arons (1974, 199 7) and Hake (2004a).

Arons, A.B. 1974. “Toward wider public understanding of science: Addendum,” *Am. J. Phys*. **42**(2): 157-158.

Arons, A.B. 1983. "Achieving Wider Scientific Literacy," *Daedalus*, Spring. Reprinted in Arons (1997). Arons’ twelve “hallmarks of science literacy” are also listed in Hake (2000b, pp. 5-7).

Arons, A.B. 1986. "Conceptual Difficulties in Science" in *Undergraduate Education in Chemistry and Physics, Proceedings of the Chicago Conferences on Liberal Education*," No. 1, M.R. Rice. ed., Univ. of Chicago, p. 23-32.

Arons, A.B. 1997. *Teaching Introductory Physics*. Wiley.

Bao, L., and E. F. Redish. 2001. "Concentration analysis: a quantitative assessment of student states." *Am. J. Phys*. **69**(7): S45-S53; online at < http://www.physics.ohio-state.edu/~lbao/papers.htm >.

Barnett, S.M. & S.J. Ceci. 2002. “When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612-637; an abstract is online at < http://www.people.cornell.edu/pages/sjc9/ >

Beichner, R.J., L. Bernold, E. Burniston, P. Dail, R. Felder, J. Gastineau, M. Gjertsen, and J. Risley. 1999. "Case study of the physics component of an integrated curriculum." *Physics Education Research Supplement to Am. J. Phys*. **67**(7): S16-S24; online at < http://www.ncsu.edu/PER/articles.htm >.

Belcher, J.W. 2003 "Improving Student Understanding with TEAL" [TEAL = Technology Enhanced Active Learning], *The MIT Faculty Newsletter* XVI(2); online as a 176 kB pdf at < http://web.mit.edu/jbelcher/www/fnlEditedLinks.pdf >.

Bernhard, J. 2000. "Does active engagement curricula give long-lived conceptual understanding? Proceedings of GIREP 2000: *Physics Teacher Education Beyond 2000*, Barcelona; online at < http://www.itn.liu.se/~jonbe > / "Publications" / "Physics Education Research," where "/" means "click on."

Bereiter, C. 2002. “Design research for sustained innovation,” in *Cognitive Studies, Bulletin of the Japanese Cognitive Science Society* **9**(3): 321-327; online at < http://ikit.org/people/bereiter.html > as a 28 kB pdf < http://ikit.org/fulltext/2002Design_Research.pdf >.

Boyer, E.L. 1990. *Scholarship reconsidered: priorities for the professoriate*. Carnegie Foundation for the Advancement of Teaching.

Boyer Commission. 1998. *Reinventing undergraduate education: A blueprint for America’s research universities*. The Boyer Commission on Educating Undergraduates in the Research University (Carnegie Foundation for the Advancement of Teaching); online at < http://naples.cc.sunysb.edu/Pres/boyer.nsf/ > / “boyer.pdf” where “/” means “click on” to obtain a 544 kB pdf.

Bransford, J. 2000. "A Tribute to Ann Brown," *Journal of Learning Sciences* **9**(1); online at < http://www.cc.gatech.edu/lst/jls/vol9no1.html#TOC >.

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.

Bridgman, P.W. 1927. *The Logic of Modern Physics*. Available as *Logic of Modern Physics (3 Centuries of Science in America),* 1980. ed. by I. Bernard Cohen, Ayer Publishers.

Bridgman, P.W. 1947. "New Vistas for Intelligence," in E.P. Wigner, ed., *Physical Science and Human Values*. Princeton University Press.

Brown, A. L. 1992. “Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences* **2**: 141-178; abstract online at < http://www.cc.gatech.edu/lst/jls/vol2no2.html#Article1 >. See also Bransford (2000).

Brown, A. L. & J. C. Campione. 1996. "Psychological theory and the design of innovative learning environments: On procedures, principles, and systems," in R. Glaser, ed. *Innovations in learning: New environments for education*, pp. 289-325. Erlbaum.

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/pubs/er/toc/3209.htm >.

Campbell, D.T., W.R. Shadish, & T.D. Cook. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Caramazza, A., M. McCloskey, and B. Green, "Naive beliefs in sophisticated subjects: Misconceptions about trajectories of objects," *Cognition* **9**: 117 (1981).

