Improving Educational Research:Toward a More Useful, More Influential, and Better-Funded Enterprise

Hugh Burkhardt and Alan H. Schoenfeld

EDUCATIONAL RESEARCHER 2003 32: 3

DOI: 10.3102/0013189X032009003

Improving Educational Research: Toward a More Useful, More Influential, and Better-Funded Enterprise

by Hugh Burkhardt and Alan H. Schoenfeld

Educational research is not very influential, useful, or well funded. This article explores why and suggests ways that the situation could be improved. Our focus is on the processes that link the development of good ideas and insights, the development of tools and structures for implementation, and the enabling of robust implementation in realistic practice. We suggest that educational research and development should be restructured so as to be more useful to practitioners and to policymakers, allowing the latter to make better-informed, less-speculative decisions that will improve practice more reliably.

Experience in other fields shows that clear practical payoffs lead to massively increased funding for all aspects of research, pure and applied. The approach advocated here would also provide a principled and constructive response to current demands for evidence-based educational practice.

I. The Current State

Roughly a decade has passed since Carl Kaestle (1993) wrote his well-known *Educational Researcher* article, "The Awful Reputation of Educational Research." Despite significant advances in theory and method, it is hard to claim that the situation has improved. Indeed, research in education may be accorded even less respect now than a decade ago. Consider, for example, the following statement from the U.S. Department of Education's Strategic Plan for 2002–2007 (2002):

> Unlike medicine, agriculture and industrial production, the field of education operates largely on the basis of ideology and professional consensus. As such, it is subject to fads and is incapable of the cumulative progress that follows from the application of the scientific method and from the systematic collection and use of objective information in policy making. We will change education to make it an evidence-based field. (p. 48)

Consider the matter of tangible support. Just how important, in dollar terms, is the research enterprise in education? Organizations in applied fields where change is recognized as important (medicine, engineering, electronics) typically spend 5% to 15% of turnover on R&D, with about 20% of R&D expenditures on basic research and 80% on design and systematic development. Here is how education compares. The U.S. House Committee on Science (1998) wrote, "currently, the U.S. spends approximately \$300 billion a year on education and less than \$30 million, 0.01 percent of the overall education budget, on education research . . . This minuscule investment suggests a feeble long-term commitment to improving our educational system" (p. 46).

Educational Researcher, Vol. 32, No. 9, pp. 3–14

We trust that the case has been made. In general, education research does not have much credibility—even among its intended clients, teachers and administrators. When they have problems, they rarely turn to research. Part of the reason, we argue, is a lack of credible models of employing educational research to shape educational practice. Part of the reason is that the traditions of educational research are not themselves strongly aligned with effective models linking research and practice, which we shall refer to in shorthand as R↔P models. We believe that if such models are adopted and shown to be successful, there is the potential for significantly increased funding for educational research.

The essence of our argument is that:

1. Educational research does not often lead directly to practical advances, although it provides useful information, insights, and ideas for improvement. Research could be more useful if its structure and organization were better linked to the practical needs of the education system.
2. The research-based development of tools and processes for use by practitioners, common in other applied fields, is largely missing in education. Such "engineering research" is essential to building strong linkages between research-based insights and improved practice. It will also result in a much higher incidence of robust evidence-based recommendations for practice, helping policymakers to make informed decisions.
3. Realigning the system to make it more educationally powerful will require significant changes in work patterns. There must be much closer coordination of effort between research, design, development, policy, and practice. Equally important, a conscious change in the academic value system will be required to induce the necessary number of educational researchers to develop the relevant skills to engage in such work.

The balance of this section explores issues of research and practice. We begin by considering six current R↔P models in education, and then three traditions of educational research.

Six Current R↔P Models in Education

Model 1: Teachers read research and implement it in their classrooms. Our experience has been consistent with the conventional wisdom: most teachers do not have time to read much research, make sense of it, and employ their understandings productively in the classroom. Doing so is a very challenging task. Given the many detailed studies of each topic and their sometimes conflicting results, how would teachers decide what changes to make? Translating research into practice is a decidedly nontrivial task (see, e.g., Magidson, 2002).

Model 2: Summary guides. Professional organizations (including discipline-based organizations, unions, and the National Research Council) regularly produce research distillations intended for teaching professionals and possibly the public. We lack evidence on the efficacy of this approach. However, summary guides provide much less explicit support than most teachers are used to receiving in published teaching materials. We believe it unlikely that this will be effective in helping them meet the far less familiar challenges that new curriculum often presents.

Model 3: General professional development. We are strongly in favor of sustained, long-term professional development for teachers. It is interesting to note, however, that in Pittsburgh, which had engaged in a sustained program of mathematics professional development, the most significant rise in student test scores came when the city also adopted new text materials consistent with the standards on which professional development was based (Briars, 2001; Briars & Resnick, 2000). Improved materials help professional development influence classroom practice.

Model 4: The policy route. The data available to policymakers often provide insights, identify problems, and suggest changes. However, because of the complex nature of education systems, the diagnosis of causes is inevitably speculative, as are the implementation decisions that follow. When education is a hot political issue, an accelerated time scale (action before the next election) and strong actions are often seen as necessary.

Recent events show all too clearly how policy can outrun the research base. In the name of "standards," for example, many states have implemented high-stakes tests for moving up in grade and for graduating from high school, despite all the evidence that such policies are unlikely to result in student knowledge gains but are likely to increase early dropout rates (Clark, Haney, Madaus, Lynch, & Lynch, 2000). Assessments of student proficiency in various states have ranged from multiple-choice tests of basic skills to portfolio evidence of student work on extended problems and projects. The formula used to identify underperforming schools in the "No Child Left Behind" act is so unwieldy that President Bush described a Michigan elementary school he visited last year as "excelling" just 3 months before it was declared below standard (Dillon, 2003).

We do not suggest that an improved research base would automatically lead to improved educational policy—the relationship between public health policy and medical research makes it clear that public policy is shaped by more than knowledge. But knowledge can make a difference. Until educational research findings are (and are perceived as) much more robust and defensible, policymakers will be free to choose findings to support their pre-selected paths. They may find themselves at the mercy of advocates who can argue for this or that approach in the absence of evidence-based consensus regarding good practice.

Model 5: The long route. There are examples of the productive dialectic between educational research and practice. It is worth examining one, albeit briefly, to see how things worked. This example also shows how contingent the outcomes can be on a series of partly fortuitous events. Starting in the 1970s, research in cognitive science produced a reconceptualization of what it means to be competent in various content domains (Gardner, 1985), among them mathematics (Schoenfeld, 1985). In 1989, the National Council of Teachers of Mathematics (NCTM) produced *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). The *Standards,* as they are known, were written for teachers but took as their grounding the conceptualization of competency that had emerged in the previous decades' research. In the 1990s, the National Science Foundation supported 12 project teams across the grade range K–12 to develop curriculum materials based on the *Standards*. These materials, with aligned assessment and professional development support, have begun to establish a significant presence in classrooms across the nation. A growing research base indicates that, in general, students taught via these curricula do approximately as well on routine skills as students taught traditional curricula, but that they do far better on assessments of conceptual understanding and problem solving (ARC Center, 2003; Senk & Thompson, 2002).

