The Research That Policy Needs

by Fritz Mosher, Susan H. Fuhrman, and David K. Cohen

The past half-century has witnessed an epochal transformation of the goals of education policy, particularly in how we judge the time-honored role of American schools in ensuring equal opportunity. The focus has shifted from inputs—that is, whether all students have reasonably equal access to the attributes of good schooling including qualified teachers, a solid curriculum, safe and well-equipped facilities, reasonable class sizes, and equitable funding—to whether virtually all students at least achieve proficiency in core knowledge and skills by the time they leave school.

In combination with such influences as the civil rights movement and international economic and political competition, education research helped to spur this shift by casting light not only on unacceptable inequities in outcomes, but also on the disappointing overall performance of American students both on the National Assessment of Education Progress (NAEP) and relative to their counterparts in other countries. True, researchers have identified interventions and policies that correlate with improved school outcomes, yet the more telling role of education policy research and evaluation has entailed examining the effectiveness of reform policies and demonstrating their limitations. Research evidence regarding what does not work has been at least as influential as evidence about effective practices and probably has had a greater impact, both by stirring a sense of crisis and by motivating the search for more promising responses.

At this juncture, we have a picture of an education system caught in transition. Standards- and outcomes-based accountability measures are in place, to be sure, but regulations and pedagogical orientations still reflect the premise that equal access to conventional inputs and resources should be sufficient to ensure equal opportunity. The system lacks fundamental knowledge about key factors—the approaches to teaching and learning, and the social organization of schools—that would be sufficient to enable substantially all students to meet or exceed

the desired standards. Acquiring such knowledge could make possible a defensible definition of "opportunity to learn." Absent such information, policymakers will be sorely tempted to define standards and proficiency in terms of the same lowest-common-denominator outcomes the system currently produces, even though such low-bar outcomes would leave many students with little chance of functioning in the world in the way the term "proficiency" should imply. The question is, what might increase the chances that research could help produce the knowledge and tools that would enable policy—and schools—to fulfill the new ambitions?

What Is Sufficient to Ensure Success?

Our reading of both the research and the current policy environment suggests several fundamental reasons why it is so difficult to develop the knowledge needed to inform policies that might enable standards-based reform to succeed.
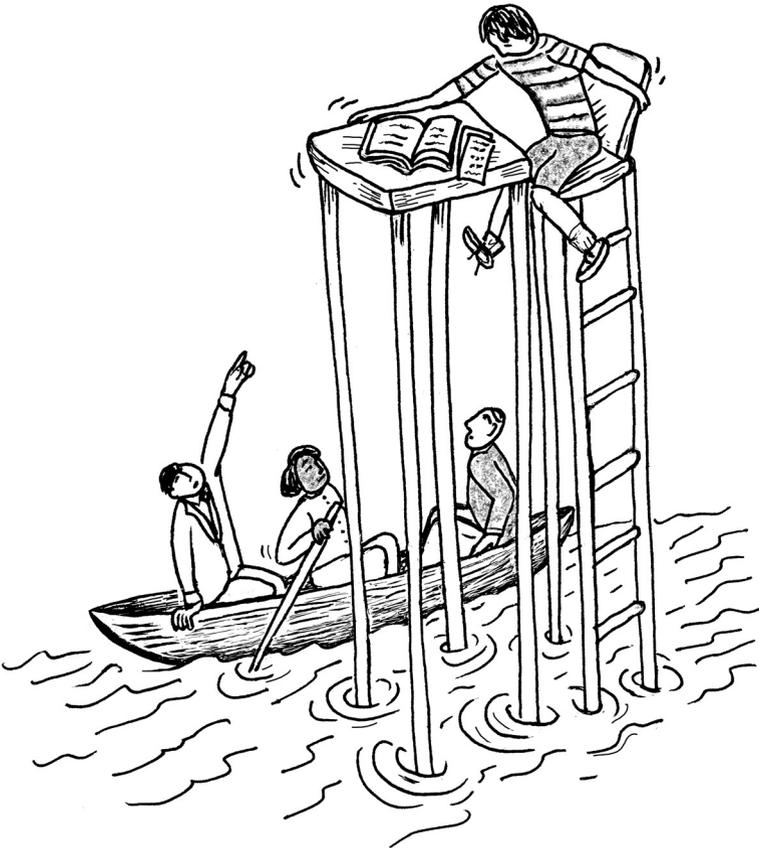
First is an inadequate conception of the goal of the system—i.e., proficiency in key subjects and skills—and how proficiency should be measured. The term proficiency better connotes skill development than it does breadth or depth of knowledge, notwithstanding the natural overlap of skills and knowledge. Proficiency also implies a level of competence that would provide substantial prospects of success when applying a skill, whether in further study, employment, citizenship, or parenthood. However, *none* of the measures used in state assessments (or even in NAEP) has any direct empirical validation for such an interpretation of current proficiency levels. At best, the measures are based on judges—curriculum specialists, experienced teachers, and sometimes parents or employers—eye-balling items in assessments and choosing those that fit their expectations of what a proficient student should be able to do (and a less proficient student could not). Those judgments are never tested against other observations of the student or against more complex assessments of a student's effectiveness in real-life settings. Neither, for the most part, are assessments aligned to any well-defined conception of how and in what order, steps, or stages knowledge and skills are acquired over time and with instruction. Current assessments provide only limited information for guiding instruction.