Carnegie Academy. 2004. Scholarship of Teaching and Learning; online at < http://www.carnegiefoundation.org/CASTL/index.htm >.

CEBP. 2004. Coalition for Evidence-Based Policy; website at < http://www.excelgov.org/displayContent.asp?Keyword=prppcHomePage >.

Champagne, A.B., L.E. Klopfer, and J.H. Anderson. 1980. "Factors influencing the learning of classical mechanics," *Am. J. Phys.* **48**(12): 1074-1079.

Champagne, A.B., & L. E. Klopfer. 1982. "A causal model of students' achievement in a college physics course," *J. Res. Sci. Teach.* **19**: 299.

Clement, J. 1982. "Students' preconceptions in introductory mechanics," *Am. J. Phys.* **50**(1): 66-71.

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/pubs/er/toc/er3201.htm >.

Cole, M. 2003. "Culture, Technology, and Development: In Memory of Jan Hawkins," *Mind, Culture, and Activity* **10**(1): 1–2; online at < http://www.leaonline.com/doi/pdfplus/10.1207/S15327884MCA1001_01 >.

Collins, A. 1992. “Toward a design science of education,” in E. Scanlon & T. O'Shea, eds., *New directions in educational technology*." Springer-Verlag; online at < http://www.edc.org/CCT/ccthome/reports/tr1.html >.

Collins, A. 1999. The changing infrastructure of education research," in E. C. Lagemann & L. S. Shulman, eds., *Issues in education research: Problems and possibilities*. Jossey-Bass; online at < http://learnweb.harvard.edu/tie/resources/collinspaper.htm >.

Collins, L.M. & J.L. Horn. 1991. *Best methods for the analysis of change*. American Psychological Association.

Cook, T.D. and D.T. Campbell. 1979. *Quasi-Experimention: Design & Analysis Issues for Field Settings*." Houghton Mifflin. See also Campbell et al. (2001)

Cook, T.D. 2001. “Sciencephobia: Why Education Researchers Reject Randomized Experiments,” *Education Next*, Fall 2001: 63-68; online at < http://www.educationnext.org/20013/index.html >.

Cook. T.D. 2002. “Randomized Experiments in Educational Policy Research: A Critical Examination of the Reasons the Educational Evaluation Community has Offered for Not Doing Them,” *Educational Evaluation and Policy Analysis*, Fall, **24**(3): 175-199; abstract online at < http://www.aera.net/pubs/eepa/abs/eepa24.htm >.

Cronbach, L.J. 1957. “The two disciplines of scientific psychology,” *American Psychologist*, **12**: 671-684; online at < http://psychclassics.yorku.ca/author.htm >.

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

Crouch, C.H. & E. Mazur. 2001. "Peer Instruction: Ten years of experience and results," *Am. J. Phys*. **69**: 970-977; online at < http://mazur-www.harvard.edu/publications.php >.

Daedalus. 1998. Theme issue **127**(4), "Education yesterday, education tomorrow." The title only is online at < http://daedalus.amacad.org/inprint.html >.

DBRC. 2003. Design-Based Research Collective, “Design-Based Research: An Emerging Paradigm for Educational Inquiry, *Educational Researcher* 32(1), 5-8; online at < http://www.aera.net/pubs/er/toc/er3201.htm >; see also at < http://www.designbasedresearch.org/ >.

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 >.

Eisenhart, M. & L. Towne. 2003. “Contestation and Change in National Policy on ‘Scientifically Based’ Education Research,” *Educational Researcher* **32**(7): 31-38; online as a 176 kB pdf at < http://www.aera.net/pubs/er/toc/er3207.htm >.

ER. 2003. *Educational Researcher*, Jan/Feb, devoted to the theme “The Role of Design in Educational Research,” online at < http://www.aera.net/pubs/er/toc/er3201.htm >.

Fagen, A.P., C.H. Crouch, & E. Mazur. 2002. “Peer Instruction: Results from a Range of Classrooms,” *Phys. Teach*. **40**(4): 206-209; online at < http://mazur-www.harvard.edu/publications.php >. See also Fagen (2003).

Fagen, A.P. 2003. *Assessing and Enhancing the Introductory Science Course in Physics and Biology: Peer Instruction, Classroom Demonstrations, and Genetics Vocabulary*, Ph.D. thesis, Harvard, online at < http://mazur-www.harvard.edu/publications.php > as a 4.4 MB pdf.