Note that the time scale for substantial R↔P impact in this case was 25 years, and that evidence on the real impact of such curricula is just beginning to accumulate. This is a very long time, partly because there exist no mechanisms at the systemic level for the efficient redesign and implementation of curricula. Equally important, the impact of research on practice was, in this case, contingent on two historically unusual events: (a) the creation of the *Standards* and the presence on the writing team of those who knew and used the research, and (b) the move by the National Science Foundation to support curriculum development. The first was a historical accident. The second was unusual and absolutely necessary. Given the marketplace and conventional forms of text development, no commercial publisher would have borne the costs of the research-based design and development process required to produce the new curricula (Schoenfeld, in press-a).

Model 6: Design experiments. 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-b). 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. Typically, however, they embody only the early ("alpha") stages of the design and refinement process (see Section IIIA below). For example, Cobb, Confrey, diSessa, Lehrer, and Schauble (2003) point to the importance of design experiments in testing and refining "local" theory, an essential component of theory-based design. The Design-Based Research Collaborative (2003, p. 8) views the promise of such work as follows: "We see four areas where design-based research methods provide the most promise: (a) exploring possibilities for creating novel learning and teaching environments, (b) developing theories of learning and instruction that are contextually based; (c) advancing and consolidating design knowledge; and (d) increasing our capacity for educational innovation." Typically, the ideas and materials produced by design experiments have been refined to the point where they appear ready to go to scale. How to do so has not been similarly problematized. Thus, the question we address in this article is: How does one refine ideas and materials so that they are robust across a wide range of contexts of implementation? The answer we propose is an adaptation of the "engineering approach" common to some other applied fields.

The Three Main Research Traditions Within Education

In Section II we examine six essential aspects of the R↔P process and current obstacles to their implementation. Before proceeding, however, it is important to identify three main research traditions within education (and academia more broadly). These research approaches, here called *humanities*, *science*, and *engineering*, have significant entailments for potential R↔P partnerships (Burkhardt, 2001).

The *humanities* approach to research is the oldest tradition in education. It may be described as "original investigation undertaken in order to gain knowledge and understanding; scholarship; the invention and generation of ideas . . . where these lead to new or substantially improved insights" (Higher Education Research Funding Council, 1999, p. 4). There is no requirement that the assertions made be tested empirically. The test of quality is critical appraisal concerning plausibility, internal consistency and fit to prevailing wisdom. The key product of this approach is critical commentary.

Much work in education (including this article!) is of this type. Ideas and analysis based on authors' reflections on their experience are often valuable. However, since so many plausible ideas in education have not worked well in practice, the lack of empirical support is a profound weakness. This has led to a search for "evidence-based education" and the significant growth in the education research community (though not in education as a whole) of the science approach described below.

The *science* approach[1] to research is also focused on the development of better insight; of improved knowledge and understanding of "how the world works," through the analysis of phenomena; and the building of models that explain them. However, this approach imposes in addition a further essential requirement—that assertions be subjected to empirical testing. The key outcomes are again assertions—but now with both arguments in support and responses to key questions that are built on empirical evidence. The common products are research journal papers, books, and conference talks. Such research provides insights, identifies problems, and suggests possibilities. However, it does not itself generate practical solutions, even on a small scale; for that, it needs to be linked to the engineering approach.

The *engineering* approach to research is directly concerned with practical impact—understanding how the world works and helping it "to work better" by designing and systematically developing high-quality solutions to practical problems. It builds on insights from other research, insofar as they are available, but goes beyond them. It can be described as "the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products, and processes, including design and construction" (Higher Education Research Funding Council, 1999, p. 4). It combines imaginative design and empirical testing of the products and processes during development and in evaluation. Key products are tools and/or processes that work well for their intended uses and users, with evidence-based evaluation.

In the educational research community the engineering approach is often undervalued. At major universities only "insight" research in the humanities or science tradition tends to be regarded as true research currency for publication, tenure, and promotion. Yet engineering research has a key role to play in making educational research as a whole more useful. In *Pasteur's Quadrant*, Stokes (1997) argues that better insights come from situating inquiry in arenas of practice where engineering is a major concern. Stokes's motivating example is Pasteur, whose work on solving real world issues contributed fundamentally to theory while addressing pressing problems such as anthrax, cholera, and food spoilage.[2] Analogous arguments have been made regarding the potential for such work in education (National Academy of Education, 1999; Schoenfeld, 1999), and serve as a justification for design experiments. Our point is that the same profitable dialectic between theory and practice can and should occur (with differing emphases on the R&D components) from the initial stages of design all the way through robust implementation on a large scale. We also argue that success will breed success: Once this approach is shown to produce improved materials that work on a large scale, more funding will become available for it. Such has been the history in other applied fields, such as medicine and consumer electronics.

In closing this section we wish to re-emphasize our main point. Although good insight-focused research identifies problems and suggests possibilities for progress, it does not itself generate reliable solutions that can be directly implemented on a large scale. To achieve that, research-based development and robust well-tested models of large-scale change are both essential. Currently, the links between these elements are at best weak and often nonexistent. How they might be strengthened, and the implications for the research community, are major themes of this article.

II. Effective Models of R↔P

We believe that six key elements of R↔P models are common to successful research-based fields of practice such as medicine and the design and engineering of consumer electronics. We list them here in an order that helps to clarify the argument. All six are essential.

1. Robust mechanisms for taking ideas from laboratory scale to widely used practice. Such mechanisms typically involve multiple inputs from established research, the imaginative design of prototypes, refinement on the basis of feedback from systematic development, and marketing mechanisms that rely in part on respected third-party in-depth evaluations. These lab-to-engineering-to-marketing linkages typically involve a strong research-active industry (for example, the drug companies, Bell Labs, Xerox PARC, and IBM).
2. Norms for research methods and reporting that are rigorous and consistent, resulting in a set of insights and/or prototype tools on which designers can rely. The goal, achieved in other fields, is *cumulativity*—a growing core of results, developed through studies that build on previous work, which are accepted by both the research community

and the public as reliable and non-controversial within a well-defined range of circumstances. (Work on the cutting edge is something else, of course, with some uncertainties and controversy in every field of research.)

3. A reasonably stable theoretical base, with a minimum of faddishness and a clear view of the reliable range of each aspect of the theory. Such a theory base allows for a clear focus on important issues and provides sound (though still limited) guidance for the design of improved solutions to important problems.
4. Teams of adequate size to grapple with large tasks, over the relatively long time scales required for sound work of major importance in both research and development.
5. Sustained funding to support the R↔P process on realistic time scales.
6. Individual and group accountability for ideas and products—do they work as claimed, in the range of circumstances claimed?

III. R↔P in Education: Barriers, Necessary Changes, and Levers for Change

In this section we examine the degree to which each of the six key elements of R↔P models is in place in education. For each we address three key questions: What are the current barriers? What changes would lower those barriers? How can such changes be brought about?

