

Educational Research: The Hardest Science of All

by David C. Berliner

Under the stewardship of the Department of Education, recent acts of Congress confuse the methods of science with the process of science, possibly doing great harm to scholarship in education. An otherwise exemplary National Research Council report to help clarify the nature of educational science fails to emphasize the complexity of scientific work in education due to the power of contexts, the ubiquity of interactions, and the problem of decade by findings interactions. Discussion of these issues leads to the conclusion that educational science is unusually hard to do and that the government may not be serious about wanting evidence-based practices in education.

Scientific Culture and Educational Research” (this issue), as well as the National Research Council (NRC) report from which it draws, are important documents in the history of educational research. I commend the authors and panelists who shaped these reports, and I support their recommendations. But it is not clear to me that *science* means the same thing to all of us who pay it homage, nor do I think that the distinctions between educational science and other sciences have been well made in either report. There are implications associated with both these issues.

Definitions of Science

I admire Richard Feynman’s (1999) definition of *science* as “the belief in the ignorance of authority” (p. 187). Unrestricted questioning is what gives science its energy and vibrancy. Values, religion, politics, vested material interests, and the like can distort our scientific work only to the extent that they stifle challenges to authority, curtailing the questioning of whatever orthodoxy exists. Unfettered, science will free itself from false beliefs or, at the least, will moderate the climate in which those beliefs exist. As politicians recognize that “facts are negotiable, perceptions are rock solid,” so there is no guarantee that science will reduce ignorance. But as long as argument is tolerated and unfettered, that possibility exists.

Another admirable definition of *science* was provided by Percy Bridgman (1947), who said there really is no scientific method, merely individuals “doing their damndest with their minds, no holds barred” (pp. 144–145). I admire Feynman’s and Bridgman’s definitions of science because neither confuses science with method or technique, as I believe happens in recent government proclamations about the nature of appropriate, and therefore fundable, educational research. World-renowned scientists do

not confuse science with method. As Peter Medawar said, “what passes for scientific methodology is a misrepresentation of what scientists do or ought to do.”

The “evidence-based practices” and “scientific research” mentioned over 100 times in the No Child Left Behind Act of 2001 are code words for randomized experiments, a method of research with which I too am much enamored. But to think that this form of research is the only “scientific” approach to gaining knowledge—the only one that yields trustworthy evidence—reveals a myopic view of science in general and a misunderstanding of educational research in particular. Although strongly supported in Congress, this bill confuses the methods of science with the goals of science. The government seems to be inappropriately diverging from the two definitions of science provided above by confusing a particular method of science with science itself. This is a form of superstitious thinking that is the antithesis of science. Feuer, Towne, and Shavelson, representing the entire NRC committee, clearly recognize this mistake, and we should all hope that they are persuasive. To me, the language in the new bill resembles what one would expect were the government writing standards for bridge building and prescription drugs, where the nature of the underlying science is straightforward and time honored. The bill fails to recognize the unique nature of educational science.

Hard and Soft Science: A Flawed Dichotomy

The distinctions between hard and soft sciences are part of our culture. Physics, chemistry, geology, and so on are often contrasted with the social sciences in general and education in particular. Educational research is considered too soft, squishy, unreliable, and imprecise to rely on as a basis for practice in the same way that other sciences are involved in the design of bridges and electronic circuits, sending rockets to the moon, or developing new drugs. But the important distinction is really not between the hard and the soft sciences. Rather, it is between the hard and the easy sciences. Easy-to-do science is what those in physics, chemistry, geology, and some other fields do. Hard-to-do science is what the social scientists do and, in particular, it is what we educational researchers do. In my estimation, we have the hardest-to-do science of them all! We do our science under conditions that physical scientists find intolerable. We face particular problems and must deal with local conditions that limit generalizations and theory building—problems that are different from those faced by the easier-to-do sciences. Let me explain this by using a set of related examples: The power of context, the ubiquity of interactions, and the problem of “decade by findings” interactions. Although these issues are implicit in the Feuer, Towne, and Shavelson article, the authors do not, in my opinion, place proper emphasis on them.