FLAG. 2004. "Field-tested Learning Assessment Guide; online at < http://www.flaguide.org/ >: ". . . offers broadly applicable, self-contained modular classroom assessment techniques (CAT's) and discipline-specific tools for STEM [Science, Technology, Engineering, and Mathematics] instructors interested in new approaches to evaluating student learning, attitudes and performance. Each has been developed, tested, and refined in real colleges and universitie classrooms." Assessment tools for physics and astronomy (and other disciplines) are at < http://www.flaguide.org/tools/tools.php >.

Francis, G.E., J.P. Adams, E.J. Noonan. 1998. “Do they stay fixed?” *Phys. Teach*. **36**(8): 488-491.

Fuller, R.G. 1993. Millikan Lecture 1992. “Hypermedia and the knowing of physics: standing upon the shoulders of giants,” *Am. J. Phys.* **61**(4): 300-304.

Fuller, R.G., ed. 2002. *A Love of Discovery: Science Education - The Second Career of Robert Karplus*. Kluwer. This is a valuable resource containing seminal papers of Karplus and his colleagues.

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

Glass, G.V. 2000. “Meta-analysis at 25,” unpublished; online at < http://glass.ed.asu.edu/gene/papers/meta25.html >.

Glass, G. V., B. McGaw, & M.L. Smith 1981. *Meta-analysis in social research*. Sage. See also Glass (2000) and Robinson (2004).

Goldberg, F. 2003. “2003 Millikan award: Research and Development in Physics Education: Focusing on Students’ Thinking,” to appear in *Am. J. Phys*.; a handout is online at < http://www.sci.sdsu.edu/CRMSE/quest/Millikan_handout.pdf > (232 kB).

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

Greeno, J.G., A. Collins, & L.B. Resnick. 1996. “Cognition and Learning,” in D.C. Berliner & R.C. Colfee, eds., *Handbook of Educational Psychology*, pp. 15-16., Macmillan

Gunstone, R.F. & R. White. 1981. “Understanding of gravity,” *Sci. Educ.* **65**, 291-299.

Hake, R.R. 1992. “Socratic pedagogy in the introductory physics lab.” *Phys. Teach*. **30**: 546-552; updated version (4/27/98) online as ref. 23 at < http://www.physics.indiana.edu/~hake >.

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 >.

Hake, R.R. 1998b. “Interactive-engagement methods in introductory mechanics courses,” online as ref. 25 at < http://www.physics.indiana.edu/~hake >. Submitted on 6/19/98 to the Physics Education Research Supplement (PERS) to *Am. J. Phys.*, but rejected on the grounds that its very transparent *Physical Review*-type data tables were “impenetrable”! PER suffers because it has no *Physical-Review*-type archival journal. 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. “What Can We Learn from the Biologists About Research, Development, and Change in Undergraduate Education?” *AAPT Announcer* **29**(4), 99 (1999); available on the web as ref. 7 at < http://www.physics.indiana.edu/~hake >. The potential of the web as a mechanism for promoting interdisciplinary synergy in education reform is emphasized and schematically pictured on page 3.

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 >. Based on an earlier libretto with the leitmotiv: “The road to U.S. science literacy begins with effective university science courses for pre-college teachers.” The opera dramatizes the fact that the failure of universities *throughout the universe* to properly educate pre-college teachers is responsible for our failure to observe any signs of either terrestrial or extraterrestrial intelligence.

Hake, R.R. 2002a. “Lessons from the physics education reform effort,” *Ecology and Society* **5**(2): 28; online at < http://www.ecologyandsociety.org/vol5/iss2/art28/ >. *Ecology and Society* (formerly *Conservation Ecology*) is a *free* "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. 2002; also online as ref. 29 at < http://www.physics.indiana.edu/~hake/ >.

Hake, R.R. 2002c. “Relationship of Individual Student Normalized Learning Gains in Mechanics with Gender, High-School Physics, and Pretest Scores on Mathematics and Spatial Visualization,” submitted to the Physics Education Research Conference; Boise, Idaho; August 2002; online as ref. 22 at < http://www.physics.indiana.edu/~hake >.