A. Robust Mechanisms for Taking Ideas From Laboratory Scale to Widely Used Practice.

In mature R↔P fields, prototypes are designed, tested, refined, and brought to scale through large-scale testing and marketing. Here, we sketch out some details of what the full R↔P engineering process for a component of a curriculum might entail.[3] Analogous approaches could be taken to assessment or professional development or, more problematically, to systemic change. The methodology we describe is in line with those used, with variations, across many fields of engineering and applied science (see, for example, Downton, 1992).

The design process includes the development of a set of goals and standards, a search for design ideas, a benchmarking process to examine in fine detail how well the materials are working, and the early piloting of experimental ideas. It proceeds with the production of alpha and beta versions for field testing and subsequent refinement prior to polishing the materials for marketing on a large scale. Goals, materials, and benchmarks are often refined during this process—particularly in the early exploratory stages. (Later revision is always more expensive, and speculative—the engineering "rule of 10" suggests that at each phase of the development process, revision is 10 times more costly than at the previous phase.)

The first phases of design (not unlike some aspects of standard curriculum development, or design experiments) often include:

- preliminary agreement on the goals for and structure of the curriculum;
- collecting and generating design ideas, within the design group (which may include teachers and researchers) and in consultation with outside experts;
- the production of draft materials, which are piloted in the classroom by lead designers and others in the design group, possibly as part of design experiments.

Revisions are undertaken until the process produces (a) a well-defined set of goals covering, for example, classroom learning activity patterns, student performance, and attitudes; (b) an alpha version of the materials that "works" when taught by team members and is considered to be in good enough shape to be used by others; and (c) an assessment that includes a curriculum-independent standards-based component and formal and informal assessments of curriculum-specific goals.

At this point, the engineering process becomes more structured and elaborate than in the typical materials development process. Alpha testing is carried out in (say) 10 classrooms that have been chosen for at least some variation in school type, student demographics, and teacher skill. Data gathered in this phase include structured observation reports by a team of observers, informal interviews with teachers and students, and the systematic sampling of student work (including performance on the standards-based and curriculum-specific assessments). These data are used for both research and product improvement. For both purposes, the design team constructs an analytic description of each teacher, how the teacher used the materials, and how the materials functioned in the classroom. Note that at this point in time, robust descriptions of teacher characteristics and how those characteristics interact with novel instructional materials and practices would be a significant contribution to the field's understanding of teacher knowledge. More broadly, studies of institutional support structures and how they affect curriculum implementation would be extremely useful. Hence, there is much to contribute to basic research. At the same time, the data contribute to:

- curriculum revisions on the basis of perceived weaknesses, observations of student use, and productive adaptations made by the teachers;
- the development and/or modification of support materials for teachers;
- preliminary descriptions of necessary conditions for successful implementation ("successful" in this case meaning at least as good as the status quo).

The revised materials are then ready for more extensive beta testing. In this round, matched pairs of between 50 and 100 classrooms with varied characteristics are chosen. The treatment (either the current curriculum or the beta version of the new curriculum) is randomly assigned. The standards-based curriculum-independent benchmarking tests are used in all classrooms to assess student performance. The data gathered allow for rigorous descriptions of curricular impact. In addition, periodic observations of instruction and the curriculum-specific assessments are used (although not as intensively as in the alpha testing) for purposes of further refinements to the curricular and support materials. Those materials, having been thoroughly vetted in a wide range of circumstances, are considered ready for widespread distribution—along with performance data that can be used for comparison purposes with other curricula. (In engineering this approach is often called the "waterfall methodology.")

Without elaborating, we note that the conditions of implementation matter a great deal. A curriculum that can "raise the floor" with a certain level of teacher support may also "lower the floor" if certain support structures are not in place. Potential users of a curriculum should know what conditions are necessary for its successful implementation, so they can make sure the conditions are in place (or if they cannot, they should choose another curriculum). It is the development team's job to discover and provide this information in the later stages of development and from use in the field. This, too, is an aspect of standard product development.

We believe that the following barriers contribute to the general absence of these processes of research-based development in education.

Barrier 1: It's (almost) nobody's job to turn insight into impact. Orchestrating the R↔P process for any major idea or product is an arduous and time-consuming job. Doing such work is not part of most academics' job description. Some product refinement does get carried out—mostly in funded development projects, which are often based in universities or in "soft money" development centers such as TERC (Massachusetts) or the Lawrence Hall of Science (California). However, the number of people engaged in such work (order of magnitude 100) is very small compared to the community of researchers (order of magnitude 10,000), and the effort is correspondingly limited. Moreover, the policies of funding agencies typically place so much pressure on the development teams (typically, to deliver a year's curriculum each year) that the in-depth probing that a good engineering research approach demands is rarely possible.

Barrier 2: "You do your thing, I do mine." The academic community in education (when compared to other applied fields) tends to see research as very much an individual or small group enterprise. The typical project scale is usually defined by that of a PhD study, or the work needed to produce a research paper with interesting insights. This fits, of course, with the notion of assigning "credit" to individuals for work that is identifiably theirs.

In contrast, the development of imaginative and robust products and processes that directly help to improve practice requires empirical testing using research methods on a substantial scale, from early pilot or design work right through to studies in the realistic conditions of implementation and its support. This can only be done well by substantial teams.

Barrier 3: A negative incentive system. Institutions vary in the ways they allocate credit to individuals for their contributions to research papers, but a general rule is that the sum of the contributions to a paper adds up to one—the more participants, the less credit one is likely to earn toward promotion and/or tenure. And, of course, the larger the project, the more difficult it is to carve out an academic identity for a line of work—the sine qua non for academics.

The contrast with some parts of industry is dramatic. Henry Pollak was Director of Mathematics and Statistics Research at Bell Laboratories in a period in which this group was highly regarded for the work it produced. Pollak argued that if one person sees 80% of what one *ought* to see in a problem, then two people with different backgrounds might see 96% of what they ought to see. Pollak instituted the following credit policy: All the contributors to a paper agree on who will be listed as co-authors, and all will get full credit for that paper. "I don't recall this ever leading to any arguments," said Pollak, "but productivity went through the roof, and the reputations of the individuals, the group, and the work sure didn't suffer" (H. Pollak, personal communication, March 28, 2003).

At research universities, at least, the situation in education is worse than suggested by the two preceding paragraphs: if a group's activity produces "instructional materials" rather than scholarly papers, there may be no credit at all to be shared! This contrasts strongly with the medical research model, where papers describing successful treatments often have dozens of authors listed, and credit is given for those who supervised clinical trials as well as those who shaped the development of the treatments. Similar policies apply in other "big" research fields, including "pure" fields like elementary particle physics and space science.

Barrier 4: The absence of a research-based industry. Educational publishers are the "big manufacturers" of educational products. Why do they not invest more in systematic research-based development? Simply put, economic forces argue against it. First, the R↔P process described above is far more costly than traditional mechanisms for producing text series. Second, evaluations of curriculum effectiveness almost never involve performance data! Manufacturers of cars or computers face evaluations in *Consumer Reports* and in specialty magazines such as *Motor Trends, Car and Driver, MacUser,* and *PC Magazine*. In contrast, publishers have no significant incentives for serious improvement. Superficial features sell products, and they are the basis for marketing. That will remain true until there are more reliable mechanisms that provide clear evidence regarding the effectiveness of different products.