The Power of Contexts

In education, broad theories and ecological generalizations often fail because they cannot incorporate the enormous number or determine the power of the contexts within which human beings find themselves. That is why the Edison Schools, Success for All, Accelerated Schools, the Coalition of Essential Schools, and other school reform movements have trouble replicating effects from site to site. The decades old Follow-Through study should have taught us about the problems of replication in education (House, Glass, McLean, & Walker, 1978). In that study, over a dozen philosophically different instructional models of early childhood education were implemented in multiple sites over a considerable period of time. Those models were then evaluated for their effects on student achievement. It was found that the variance in student achievement was larger within programs than it was between programs. No program could produce consistency of effects across sites. Each local context was different, requiring differences in programs, personnel, teaching methods, budgets, leadership, and kinds of community support. These huge context effects cause scientists great trouble in trying to understand school life. It is the reason that qualitative inquiry has become so important in educational research. In this hardest-to-do science, educators often need knowledge of the particular—the local—while in the easier-to-do sciences the aim is for more general knowledge. A science that must always be sure the myriad particulars are well understood is harder to build than a science that can focus on the regularities of nature across contexts. The latter kinds of science will always have a better chance to understand, predict, and control the phenomena they study.

Doing science and implementing scientific findings are so difficult in education because humans in schools are embedded in complex and changing networks of social interaction. The participants in those networks have variable power to affect each other from day to day, and the ordinary events of life (a sick child, a messy divorce, a passionate love affair, migraine headaches, hot flashes, a birthday party, alcohol abuse, a new principal, a new child in the classroom, rain that keeps the children from a recess outside the school building) all affect doing science in school settings by limiting the generalizability of educational research findings. Compared to designing bridges and circuits or splitting either atoms or genes, the science to help change schools and classrooms is harder to do because context cannot be controlled.

The Ubiquity of Interactions

Context is of such importance in educational research because of the interactions that abound. The study of classroom teaching, for example, is always about understanding the 10th or 15th order interactions that occur in classrooms. Any teaching behavior interacts with a number of student characteristics, including IQ, socioeconomic status, motivation to learn, and a host of other

factors. Simultaneously, student behavior is interacting with teacher characteristics, such as the teacher's training in the subject taught, conceptions of learning, beliefs about assessment, and even the teacher's personal happiness with life. But it doesn't end there because other variables interact with those just mentioned—the curriculum materials, the socioeconomic status of the community, peer effects in the school, youth employment in the area, and so forth. Moreover, we are not even sure in which directions the influences work, and many surely are reciprocal. Because of the myriad interactions, doing educational science seems very difficult, while science in other fields seems easier.

I am sure were I a physicist or a geologist I would protest arguments from outsiders about how easy their sciences are compared to mine. I know how “messy” their fields appear to insiders, and that arguments about the status of findings and theories within their disciplines can be fierce. But they have more often found regularities in nature across physical contexts while we struggle to find regularities across social contexts. We can make this issue about the complexity we face more concrete by using the research of Helmke (cited in Snow, Corno & Jackson, 1995). Helmke studied students' evaluation anxiety in elementary

and middle school classrooms. In 54 elementary and 39 middle school classrooms, students' scores on questionnaires about evaluation anxiety were correlated with a measure of student achievement. Was there some regularity, some reportable scientific finding? Absolutely. On average, a negative correlation of modest size was found in both elementary and middle school grades. The generalizable

finding was that the higher the scores on the evaluation anxiety questionnaire, the lower the score on the achievement test.

But this simple scientific finding totally misses all of the complexity in the classrooms studied. For example, the negative correlations ran from about $-.80$ to zero, but a few were even positive, as high as $+.45$. So in some classes students' evaluation anxiety was so debilitating that their achievement was drastically lowered, while in other classes the effects were nonexistent. And in a few classes the evaluation anxiety apparently was turned into some productive motivational force and resulted in improved student achievement. There were 93 classroom contexts, 93 different patterns of the relationship between evaluation anxiety and student achievement, and a general scientific conclusion that completely missed the particularities of each classroom situation.

Moreover, the mechanisms through which evaluation anxiety resulted in reduced student achievement appeared to be quite different in the elementary classrooms as compared to the middle school classrooms. It may be stretching a little, but imagine that Newton's third law worked well in both the northern and southern hemispheres—except of course in Italy or New Zealand—and that the explanatory basis for that law was different in the two hemispheres. Such complexity would drive a physicist crazy, but it is a part of the day-to-day world of the educational researcher. Educational researchers have to accept the embedded-

**We do our science
under conditions
that physical scientists
find intolerable.**

ness of educational phenomena in social life, which results in the myriad interactions that complicate our science. As Cronbach once noted, if you acknowledge these kinds of interactions, you have entered into a hall of mirrors, making social science in general, and education in particular, more difficult than some other sciences.