Hake, R.R. 2003a. “Spare Me That Passive-Student Lecture,” post of 1 Oct 2003 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. 2003b. “Re: Lively Academic Listservs,” online at
< http://listserv.nd.edu/cgi-bin/wa?A2=ind0312&L=pod&P=R3771 >. Post of 5 Dec 2003 to POD, PhysLrnR, Physhare. Later sent to AERA-D, EvalTalk, and ASSESS, and Math-

Hake, R.R. 2004a. “Socratic Dialogue Inducing (SDI) Labs website
< http://www.physics.indiana.edu/~sdi >. Contains a brief description of SDI Labs with pertinent references and 9 SDI Lab Manuals that may be freely downloadcd.

Hake, R.R. 2004b. “The Arons Advocated Method,” submitted to *Am. J. Phys.* on 24 April 2004; online as reference 31 at
< http://www.physics.indiana.edu/~hake >, or download directly as a 144 kB pdf by clicking on
< http://www.physics.indiana.edu/~hake/AronsAdvMeth-8.pdf >.

Hake, R.R. 2004c. “Re: Measuring Content Knowledge,” online at
< http://listserv.nd.edu/cgi-bin/wa?A2=ind0403&L=pod&O=D&P=16472 >. Post of 14 Mar 2004 to ASSESS, Biopi-L, Chemed-L, EvalTalk, Physhare, Phys-L, PhysLnrR, POD, and STLHE-L.

Hake, R.R. 2004d. "Direct Science Instruction Suffers a Setback in California - Or Does It?", contributed to the 129th National AAPT meeting in Sacramento, CA, 1-5 August 2004; online as reference 33 at < http://www.physics.indiana.edu/~hake >, or download directly as a 420 kB pdf by clicking on < http://www.physics.indiana.edu/~hake/DirInstSetback-041104f.pdf >.

Halloun, I. & D. Hestenes. 1985a. “The initial knowledge state of college physics,” *Am. J. Phys.* **53**:1043-1055; online at < http://modeling.asu.edu/R&E/Research.html >.

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. & D. Hestenes**.** 1987. “Modeling instruction in mechanics.” *Am. J. Phys.* **55**: 455-462.

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.)

Hammer, D., A. Elby, R.E. Scherr, E.F. Redish. 2004. "Resources, framing, and transfer," to appear in J. Mestre . ed. *Transfer of Learning: Research and Perspectives*. Information Age Publishing; online as 184 kB pdf at
< http://www.physics.umd.edu/perg/papers/redish/index.html >.

Handelsman, J., D. Ebert-May, R. Beichner, P. Bruns, A. Chang, R. DeHaan, J. Gentile, S. Lauffer, J. Stewart, S.M. Tilghman, W.B. Wood. 2004. "Scientific Teaching," *Science* **304** (23): 521-522, April; online (free to AAAS members only) at
< http://www.sciencemag.org/content/vol304/issue5670/index.shtml#policyforum >;
Supporting Online Material (SOP) material may be freely downloaded at
< http://www.sciencemag.org/cgi/data/304/5670/521/DC1/1 >. The complete article may be downloaded for free at Handelsman's homepage
< http://www.plantpath.wisc.edu/fac/joh/joh.htm > / "Review Articles" where "/" means "click on," or more directly by simply clicking on
< http://www.plantpath.wisc.edu/fac/joh/scientificteaching.pdf > (100kB).

Hawkins, J. & A. Collins. unpublished. *Design Experiments: Using Technology to Restructure Schools*. Cambridge University Press. [See Cole (2003).]

Heisenberg, W. 1999. *Physics & Philosophy: The Revolution in Modern Sciences*. Prometheus Books (reprint edition - originally published as *Physik und Philosophie*).

Hestenes, D. 1987. "Toward a Modeling Theory of Physics Instruction," *Am. J. Phys.* **55**, 440-454.

Hestenes, D. 1992. "Modeling Games in the Newtonian World," *Am. J. Phys.* **60**(8): 732-748; online at < http://modeling.asu.edu/R&E/Research.html >.

Hestenes, D. 1998. "Guest comment: Who needs physics education research!?" *Am. J. Phys.* **66**(6): 465-467 online at < http://modeling.asu.edu/R&E/Research.html >.

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).

Hestenes, D. 2003. “Oersted Medal Lecture: Reforming the mathematical language of physics,” *Am. J. Phys*. **71** (2): 104-121; online at < http://modelingnts.la.asu.edu/html/overview.html >.