Barrier 5: The absence of commercial incentives for implementing change. Novel materials are often more expensive to produce than traditional materials. In addition, sales and support costs are much higher. If materials are nonstandard, a sales force will have to be trained to understand them and sell them appropriately. Customers will often require substantial professional development in order to use the materials successfully (and an absence of such support may produce "disaster reports," damaging the publisher's credibility). A new product may compete directly with a publisher's established products. Unless there is reason for a publisher to believe that new materials will become the next "big thing," nonstandard materials are likely to be published as alternative "special products" to the publishers' main lines, and not given the kind of marketing support that other industries give to new products (consider consumer electronics or medicine as examples!).

Given these barriers, the following changes would be productive.

Change 1: More research in Pasteur's Quadrant. Educational research has reached the point, unimaginable a mere 25 years ago, where it is now possible to conduct fundamental research amidst principled attempts to affect practice for the better. Indeed, an argument can be made that much more work of this type must be conducted to obtain valid findings. The classic "do basic research in the laboratory and then apply it" model of applying research to practice is context insensitive, while almost all educational interventions are context sensitive. This being the case, more research *in* practice is essential to understand which contextual factors are critical and which are not. As noted above, design experiments (Kelly, 2003) are primary examples of research that resides in what Stokes (1997) refers to as "Pasteur's quadrant"—the space of studies that offer potentially significant contributions to both theory and practice. We propose expansions of that enterprise to larger scale.

Change 2: More engineering research. Researchers in education can contribute a great deal to the engineering research approach—the imaginative design and systematic research-based development of educational materials and their implementation. Because some of this work is currently less prestigious than insight-based research, making the engineering process a reality will involve changes in value systems affecting Barriers 1, 2, and 3 (above). If credit is not allocated for design, large-scale data gathering, and principled refinement on the basis of data, academic researchers will have no incentives to do such work. Making such changes in the value system may not require as great a leap as it might seem: within the university system medicine, engineering, and "big science" departments offer alternative models of credit allocation for faculty working on large projects.

Change 3: Offering the field robust models of R↔P. Existence proofs matter. The medium-term funding of some such large-scale projects or, better, a number of collaborative institutes to carry out such projects, would be an excellent catalyst for the growth of the engineering enterprise.

Potentially effective levers that would encourage such changes include the following: offering training for researchers in the broader range of skills involved in educational engineering; identifying and rewarding outstanding designers of educational materials and processes; re-balancing the academic value system, in recognition of the importance of engineering research and development; funding to support such work (see section E); and the creation of appropriate organizational structures to foster it (see section D). An emerging literature on issues of implementation fidelity, scalability, and sustainability (see, e.g., Blumenfeld, Fishman, Krajcik, Marx, & Soloway, 2000; Elmore, 2000; Fishman, Honey, Hug, Light, Marx, & Carrigg, 2003; Gomez, Fishman, & Pea, 1998; Honey & McMillan-Culp, 2000; Means & Penuel, 2003) discusses mechanisms for increasingly robust implementation of educational practices and materials.

B. Issues Pertaining to Research

The changes discussed above imply changes in the methodologies, norms, and values of research in education. We need methods that are rigorous enough to provide results on which designers and other researchers can rely. This in turn requires cumulativity through studies that build on previous work. As before, we discuss barriers and levers for change.

Barrier 1: Building on quicksand, or without design specs. Schoenfeld (2002) offers a scheme for categorizing the impact of studies in education. He argues that it might be useful to evaluate the claims made in such studies along (at least) these three dimensions:

(a) *Trustworthiness*: How well substantiated is each claim?
(b) *Generality*: To how wide a set of circumstances is the statement claimed to apply?
(c) *Importance*: What contribution does this paper make to theory, methods, or practice?

Generally speaking, papers (and even bodies of literature) tend to score well on one of these dimensions and poorly on the others. Many individual studies, for example, score reasonably well on *trustworthiness*, presenting specific findings that are reasonably well warranted by the evidence given. Often, it is claimed that those findings are representative of a broad class of phenomena, which implies their *importance* and *generality*. But, a rigorous evidentiary warrant for these claims is rarely included with the research reported. Thus, importance and generality remain hypothetical.

The point is that most such studies, while rigorous enough to be published and while providing some form of insight, tend not to provide adequate information to allow for replication and extension; they tend not to make the substantive (rather than suggested) case for generality. In pragmatic terms, papers that often claim to have instructional implications do not offer enough for a designer to use as more than a possible source of ideas, whose validity must then be established from scratch in the domain of use.

Let us examine the implications for cumulativity. Often a topic is claimed to have theoretical importance or generality. For such a topic one can find large numbers of studies, many of which seem to be trustworthy on an individual level. Yet it is often the case that, collectively, the studies do not cohere. As an example, consider the case of "advance organizers," a construct introduced by Ausubel (1968). Hundreds of papers have been published on the topic. Among them are many rigorously conducted individual studies, in which the effects of a particular advance organizer were examined. Yet at the metalevel, no robust conclusions have been drawn from this literature. Why? The very notion of advance organizer was heuristic and descriptions thereof were underspecified. Individual experimenters made up their own advance organizers according to their best understandings of the concept. The studies, then, did not produce findings about advance organizers. They produced results regarding individual researchers' attempts to study what they thought advance organizers might be. Cumulatively speaking, there was no trustworthiness, and thus no generality.

Advance organizers are a case in point. Absent self-conscious attempts on the part of the field to use standard definitions, methods, and benchmarks (a hallmark of engineering research), the research whole is less than the sum of the parts. This generalizes to Barrier 2.

Barrier 2: Most research investigates treatments, but claims to infer general principles. To establish principles, one must check that the observed phenomena persist across a well-defined range of other potentially important variables (such as treatment, designer, topic, teacher, student). A substantial body of evidence of this type would give designers a firm basis for using the principle revealed by the research. Obtaining that body of evidence requires large studies, involving both extension and replication—an unfashionable research activity in education.

Barrier 3: Treatments that are both robust and well researched are rare. Why are the treatments used in careful research studies, teaching experiments, for example, so rarely published for use by practitioners? Researchers naturally prefer to custom design the treatments they investigate to fit the purpose of their inquiry. But they rarely take a treatment through the systematic development process that would ensure its robust effectiveness in use by others, and thus enhance its potential large-scale impact. Indeed, the design and development of the treatment often receives less attention than do the data gathering and analysis intended to yield insights about it.

Why? Academic appointment and advancement criteria tend to produce supervisors who are much more proficient in the analytic skills of insight research than in those of engineering design and development. There may be a belief that design details are unimportant, so that it is the principles behind the treatment's design that are being researched (but note Barrier 2 above). Time pressure will often be a factor.