The result is that public discussion of academic standards and learning outcomes takes place in an empirical vacuum. There is no way to determine how near, or how far, the goal of student proficiency might be, nor is there a shared basis for considering what the tradeoffs might be for setting the goal higher or lower when people legitimately differ on costs and benefits. We simply cannot tell, for instance, whether schools embracing current best practices can in fact succeed with most children; whether exceeding the typical time or effort might help a lot or a little; whether success might require fundamentally new knowledge; or what we, as well as the least-performing students, will lose if we settle for a lesser standard of proficiency. Research that could better inform outcomes-based policy desperately requires development of better measures of the outcomes themselves. Such work will require a lengthy iterative process that explores basic work on core subjects and skills, the ways they are learned, and the role those subjects and skills play in effective performance in the world. Still, it is time to get started.

Second, research shows that the teaching side of the teaching/learning interaction in instruction is crucial, but we know too little about what makes it so. One implied imperative of standards-based reform and accountability is to adapt instruction to meet the needs of children who are not on the path to meeting standards. Policy therefore needs to affect what actually happens between teachers and students in specific classrooms. But we have little direct evidence about what teach-

ers do in those rooms and how it affects student learning. In the United States, much of both policy and research stops at the classroom door, and we are left to study proxy measures. Recently, a growing literature has developed based on direct classroom observation, analysis of video records of instruction, teacher self-reports, teacher logs, and the collection of instructional artifacts. These studies explore teachers' pedagogical behavior and decision-making more directly, but as yet there is no clear agreement about the crucial behaviors to observe or how to sample them reliably and validly.

This work has to be pushed much further if we are to learn whether reform policies, pre- and in-service teacher education, or other efforts to improve student outcomes are having any effect on teacher behaviors. Unless we develop a common vocabulary and a common set of tools for studying instruction, most of our research will be ineffective. It also will be important to tie work on instruction much more closely to specific content, subjects, and skills, since pedagogical-content knowledge is likely to prove to be a crucial element underlying effective instruction. The growing enthusiasm for formative assessment and the use of evidence in instruction also may help to call attention to the decisions that teachers make concerning their students and whether those decisions focus on student progress in specific subjects and skills.

Third, we need to solve the problem of finding real-school settings for conducting multivariate research and, more particularly, development at a scale and duration that will produce usable, proven knowledge, tools, and policies for teaching practice. This is easier said than done.