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

IES. 2004. Institute for Education Sciences, *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*, online as a 140 KB pdf at < http://www.ed.gov/about/offices/list/ies/index.html >.

ISPEG. 2004. Iowa State Physics Education Research Group, listing of U.S. physics education groups with web homepages; online at < http://www.physics.iastate.edu/per/links/index.html >.

Johnson, M. 2001. “Facilitating high quality student practice in introductory physics.” *Physics Education Research Supplement to American Journal of Physics* **69**(7): S2-S11.
Karplus, R. 1981. “Autonomy and Input,” *Am. J. Phys.* **49**(9): 811-814. (Oersted Medal Award). Also in Fuller (2002), pp. 90-95.

Kelly, A.E. 2003. “Research as Design,” *Educational Researcher* **32**(1): 3-4; online at < http://www.aera.net/pubs/er/toc/er3201.htm >.

Laws, P. 1997. “Millikan Lecture 1996: Promoting active learning based on physics education research in introductory physics courses.” *Am. J. Phys.* **65**(1): 13-21.

Lobato, J.E., A. Ellis, & R. Munoz. 2002. "Paradox or possibility: The Generalization of Situated Reasoning." Sesson 20.13 , p. 136-137, program of the annual meeting of the American Educational Research Association, New Orleans, LA, April.

Labato, J.E. 2003. “How Design Experiments Can Inform a Rethinking of Transfer and Vice Versa” *Educational Researcher* 32(1): 17-20; online at < http://www.aera.net/pubs/er/toc/er3201.htm >.

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 >.

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 >.

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

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

Maloney, D. P. 1984. "Rule-governed approaches to physics: Newton’s Third Law," *Phys. Educ.* **19**: 37-42.

McCloskey, M., A. Caramazza, and B. Green. 1980. “Curvilinear motion in the absence of external forces,” *Science* **210**: 1139-1141.

McCloskey, M. 1983. "Intuitive physics," *Sci. Am.* **249** (April): 122-130.

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 >.

McDermott, L.C. 1991. “Millikan lecture 1990: What we teach and what is learned: Closing the gap,” *Am. J. Phys*. **59**(4): 301-315.

McDermott, L.C. 2001. “Oersted Medal Lecture 2001: Physics Education Research – The Key to Student Learning,” *Am. J. Phys*. **69**(11): 1127-1137. abstract online at < http://www.phys.washington.edu/groups/peg/pubsa.html#abs1a >.

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

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

Meltzer, D.E. 2002b. “Normalized Learning Gain: A Key Measure of Student Learning,” Addendum to Meltzer (2002a); online as article #7 (addendum) at
< http://www.physics.iastate.edu/per/articles/index.html >.

Meltzer, D.E. & K. Manivannan. 2002. “Transforming the lecture-hall environment: The fully interactive physics lecture,” *Am. J. Phys*. **70**(6): 639-654; also online as article #6 at
< http://www.physics.iastate.edu/per/articles/index.html >.

Minstrell, J. 1982. “Explaining the ‘at rest’ condition of an object,” *Phys. Teach*. **20**: 10-14.

Minstrell, J. 2001. "Facets of students' thinking: Designing to cross the gap from research to standards-based practice." In K. Crowley, C. Schunn, & T. Okada, eds., *Designing for science: Implications from everyday, classroom, and professional settings*, pp. 415-444. Erlbaum.

Mosteller, F. B. Nave, & E.J. Miech. 2004. “Why We Need a Structured Abstract in Education Research,” *Educational Researcher* **33**(1): 29-34; online at
< http://www.aera.net/pubs/er/toc/er3301.htm >.

NCSU. 2002. “Assessment Instrument Information Page,” Physics Education R & D Group, North Carolina State University; online at < http://www.ncsu.edu/per/TestInfo.html >.

NSDL. 2004. National Science Digital Library (NSF), information online at
< http://www.ehr.nsf.gov/due/programs/nsdl/ >. The initial version is at < http://nsdl.org/ >.

Olsen, D.R. 2004. “The Triumph of Hope Over Experience in the Search for 'What Works': A Response to Slavin,” *Educational Researcher* **33**(1): 27-38; online at
< http://www.aera.net/pubs/er/toc/er3301.htm >.

Pea, R. 1999. “New Media Communications Forums for Improving Education Research and Practice,” in E. C. Lagemann & L.S. Shulman (eds.) *Issues In Education Research* (Jossey-Bass); online at < http://www.sri.com/policy/ctl/html/context5.html >.