Barrier 4: Doing your own thing (again). The individualistic value system underlying academic credit allocation tends to limit the scope of investigations, not only in scale but also in how far they can combine generality and trustworthiness. Tacitly, there are pressures against standardization of treatments or probe-instruments, with a premium in prestige and satisfaction for inventing your own rather than using, perhaps with fine-tuning, treatments and research tools that already exist and are nearly as good.

Apart from limiting the scope and reliability of research, this hinders comparability. Imagine how slowly science would progress if everyone invented his or her own methods, concepts, and units! It is said that there were 250,000 different units of length and weight in use in France before the revolution of 1789. This may have been unproblematic as long as no town needed to communicate intelligibly with another—but standardization was necessary for progress.

Barrier 5: Intra-communal disputation. Progress in education depends on a body of reliable research-based information. Such information must include not only useful insights but also their range of demonstrated validity. In contrast to fields where the research community is influential, the educational research community has produced no common core of generally accepted results; nor does it seem to have effective mechanisms for doing so. As a result, education has no collective voice to counterbalance less expert commentators. Indeed, many "common sense" results that are widely accepted by the public (e.g., "retention works") fly in the face of modern research.

Paradigm wars that seek to establish a single theoretical perspective or methodology as superior are only one example of this unproductive disputation (see, e.g., Anderson, Reder, & Simon, 1996, for a relatively recent example; for a historical example, recall the behaviorists' attempts to outlaw all "mentalism"). Why should a field be taken seriously by outsiders when major camps within the field disparage each other's work?

Given the current political context, it is essential for the research community to delineate the many good ways of doing high-quality research, and then live up to the standards it sets. Science advances by testing hypotheses from all credible viewpoints, not by applying predetermined methods (e.g., randomized controlled trials) independent of context. The goal is to provide rigorous, evidence-based warrants for one's claims; the idea is to match the method(s) with the issue at hand, and to only draw conclusions warranted by each method or the methods in combination (see, e.g., National Research Council [NRC], 2002; Schoenfeld, 2002).

Some potentially productive changes include the following. (We note that all of them are in line with Lagemann's [2002] and others' calls for more usable knowledge in education.)

Change 1: Team research on substantial projects. Research that will have an impact on large-scale practice must be of much larger scale than the typical studies conducted today (National Academy of Education, 1999) or even the work of a small group of academics over a year or two. The same is true for establishing more general principles with a known range of validity.

Change 2: Closer links between researchers and development groups. Many of the problems discussed above will be mitigated if more researchers choose to work with development teams directly on the treatments they design, both during development and afterward. This provides opportunities for in-depth analysis of potentially important and well-developed treatments. It provides a much more direct route from research to improved practice. It will also reveal more basic insights, unclouded by the limitations of underdeveloped treatments.

Change 3: Studies of widely available treatments. Comparative in-depth studies of alternative treatments (e.g., all the NSF middle school mathematics curricula plus some standard texts, some with radically different philosophical bases) would give better information on bottom-line matters, such as which curricula produce which outcomes (linked to standards) under which conditions. They would also provide rich feedback from use in practice to inform the further improvement of the curricula. Engineers call this process *successive iterative refinement*. Such studies would also illuminate many deeper research issues, including the treatments *versus* principles issue—how much the outcomes depend on common principles and how much on more detailed matters of design.

Change 4: More standardization of detailed methods and instruments. Although progress demands new ideas and research tools, standardization has important benefits in reducing the uncertainties of comparability studies. Successive refinement, widely used in other fields, can yield robust tools. A requirement to justify *not* using established instruments and methods would provide pressure in this direction.

Change 5: The reward system. To prosper, the kind of engineering research we advocate must be recognized as full-value research currency by the academic community where it counts—in appointment, tenure, and promotion procedures.

For example, creative design with systematic development and evaluation needs to be given as much credit as insight-enhancing papers in prestigious journals. Such products make comparable (though different) intellectual demands on their producers. Currently, evaluators may well get more academic credit than those who create the materials—just as, at one time in the humanities, critics got more academic credit than authors or composers. This is no longer true in the humanities, the arts, or in engineering—or even in science, where many Nobel prizes are awarded for inventions. Artists, composers, accountants, and engineers are now appointed to academic positions, and their research is judged on its contributions to improving practice.

Major levers for change overlap with those discussed in section A. Additional levers might be:

- providing training for researchers in the broad range of skills involved in impact-focused research;
- funding programs that support work that contributes to both research and practice, and supporting organizational structures that are strong in this respect;
- encouragement by journal editors and reviewers of work that explores materials and insights in a wide range of contexts, and that builds in meaningful ways on prior research;

- evaluation and recognition of first-rate work published through other less-permanent channels such as the web and, complementing this, of scholarly reviews of such work;
- inclusion in criteria for publication or promotion the discussion of the likely impact of the work on practice.

C. A Reasonably Stable Theoretical Base

Many barriers to achieving a stable theoretical base were discussed above. These include a tendency in the field toward grand theories and claims for them, intra-communal disputation, a reward structure that favors iconoclasm and doing your own thing, and a lack of incentives for programmatic work aimed at refining theory in detail (e.g., describing the conditions/contexts in which certain ideas or treatments have been shown to "work"). The changes needed include the following.

Change 1: Seeing the big picture; putting theory in its place(s). Over the past few decades educational research has made significant theoretical progress, as various interdisciplinary fields have produced much more comprehensive and synthetic explanations of education-related phenomena than the fields that they drew from. For example, cognitive science and sociocultural theory have provided more robust and detailed descriptions of cognitive and social phenomena than many of their constituent disciplines. Yet, we have a long way to go, with regard to the construction of more robust, more encompassing theories (see, e.g., Schoenfeld, 1999), and in understanding which aspects of which theories apply strongly in which contexts.

Most of the theories that have been applied to education are quite broad. They lack what might be called "engineering power." To put it a different way, they lack the specificity that helps to guide design, to take good ideas and make sure that they work in practice. This, of course, is not unique to education. Newton's Laws of Motion, which describe the effects of forces on motion, are the paradigmatic example of a grand theory. Aeronautical engineering uses Newton's Laws of Motion. But it also uses a great deal of more local, phenomenological theory such as that which describes the elastic properties of aluminum alloys. This kind of local theory provides reliable input to design of limited but known range—information that is every bit as essential for effective engineering as the grand theory. It is essential to know the limits of each piece of theory as much as its strengths.

In aeronautical engineering, the diverse elements of theory are cumulatively strong. Those who know the theory can design an airplane at a computer and build it. The airplane will fly—efficiently. (It will still be flight-tested extensively to optimize its design, and to detect and remedy problems.) In medicine, theory is moderately weak, but getting stronger. Advances in surgery are firmly based on phenomenological knowledge of anatomy and physiology. In contrast, for the design and development of new drugs theory is weak. Despite all that is known about physiology and pharmacology, new drugs are mainly found by testing the effects of very large numbers of naturally occurring substances; they are chosen intelligently, based on analogy with known drugs, but the effects are not predictable and the search is wide. However, as fundamental work on DNA (surely a grand theory) has advanced, and with it the theoretical understanding of biological processes, designer drugs with much more theoretical input have begun to be developed.