Time for a Full-Court Press

Education research over the years has identified multiple factors that are associated with increased student performance. Some arguably are necessary, but none, either separately or together, has produced widespread, demonstrable success for most or all students, or rivaled family background or social status in significance. Even research based on current performance measures (which clearly fall short of assuring that students will function effectively in real-life situations) fails to identify combinations of factors that would enable most students to meet those standards. While searching large data sets to identify correlations between variations in educational inputs and student outcomes can help to suggest design ideas for potentially effective interventions, it is likely that the naturally occurring variation in American schools will not include all the factors, the requisite levels of effort, and certainly not the combinations of factors that will be required for success.

Now is the time to design and develop best-bet instructional interventions—what Cohen, Raudenbush, and Ball (2003) call "instructional

regimes"—that combine hypotheses about promising approaches derived from large-scale correlational studies, new evidence stemming from basic research and well-supported theory, and the best wisdom of practice. These in turn should be tested in schools serving at-risk populations at the same time there is a full-court press to ameliorate other conditions that disrupt learning. The goal is to discover what combinations seem sufficient to enable most students at least to meet reasonable standards.

A notable example of such a strategy is Success for All, the work of Robert E. Slavin and his colleagues (detailed in the authors' *The State of Education Policy Research*, and elsewhere). We need many more such examples, sustained over comparable periods of time and across a range of school settings. We also need for work of this sort to be formatively evaluated, both to make running improvements in designs and to identify problems that may require additional attention from funders and researchers, and that then could inform and improve future designs.

This advice may be difficult to act on, and the authors think that the reasons behind the difficulty are the reasons why education research has such a poor track record. In many fields, investigators can translate results of basic laboratory work into designs and test them relatively quickly in working settings—modifying and re-testing if necessary. Such combinations of science and engineering are well documented. However, the time required both to act and to see results is much greater in education (and other social institutions). There tend to be many more levels at which conditions must be held equal or varied in measurable and replicable ways.

Consider these factors:

- curriculum or pedagogy
- levels of funding available to schools
- background, experience, and training of teachers and the in-service training, time, and support available to them
- quality and behavior of school and district leadership
- coherence of curriculum and degree of consistency between curricular and pedagogical expectations and the criteria for student, teacher, school, and system accountability
- degree of stress from the latter
- social backgrounds of students and parents and levels of community resources and stress
- positive and perverse incentives for performance or lack of incentives, both in the system and in the local and student cultures

When we urge a full-court press, we are suggesting that the most promising interventions need to push these variables toward benign

ranges—that is, toward conditions that are not so bad as to make it unlikely that students can succeed, but not so costly or unusual as to make the schools unique, unaffordable, or politically unsupportable—while solving instructional problems tied to the cognitive and social needs of individual students and groups of students.

Researchers and educators do not know how to solve the problem of finding authentic settings in which to test these complex forms of design and intervention. Nonetheless, it is a problem that needs at least a partial solution before there is much chance of obtaining the knowledge we need in the form in which we need it. Perhaps we will decide to look beyond the current realities of American schools to identify hypotheses about what may happen if those realities change radically. For instance, what if state departments of education were to assume roles more like those played by ministries of education in some other countries, with the effect of providing a more coherent setting for experimentation with instructional regimes than is now possible in the fragmented American system? Or are there ways to take advantage of the charter school movement to encourage the development of subsystems within which real experimentation might be carried out? These possibilities strike the authors as a high-priority set of issues that should be the focus of serious discussion among researchers, policymakers, and funders of education research. (A 2006 *Education Week* essay by Paul Hill makes parallel points directed toward the Bill & Melinda Gates Foundation.)

Strategic Focus on Key Goals, Big Problems

And that brings this essay to a fourth point. None of its first three concerns can be addressed effectively unless those who fund and manage education research take a much more strategic view of what they are doing. By "strategic" we mean that funders and managers should focus on the key goals of practice or on particular big problems that seem to impede attainment of those goals. They should try to determine whether the current understanding of factors affecting the focal areas can support the design of practical approaches and solutions that might have big effects in moving practice toward the goals. If current understanding does not seem sufficient to support such work, funders and managers should consider providing programmatic support for basic disciplinary or multidisciplinary studies in areas that show promise of identifying understandings with potential to inform new designs that might produce big effects. (When we refer to "big effects," we mean, for instance, effects as large as those accomplished by the demonstration that an underlying ability called phonological awareness is key to a child's ability to develop fluent decoding in early reading and that interventions explicitly calling attention to and providing practice in the

alphabetic principle can help bring children who are low on this ability into the normal range of ability to decode.)