Powell, K. 2003. “Spare me the lecture,” *Nature* **425**, 18 September, pp. 234-236; online as a 388K pdf at < http://www.nature.com./cgi-taf/DynaPage.taf?file=/nature/journal/v425/n6955/index.html >, scroll down to "News Feature": "US research universities, with their enormous classes, have a poor reputation for teaching science. Experts agree that a shake-up is needed, but which strategies work best? Kendall Powell goes back to school." For some commentary on this report see Hake (2003a).

Rabello, N.S., D.A. Zollman, A.R. Allbaugh, P.V. Englehardt, K.E. Gray, Z. Hrepic, & S.F. Itza-Ortiz. 2004, “Dynamic Transfer: A Perspective From Physics Education Research,” to appear in J. Mestre, ed., *Transfer of Learning: Research and Perspectives*. Information Age Publishing; online as 1.9 MB pdf at < http://web.phys.ksu.edu/ >.

Reif, F. 1995. “Millikan lecture 1994: understanding and teaching important scientific thought processes,” *Am. J. Phys*. **63**(1): 17-32.

Redish, E.F. 1994. “Implications of cognitive studies for teaching physics.” *Am. J. Phys.* **62**(9): 796-803; online at : < http://www.physics.umd.edu/rgroups/ripe/perg/cpt.html >.

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

Redish, E.F. 2003. *Teaching Physics With the Physics Suite*. John Wiley.

Redish, E.F. 2004. “A Theoretical Framework for Physics Education Research: Modeling Student Thinking,” to be published in *The Proceedings of the Enrico Fermi Summer School in Physics, Course CLVI (Italian Physical Society)*; online as 724 kB pdf at < http://www.physics.umd.edu/perg/papers/redish/index.html >.

Redish, E. F., J. M. Saul, and R. N. Steinberg. 1997. “On the effectiveness of active-engagement microcomputer-based laboratories,” *Am. J. Phys.* **65**(1): 45-54; online at < http://www.physics.umd.edu/rgroups/ripe/perg/cpt.html >.

Redish, E.F. & R.N. Steinberg. 1999. “Teaching physics: figuring out what works,”*Phys. Today* **52**(1): 24-30; online at < http://www.physics.umd.edu/rgroups/ripe/perg/cpt.html >.

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

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

Robinson, D.H. 2004. “An Interview with Gene V. Glass,” *Educational Researcher* **33**(3): 26-30; online as a 68 kB pdf at < http://www.aera.net/pubs/er/toc/er3303.htm >.

Saul, J.M. 1998. *Beyond problem solving: evaluating introductory physics courses through the hidden curriculum*. Ph.D. thesis, Univ. of Maryland.

Savinainen A. & P. Scott. 2002a. “The Force Concept Inventory: a tool for monitoring student learning,” *Phys. Educ*. **37**: 45-52; online at < http://kotisivu.mtv3.fi/oma/physics/downloads.html >.

Savinainen A. & P. Scott. 2002b. “Using the Force Concept Inventory to monitor student learning and to plan teaching,” *Phys. Educ*. **37**: 53-58; online at < http://kotisivu.mtv3.fi/oma/physics/downloads.html >.

Schoenfeld, A. H. 2002. “Research methods in (mathematics) education,” in L. English, ed., *Handbook of international research in mathematics education* (pp. 435–488). Erlbaum.

Schoenfeld, A. H. (in press). “Design experiments,” in P.B. Elmore, G. Camilli, & J. Green, eds., *Complementary methods for research in education*. AERA.

Scriven, M. 2004a. “Causation,” unpublished; online at < http://www.arts.auckland.ac.nz/staff/index.cfm?S=STAFF_mscr004 > / “Research” /”Work In Progress” / “Causation”, where "/" means "click on."

Scriven, M. 2004b . “Subspecies of Discussion Lists,” private communication of 26 January to R.R. Hake, G. Glass, and L. Rudner.

Shavelson, R.J. & L. Towne. 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/pubs/er/toc/er3201.htm >.

Slavin, R. 2002. “Evidence-Based Education Policies: Transforming Educational Practice and Research,” *Educational Researcher* **31**(7): 15-21; online at < http://www.aera.net/pubs/er/toc/er3107.htm >.