Education lags far behind in the range and reliability of its theories. By overestimating theories' strength (or perhaps better, by not constraining their application appropriately) damage has been done. This is the case for theories as different as behaviorism or constructivism. In the case of behaviorism, a very narrow conception of knowledge acquisition banished a wide range of important cognitive processes from consideration. In the case of constructivism, a naïve faith in "cooperative learning" as a mechanism for generating productive conversations has sometimes resulted in pleasant but content-free discussions. It is not that behaviorism or constructivism is "wrong"; indeed, each is "right" in some important ways. The harm comes from overestimating their generality and power, and underestimating the need to specify the contexts in which they are effective and the steps necessary to implement them successfully.

To summarize, general theories are weak, providing only general guidance for design; nonetheless they receive the lion's share of attention in the research literature. Local or phenomenological theories based on experiment are seen as less important or prestigious than general theory but are currently more valuable in design (as well as easier to establish). We need both, and a value system that rewards both.

Change 2: Robust consensus building and public dissemination of well-substantiated results. Cutting-edge research that appears in the *New England Journal of Medicine* is widely reported; everyday consensus dealing with significant health issues is disseminated on a regular basis by publications such as the *Berkeley Wellness Letter*. If a politician makes a suggestion that violates contemporary medical wisdom, the American Medical Association is sure to pounce—with effect. In contrast, political expediency almost always trumps educational research: consider many states' rush to high-stakes exams with consequences such as retention in grade, despite the clear evidence that the main effect of retention is to increase dropout rates.

It wasn't always this way in medicine: there is a long history of quackery, patent medicines, and other dubious practices. Medicine is hardly free of such things today—but they are at the fringe, and there is a public perception that medical consensus is robust. Education must work toward a similar state. It is a sign of abdication of professional education societies' role in this regard that the most important consensus documents regarding educational substance (for example *How People Learn* [NRC, 1999a], *Preventing Reading Difficulties in Young Children* [NRC, 1999b], and *Adding it Up* [NRC, 2001]) in recent years have come from the National Research Council. While we as a field should welcome such efforts on the part of our NRC colleagues, we should also embark on a major effort to work, fieldwide, toward consensus statements regarding fundamental issues in education. It is vitally important to establish a gradually growing core of research results that are generally accepted within the research community, and publicized as such. However difficult, this will be a critical step in improving the status of educational research, and minimizing the influence of bias and special interests. It will need additional meta-analytical work along the lines of the Cochrane Collaboration in medicine[4] and the Campbell Collaboration in education.[5]

We wish to stress that a consensus on *findings* need not come at the cost of methodological pluralism. We are in complete sympathy with the NRC's report *Scientific Research in Education* (NRC, 2002) that, as in all science and engineering, there is a wide range of ways of conducting high-quality research in education—and that triangulation using multiple methods is one fundamental way to establish robust findings.

D. *Teams of Adequate Size*

As noted often above, large teams encompassing a wide range of design, analysis, and theoretical skills are necessary in order to grapple with large-scale engineering tasks in education. Some barriers to building the desired infrastructure are as follows:

Barrier 1: The personal costs of collaboration. Working intensively as part of a team involves significant loss of autonomy, organizational complexities and the concomitant expenditure of time and energy, and a possible loss of individual status in authorship (and perhaps identity!).

Again, we observe that "big" science, engineering, and medicine have value systems that accommodate the work of big teams, and yet give appropriate credit to individuals. The main compensations for the costs described in the previous paragraph are the much greater impact of their research on the field, and the greater support that follows. What we are recommending might be called "big education." It needs similar organizational forms.

Barrier 2: Currently limited funding for such major enterprises. The absence of steady funding to support large R&D teams is all too clear.

Barrier 3: Institutional structures inhospitable to such teams. Can you name a major university that guarantees the stable, long-term employment (whether under the name of tenure or long-term contracts) of a team of designers to work in partnership with insight-focused researchers?

The following changes would be productive.

Change 1: The creation of existence proofs. It would be valuable for a funding agency, as a "proof of concept" experiment, to support a few large engineering efforts (including a curricular implementation and refinement study as described in IIIA) to test the effectiveness of the engineering approach in education. Once such work is shown to have demonstrable impact, the climate may be more welcoming for other such efforts.

Change 2: Vertical integration of research effort. By vertical integration we mean a well-established flow of ideas and research results in both directions between small-scale exploratory studies, systematic development of tools and processes for their use, implementation initiatives, and comparative evaluation in depth.

This can sometimes be achieved through opportunistic collaborations that serve all concerned (e.g., the school district, the development groups, and the researchers). However, it is more likely that stable teams and collaborations will be productive—it takes time to develop trust and to learn to collaborate effectively. Equally important, it takes a long time for a collaborative team to discover and document the complexities of large-scale implementation.

The changes we describe may be thought of as a rebalancing of overall research efforts across different levels of research and development, which are delineated in Table 1. Note that the R↔P foci are different at each level, but that research and development are deeply intertwined through the entire process, from initial learning studies to large-scale systemic implementation.

Currently, nearly all detailed classroom research, including design experiments, is at the individual teacher (IT) level. A better balance across the levels is needed, if research and practice are to benefit from each other as they could. It goes without saying that much larger teams than currently exist will be necessary to explore issues at the RT and SC levels. Note that such studies would

Table 1
Four Levels of R&D

Level	Variables	Typical Research and Development Foci
Learning (L)	Student Task	R: Concepts, skills, strategies, metacognition, beliefs D: Learning situations, probes, data capture
Individual Teacher (IT)	Instruction Student Task	R: Teaching tactics and strategies, nature of student learning D: Classroom materials that are OK for some teachers
Representative Teachers (RT)	Teacher Instruction Student Task	R: Performance of representative teachers with realistic support. Basic studies of teacher knowledge and competency D: Classroom materials that "work" for most teachers
System Change (SC)	System School Teacher Instruction Student Task	R: System change D: Tools for Change (i.e., materials for: classrooms, assessment, professional development, community relations)

provide the real "gold standard" for educational research—detailed documentation of what really happens when a reasonably well-defined instructional "treatment" is implemented in practice, with full descriptions of the impact of typical usage, contexts that are productive and problematic, and "side effects" that can be anticipated.

Levers for change include the systemic encouragement and rewarding of such work as academic "coin of the realm" and the funding of such enterprises on the scale and time scale required (see section E).

E. Sustained Funding

An effective R↔P program will require substantial and sustained funding. Recall that in 1998 the United States spent less than $30 million, 0.01% of its overall $300 billion education budget, on basic education research. In contrast, the Pfizer Pharmaceutical Corporation describes its expenses for basic R&D on animal health care as follows:

> We're a world leader in animal health care. We're devoted to animals their entire lives, from their first vaccination to medications for older pets. We lead the industry in research, *spending over two hundred million dollars a year,* looking for new treatments designed specifically for animals. In fact, we're the people who introduced the first arthritis medication in the U.S. specifically for dogs. (*Smithsonian*, June 1999, pp. 14–15)

Where there's a will—or at least a profit—there's a way. In fields where there is an infrastructure that supports the various levels of R&D from small- to large scale, there are mechanisms for bringing ideas into practice. Why is there no investment on a comparable scale in education? It may be that that there is no public confidence that a research-based approach can deliver improvements in education commensurate with such investment. Where are the examples, like those in medicine (e.g., aseptic surgery, anesthetics, antibiotics) that show the power of this approach in education?