In addition, if design work encounters problems that seem to require new knowledge, funders and managers should devote resources to basic efforts to understand and remove those impediments. Over time, the combination of targeted funding and strategic management has real potential to generate interacting cycles of basic work and design work, with the outcomes of one cycle informing and establishing priorities for the other. This approach requires finding reliable ways to monitor progress and problems in both implementation and design, as well as shifting, balancing, and orchestrating resources as needed. Fixing the problems with student assessment clearly will require both fundamental and practical design work, interacting in a reciprocal relationship and playing out over time. Likewise, dissecting what actually happens in instruction depends on understanding which instructional elements are key to making big differences in student outcomes. At the same time, acting strategically involves a tradeoff—focusing on a few problems versus spreading resources across more areas. Striking such a balance means making serious estimates of progress on a problem and weighing the anticipated gains against the unexpected progress that casting a wider net might reveal.

The question of how and where to make these strategic judgments is a formidable design problem in its own right. Clearly the major research funders—both federal sources and private foundations, as well as state and commercial sources, should they decide to expand their roles—must take major responsibility, although success may prove elusive without better interaction between the funders and the field. Peer review is a crucial mechanism for ensuring basic quality and adherence to methodological standards and disciplinary relevance, but it is not a sufficient solution to engaging the field in making judgments on strategic priorities.

One solution might involve developing research-management institutions of sufficient scale and resources to interact with universities, other research organizations, and schools or school systems. Such institutions are not likely to appear spontaneously; they will have to be nurtured by the very funders who then will need them as partners in making strategic judgments. Certainly there must be talented research managers across the country who can be tapped as partners. As an example, more thought should be given to the roles that the federally funded National Research and Development Centers and Regional Education Laboratories could play along these lines.

The Science Behind Educational Inputs and Outputs

Finally, the authors think that the public discourse on these issues has been clouded by a narrow, or imbalanced, understanding of what science

entails. By endorsing "scientifically based" or "scientifically valid" research, the federal laws establishing No Child Left Behind and the Institution of Education Sciences both encourage educators to adopt approaches consistent with such principles. The language of the legislation and its implementation by the U.S. Department of Education seem to call for sound scientific methodology: careful and replicable observations of phenomena and relationships that employ an array of methods appropriate to the questions being studied. Nevertheless, the result has been a clear bias toward experimental evaluations of already identified and currently available pedagogical approaches, curricula, materials, tools, and education policies.

That bias is understandable. One of the main reasons education research has been considered weak is its failure to use methods that allow rigorous causal inferences about the relationships between education inputs and outcomes. It is reasonable to expect that policymakers and educators would find information from real settings about what works, and when, more useful than scholarly findings about the relationships among particular education variables when "other things are held equal," or the results of laboratory experiments.

But there is a catch, as mentioned above: The emphasis on summative evaluation of interventions using some form of randomized trials, or equivalent controls, places a premium on working with interventions that already are identified. Education interventions tend to be complex, however, so that even a demonstrated cause-effect relationship can leave uncertainty about what within the intervention caused the result. Definitive results from a single, large randomized trial are unlikely; the real need is for an extended series of studies spread over varied times and settings. Such studies might eventually isolate effective elements and the extent to which their value can be generalized. Randomized trial studies are expensive, often requiring payment of incentives to induce schools to participate. And they tend to be rigid: to ensure control, a particular version of the intervention has to be locked in during the experiment. For both reasons the experimental periods tend to be relatively short, often one to three years. Yet the effects in education, if they are there, might reasonably be expected only to show up and play out in full over much longer time periods. These short-term and rigid trials work against the kinds of extended and formatively varied efforts that might eventually yield more conclusive and useful results.