Slavin, R. 2004. “Education Research Can and Must Address ‘What Works’ Questions,” *Educational Researcher* **33**(1): 27-38; online at < http://www.aera.net/pubs/er/toc/er3301.htm >. See also Slavin (2002).

Steinberg, R. N., & K. Donnelly. 2002. “PER-based reform at a multicultural institution,” *Phys. Teach.* **40**(2): 108-114; online at < http://www.aapt.org/tpt/toc_feb02.html >.

Stokes, D. E. (1997). *Pasteur’s quadrant: Basic science and technological innovation*. Brookings Institution Press.

Stokstad, E. 2001. “Reintroducing the Intro Course,” *Science* **293**: 1608-1610, 31 August 2001: online at < http://www.sciencemag.org/ > : “Physicists are out in front in measuring how well students learn the basics, as science educators incorporate hands-on activities in hopes of making the introductory course a beginning rather than a finale.”

Sundberg, M.D. 2003. “Strategies to Help Students Change Naive Alternative Conceptions about Evolution and Natural Selection,” Reports of the National Center for Science Education **23**(2): 23-26, soon to be online at < http://www.natcenscied.org/newsletter.asp?curiss=35   >.

Thornton, R.K. 1995. “Conceptual Dynamics: Changing Student Views of Force and Motion,” in *Thinking Physics for Teaching*, Bernardini, C, C. Tarsitani, & M. Vicentini, eds. Plenum.

Thornton, R.K. & D.R. Sokoloff. 1998. “Assessing student learning of Newton's Laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula,”*Am. J. Phys.* **66**(4): 338-352. See also Thornton (1995).

Toulmin, S. E. (2001). *Return to reason*. Cambridge, MA: Harvard University Press, especially the chapter “Practical Reason and the Clinical Arts.”

Trowbridge, D.E. & L.C. McDermott. 1980. "Investigation of students' understanding of the concept of velocity in one dimension," *Am. J. Phys*. **48**(12): 1020-1028.

Trowbridge, D.E. & L.C. McDermott. 1981. "Investigation of students' understanding of the concept of acceleration in one dimension," *Am. J. Phys*. **49**(3): 242-253.

UMPERG 2004. University of Maryland Physics Education Research Group: Papers on the Web; online at < http://www.physics.umd.edu/perg/perow.htm >.

USDE. 2002. “Strategic Plan 2002- 2007,” U.S. Dept. of Education; online as a 3.6 MB pdf at < http://www.ed.gov/about/reports/strat/plan2002-07/index.html >. See also Whitehurst (2003).

Van Domelen, D.J. & A. Van Heuvelen. 2002. “The effects of a concept-construction lab course on FCI performance,” *Am. J. Phys*. **70**(7): 779-780.

Van Heuvelen, A. 2001. “Millikan Lecture 1999: The Workplace, Student Minds, and Physics Learning Systems,” *Am. J. Phys.* **69**(11): 1139-1146.

Viennot, L. 1979. “Spontaneous reasoning in elementary mechanics,” *European Journal of Science Education* **1**: 205-221.

Whitehurst, G. 2003. “The Institute of Education Sciences: New Wine, New Bottles, a Presentation by IES Director Grover (Russ) Whitehurst,” online at < http://www.ed.gov/rschstat/research/pubs/ies.html >.

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

Wilson K.G. and B. Daviss, 1994. *Redesigning Education*, Henry Holt; a description is online at < http://www-physics.mps.ohio-state.edu/~kgw/RE.html >.

Wilson, K.G. & C.K. Barsky. 1998. “Applied Research and Development: Support for Continuing Improvement in Education,” *Daedalus* **127**(4): 233-258.

Zaritsky, R., Kelly, A.E., Flowers, W., Rogers, E., & P. Patrick. 2003. “Clinical Design Sciences: A View From Sister Design Efforts,” *Educational Researcher* **32**(1), 33-34; online at < http://www.aera.net/pubs/er/toc/er3201.htm >.

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

Zimmerman, D.W. & R.H. Williams. 1982. “Gain scores in research can be highly reliable,” *Journal of Educational Measurement* **19**: 149-154.

Zollman, D.A. 1996. “Millikan lecture 1995: Do they just sit there? Reflections on helping students learn physics,” *Am. J. Phys.* **64**(2): 114-119; online at < http://www.phys.ksu.edu/perg/papers/millikan.html >.