Such examples are tentatively in the wings, in areas such as literacy and mathematics. In mathematics, for example, we have the 25-year movement from basic research to large-scale curricular implementation of new curricula—recall the discussion of "Model 5: the long route" in Section I. Given this, the detailed study of such treatments is a logical "next step" in the engineering process.

Needed changes and levers. The necessary change is simple: more money, targeted in specific directions.

Educational research is not necessarily a zero sum game: the right kinds of research could become a significant growth industry. In this regard, it is worth looking back 100 years at medicine. At the turn of the 20th century, there was negligible funding for medical research. The field barely existed, and certainly hadn't proved itself. Over the course of the 20th century, as medical research matured, it produced demonstrable and robust results. The rest, as they say, is history. Other fields that were seen to have practical payoff (e.g., electronics, nuclear physics, molecular biology) have shown similar patterns of growth in funding, including funding for basic research.

Evidence of payoff is the key lever, hence, our proposal for existence proofs. These would be good at the level of individual products, better yet at the institution level. Modern medical R&D was catalyzed by the Flexner (1910) report. New models of medical R↔P capability such as the Johns Hopkins University medical school altered the direction of the field irrevocably. Similarly, education will need some pioneering institutions to provide existence proofs of how educational research in and into practice can be done, at the highest levels of quality.

Evidence of the field's increased capacity to "deliver" on R↔P projects is a second lever. We will be seen as much more credible by both the public and funding agencies once we come to a consensus on core knowledge and methods, and refine and elaborate on theories (both local and global) in ways that are robust and coherent (cf., Section C). R↔P proposals are much more likely to be well received when they are supported by well-developed theoretical and methodological frameworks. And, the value system in the research community must reflect the importance of practical impact, just as it must recognize important new insights that combine trustworthiness with proven generality.

Finally, it should be noted that first steps need not come at significant cost to current research efforts. There are some large pots of money available from the federal government for systemic change as well as for basic research (see, for example, the National Science Foundation's Research on Learning and Education (ROLE) program[6]). It would be interesting to see a consortium of universities propose a design experiment, in concert with some number of school districts, that would achieve some of the goals discussed here.

F. Individual and Group Accountability

Why discuss accountability? Standards for publishing research are stringent, and promotion and tenure decisions reflect very strict forms of cumulative individual accountability.

The criteria for such judgments are currently inward looking. If education research is to become more useful, more influential, and better funded, it must change. It must recognize and reflect society's priority: substantial improvements in the performance of the education system. This is not unusual in other fields: Can you imagine a medical or engineering school that was not held accountable for "making a difference" in some important ways? Nor is it in conflict with original "blue skies" thinking. At some point even think tanks have to justify some proportion of what they do as leading to new or better things.

It goes without saying that real or potential influence cannot be the sole criterion for judging a body of work—it is not, now, in "applied" units of universities or in industrial laboratories. (Recall that Bell Laboratories produced a number of Nobel Prizes for fundamental research!) But it should play an increasing role, if educational research is to make a difference. Once again, we firmly believe that the research community is not facing a zero sum game: increased funding, recognition, and respect from society at large will come as the educational enterprise develops the capacity to address research in and into practice. Clearly, visible practical impact generates resources, as the history of medicine, science and engineering over the past century attest.

Concluding Comments

Educational research is at a potential turning point. One can argue that the low status of educational research is inevitable, a matter of context. Just about everybody, having gone to school, thinks he or she is an expert on education—"funding is abysmally low," "practical problems are intractable," "the Feds don't understand us," and so on. With all due respect, these are excuses. If people feel they are entitled to render judgments on educational issues, it is because we have not taught them otherwise. Why haven't we? Alas, many of the barriers to progress identified in this article can only be seen as self-inflicted wounds. Intra-communal disputation has kept us from rallying behind any number of potential consensus statements and speaking with authority. That and a lack of attention to coherent theory building leave us looking balkanized and incoherent, the whole of education being less than the sum of its parts. It also leaves us vulnerable to attack from outside—powerful politicians, and some academics, who understand little of what educational research is all about feel empowered to tell us how to go about our business. Most important, the decoupling of research from practice leaves us both ineffective and vulnerable.

Change is possible, and it would be much to our advantage. Attending to theory in the proper ways will enhance both our work and the reputation of the field. But theory qua theory will take us only so far (and not far enough). Positioning ourselves so that we can make progress on fundamental problems of practice will make the big difference.

The analogy with medicine a century ago is profitable. Read the Flexner (1910) report! It is fascinating in its own terms, but also because of the ways in which it can inform us. The report established the basis for changes in the medical profession over the course of the 20th century. Once balkanized and disputatious, medicine began slowly to cohere into a discipline. A few pioneering institutions redefined the relationship between research and practice, moving toward much tighter linkages. Standards rose, values changed. And, as the impact of medical scholarship increased, a richer infrastructure—both in terms of funding and in terms of teams capable of tackling increasingly large problems—developed. We owe ourselves, and the nation, no less.

NOTES

In the spirit of a true collaboration, each of the authors did at least 75% of the work in producing this article. The order of their names was determined by the toss of a coin.

[1] What counts as "scientific research in education" is now hotly contested (see, e.g., NRC [2002] and Jacob & White [2002]).

[2] Stokes invokes the following idea: Research may score high or low on "contributions to theory" and high or low on "contributions to practice." Thus, any piece of research falls into one of four quadrants. Bohr focused on theoretical work without specific attention to applications—the "high/low" quadrant. Edison paid little attention to theory and focused on major practical applications—"low/high." Pasteur's work made major theoretical contributions and was intended to be highly practical. The "high/high" quadrant is named after Pasteur.

[3] Note that the process begins with curriculum components—chunks of instruction that are substantial (say a number of weeks' worth of instruction) but manageable in scope.

[4] Information regarding the Cochrane Collaboration in medicine can be found at http://www.cochrane.org/

[5] Information regarding the Campbell Collaboration in education can be found at http://www.campbellcollaboration.org/

[6] Information regarding the ROLE program can be found at http://www.nsf.gov/pubs/2000/nsf0017/nsf0017.html

REFERENCES

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, *25*(6), 5–11.

ARC Center (2003). Full report of the Tri-State Student Achievement Study. Retrieved September 12, 2003, from http://www.comap.com/elementary/projects/arc/tri-state%20achievement%20full%20report.htm

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt-Reinhardt-Winston.

Blumenfeld, P., Fishman, B., Krajcik, J., Marx, R., & Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist*, *35*(3), 149–164.

Briars, D. (March, 2001). *Mathematics performance in the Pittsburgh public schools.* Presentation at a Mathematics Assessment Resource Service conference on tools for systemic improvement, San Diego, CA.