This all suggests that serious attention should be given to whether the resources and time made available for the study will allow promising interventions a real opportunity to show what they can do. It also suggests that more attention and investment should be aimed at fundamental work designed to identify basic factors and influences, which in new

combinations might inform interventions that would have much bigger effects. Attention and investment also should be aimed at iterative attempts to design, try, and modify interventions based on these new fundamental insights, as adapted by the necessary input of talented designers and accomplished practitioners. Such a process might eventually build interventions whose promise would justify the even bigger investments required to test their causal effectiveness rigorously.

Science and research are motivated both by the desire to understand and explain and by the need to inform action. The two are not mutually exclusive, and individual researchers may hold both motives in varying degrees. The balance is "basic" when it shifts toward understanding; it is "applied," or even is considered "engineering," when it emphasizes use. Frequently the balance may involve a true equilibrium, as in Donald Stokes's point that Pasteur's efforts to keep wine from going bad laid the foundations for bacteriology. Stokes was advocating increased funding for use-oriented or mission-related basic research, because such research often promises ultimate solutions to high-priority problems, rather than just supporting basic work, wherever it might lead.

Reinforcing that inclination, much of what education policy researchers do can be characterized as a search for a use-oriented understanding of how education systems work. Researchers seek insights and hypotheses in the array of associations presented to them by the real world of schools (although the methods used to identify and weigh those relationships are becoming increasingly more mathematical and sophisticated). A substantial part of the enterprise still involves refinement of definitions, that is, efforts to identify ways to measure aspects of the education experience that seem important because they yield strong associations with other matters of concern. Researchers also report the facts about education matters—how they are distributed, or maldistributed, among students and schools, and whether they change when policymakers and educators take steps intended to change them.

So the authors are happy to encourage an emphasis on use-oriented basic work, but *The State of Education Policy Research* joins a growing body of opinion suggesting that education lacks a strong tradition of engineering and design, as well as the institutional infrastructure and funding needed to support them. It is time for all parties involved to pay much more serious attention to that problem if they hope to do as well for all children as they profess.

A Cautionary Note

We end on a cautionary note. We have tried to identify what it might take to develop knowledge that can help education policymakers and schools attain their goals. Little in the history of the relationship between

research and policy suggests that, even if such knowledge existed, policy-makers or schools would necessarily put it to use. We hope for a better outcome. If our recommendations are embraced, new knowledge will not be buried in journals but instead embodied in tools, materials, policies, and practical advice designed for and tested in use, so that their relevance to practice becomes obvious. We further assume that effective and strategically organized research and development will, over time, result in interventions that demonstrate bigger effects in outcomes with the full range of students. It ought to be more difficult to ignore such evidence of effectiveness, or to go back to picking and choosing among the more anemic and ambiguous results of current studies. Finally, we hope that education researchers learn to advocate more effectively on behalf of both the integrity and the promise of their own work. None of the aforementioned would guarantee that their work would be picked up in the reality of practice, but again, it would be a start.

References

Cohen, D., S. Raudenbush, and D. Ball. 2003. "Resources, Instruction, and Research." *Educational Evaluation and Policy Analysis* 25, no. 2: 119–142.

Hill, P. T. 2006. "Money, Momentum, and the Gates Foundation: What Will Warren Buffett's Gift Mean to U.S. Schools?" *Education Week* 25, no. 44 (August 9): 34, 44.

Frederic A. (Fritz) Mosher, an independent consultant on education policy and research planning, management, and funding, was a program specialist and policy analyst with Carnegie Corporation of New York for thirty-six years.

Susan H. Fuhrman is president of Teachers College, Columbia University, and chair of the Consortium for Policy Research in Education (CPRE).

David K. Cohen is John Dewey Collegiate Professor of Education and Walter H. Annenberg Professor of Education Policy at the University of Michigan School of Education.

This essay is drawn from The State of Education Policy Research *(Taylor and Francis Associates, 2007), co-edited by the authors with the support of the Spencer Foundation.*