Briars, D., & Resnick, L. (2000). *Standards, assessments—and what else? The essential elements of standards-based school improvement*. Pittsburgh, PA: University of Pittsburgh.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*(2), 141–178.

Burkhardt, H. (1988). The roles of theory in a "systems" approach to mathematical education, article in honor of prof. Hans-Georg Steiner's 60th birthday. *International Reviews on Mathematical Education, ZDM 5*, 174–177.

Burkhardt, H. (2001, December). *Styles or research: Insight and impact*. Paper given at the ICMI Algebra Conference, University of Melbourne, Melbourne, Australia. Retrieved February 20, 2003, from http://www.nottingham.ac.uk/education/MARS/papers/

Burkhardt, H., Fraser, R., Coupland, J., Phillips, R., Pimm, D., & Ridgway, J. (1988). Learning activities & classroom roles with and without the microcomputer. *Journal of Mathematical Behavior*, *6*, 305–338.

Burkhardt, H., Fraser, R., & Ridgway, J. (1990) The dynamics of curriculum change. In I. Wirszup & R. Streit (Eds.), *Developments in school mathematics around the world, Vol. 2* (pp. 3–30). Reston, VA: National Council of Teachers of Mathematics.

Clark, M., Haney, W., Madaus, M., Lynch, C., & Lynch, P. (2000). High stakes testing and high school completion. National Board on Educational Testing and Public Policy Statements, *1*(3). Retrieved September 10, 2003, from http://www.bc.edu/research/nbetpp/publications/v1n3.html

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin, Germany: Springer.

Design-Based Research Collaborative. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8.

Dillon, S. (2003, February 6). Thousands of schools may run afoul of new law. *New York Times*. Retrieved September 12, 2003, from http://www.nytimes.com/2003/02/16/education/16EDUC.html

Downton, A. C. (Ed.). (1992). *Engineering the human-computer interface*. New York: McGraw Hill.

Elmore, R. F. (2000). *Building a new structure for school leadership*. Washington, DC: Albert Shanker Institute.

Fishman, B., Honey, M., Hug, B., Light, D., Marx, M., & Carrigg, F. (2003, April). *Exploring the portability of reform: One district's approach to adaptation*. Paper presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, IL.

Flexner, A. (1910). *Medical education in the United States and Canada: A report to the Carnegie Foundation for the Advancement of Teaching* (Bulletin No. 4). New York: Carnegie Foundation for the Advancement of Teaching.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Gomez, L., Fishman, B., & Pea, R. (1998). The CoVis project: Building a large-scale science education testbed. *Interactive Learning Environments*, *6*(1–2), 59–92.

Higher Education Funding Council. (1999). *Guidance on submissions research assessment exercise*, Paragraph 1.12. London: Higher Education Funding Council for England and Wales 1999. Retrieved September 12, 2003, from http://www.hero.ac.uk/rae/

Honey, M., & McMillan-Culp, K. (2000). Scale and localization: The challenge of implementing what works. In M. Honey & C. Shookhoff (Eds.), *The wingspread conference on technology's role in urban school reform: Achieving equity and quality* (pp. 41–46). New York: EDC.

House Committee on Science. (1998). *Unlocking our future: Toward a new national science policy. A report to Congress by the House Committee on Science*. Washington, DC: Author. Retrieved July 9, 1999, from http://www.house.gov/science/science_policy_report.htm

Jacob, E., & White, C. S. (Ed.). (2002). Theme issue on scientific research in education [Special issue]. *Educational Researcher*, *31*(8).

Kaestle, C. (1993). The awful reputation of education research. *Educational Researcher*, *22*(1), 23–31.

Kelly, A. E. (Ed.). (2003). Theme issue on the role of design in educational research [Special issue]. *Educational Researcher*, *32*(1).

Lagemann, E. (2002). *An elusive science: The troubling history of education research*. Chicago: Chicago University Press.

Magidson, S. (2002). Teaching, research, and instructional design: Bridging communities in mathematics education. *Dissertation Abstracts International A* 63/09, p. 3139. (UMI No. AAT 3063466)

Means, B., & Penuel, W. (2003). *Implementation variation and fidelity in an inquiry science program: An analysis of GLOBE data reporting patterns*. Menlo Park, CA: SRI International.

National Academy of Education. (1999). *Recommendations regarding research priorities: An advisory report to the National Educational Research Policy and Priorities Board*. New York: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Research Council. (1999a). *How people learn* (J. Bransford, A. Brown, & R. Cocking, Eds.). Washington, DC: National Academy Press.

National Research Council. (1999b). *Preventing reading difficulties in young children* (C. Snow, M. Burns, & P. Griffin, Eds.). Washington, DC: National Academy Press.

National Research Council. (2001). *Adding it up*. (J. Kilpatrick, J. Swafford, & B. Findell, Eds.). Washington, DC: National Academy Press.

National Research Council. (2002). *Scientific research in education*. (R. J. Shavelson & L. Towne, Eds.). Washington, DC: National Academy Press.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

Schoenfeld, A. H. (1999). Looking toward the 21st century: Challenges of educational theory and practice. *Educational Researcher*, *28*(7), 4–14.

Schoenfeld, A. H. (2002). Research methods in (mathematics) education. In L. English (Ed.), *Handbook of international research in mathematics education* (pp. 435–488). Mahwah, NJ: Erlbaum.

Schoenfeld, A. H. (in press-a). Design experiments. In P. B. Elmore, G. Camilli, & J. Green (Eds.), *Complementary methods for research in education*. Washington, DC: AERA.

Schoenfeld, A. H. (in press-b). Math wars. In B. C. Johnson & W. L. Boyd (Eds.), *2004 politics of education yearbook*.

Senk, S. L., & Thompson, D. R. (Eds.). (2002). *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Erlbaum.

Smithsonian Magazine (1999). *30*(3).

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technical innovation*. Washington, DC: Brookings.

U.S. Department of Education (2002). *Strategic plan for 2002–2007*. Retrieved September 12, 2003, from http://www.ed.gov/pubs/stratplan2002-07/index.html

U.S. Department of Education (2003). *Conference on scientific evidence in education*. Retrieved September 12, 2003, from http://www.excelgov.org/displayContent.asp?Keyword=prppcEvidence

AUTHORS

HUGH BURKHARDT leads the MARS Shell Centre Team, School of Education, University of Nottingham, Jubilee Campus, Nottingham NG8 1BB, UK; Hugh.Burkhardt@nottingham.ac.uk. Since he joined the Shell Centre for Mathematical Education at the University of Nottingham as Director in 1976, he and his colleagues have worked to develop the "engineering research" approach to R↔P. He is still an occasional theoretical physicist and applied mathematician.

ALAN H. SCHOENFELD is the Elizabeth and Edward Conner Professor of Education in the Graduate School of Education, University of California, Berkeley, 94720-1670; alans@socrates.berkeley.edu. President of AERA in 1999–2000, Schoenfeld has done basic research on mathematical thinking and teaching. He is collaborating with local schools to improve mathematics instruction, focusing on issues of diversity.

Manuscript received April 4, 2003
Final revision received September 12, 2003
Accepted September 18, 2003