Decade by Findings Interactions

There is still another point about the uniqueness of educational science, the short half-life of our findings. For example, in the 1960s good social science research was done on the origins of achievement motivation among men and women. By the 1970s, as the feminist revolution worked its way through society, all data that described women were completely useless. Social and educational research, as good as it may be at the time it is done, sometimes shows these “decade by findings” interactions. Solid scientific findings in one decade end up of little use in another decade because of changes in the social environment that invalidate the research or render it irrelevant. Other examples come to mind. Changes in conceptions of the competency of young children and the nature of their minds resulted in a constructivist paradigm of learning replacing a behavioral one, making irrelevant entire journals of scientific behavioral findings about educational phenomena. Genetic findings have shifted social views about race, a concept now seen as worthless in both biology and anthropology. So previously accepted social science studies about differences between the races are irrelevant because race, as a basis for classifying people in a research study, is now understood to be socially, not genetically, constructed.

In all three cases, it was not bad science that caused findings to become irrelevant. Changes in the social, cultural, and intellectual environments negated the scientific work in these areas. Decade by findings interactions seem more common in the social sciences and education than they do in other scientific fields of inquiry, making educational science very hard to do.

Conclusions

The remarkable findings, concepts, principles, technology, and theories we have come up with in educational research are a triumph of doing our damndest with our minds. We have conquered enormous complexity. But if we accept that we have unique complexities to deal with, then the orthodox view of science now being put forward by the government is a limited and faulty one. Our science forces us to deal with particular problems, where local knowledge is needed. Therefore, ethnographic research is crucial, as are case studies, survey research, time series, design experiments, action research, and other means to collect reliable evidence for engaging in unfettered argument about education issues. A single method is not what the government should be promoting for educational researchers. It would do better by promoting argument, discourse, and discussion. It is no coincidence that early versions of both democracy and science were invented simultaneously in ancient Greece. Both require the same freedom to argue and question authority, particularly the government.

It is also hard to take seriously the government’s avowed desire for solid scientific evidence when it ignores the solid scientific evidence about the long-term positive effects on student learning of high-quality early childhood education, small class size, and teacher in-service education. Or when it ignores findings about the poor performance of students when they are retained in grade, assigned uncertified teachers or teachers who have out-of-field teaching assignments, or suffer a narrowed curriculum because of high-stakes testing.

Instead of putting its imprimatur on the one method of scientific inquiry to improve education, the government would do far better to build our community of scholars, as recommended in the NRC report. It could do that by sponsoring panels to debate the evidence we have collected from serious scholars using diverse methods. Helping us to do our damndest with our minds by promoting rational debate is likely to improve education more than funding randomized studies with their necessary tradeoff of clarity of findings for completeness of understanding. We should never lose sight of the fact that children and teachers in classrooms are conscious, sentient, and purposive human beings, so no scientific explanation of human behavior could ever be complete. In fact, no unpoetic description of the human condition can ever be complete. When stated this way, we have an argument for heterogeneity in educational scholarship and for convening panels of diverse scholars to help decide what findings are and are not worthy of promoting in our schools.

The present caretakers of our government would be wise to remember Justice Jackson’s 1950 admonition: “It is not the function of our government to keep the citizens from falling into error; it is the function of the citizen to keep the government from falling into error.” Promoting debate on a variety of educational issues among researchers and practitioners with different methodological perspectives would help both our scholars and our government to make fewer errors. Limiting who is funded and who will be invited to those debates is more likely to increase our errors.

REFERENCES

- Bridgman, P. W. (1947). New vistas for intelligence. In E. P. Winger (Ed.), *Physical science and human values*. Princeton, NJ: Princeton University Press.
- Feynman, R. P. (1999). *The pleasure of finding things out*. Cambridge, MA: Perseus.
- House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. (1978). No simple answer: Critique of follow through evaluation. *Harvard Educational Review*, 48, 128–160.
- Snow, R. E., Corno, L., & Jackson, D. (1995). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243–310). New York: Macmillan.

AUTHOR

DAVID C. BERLINER is Regents’ Professor of Education, Arizona State University, Tempe, AZ 85287; berliner@asu.edu. His research interests include the study of teaching, teacher education, and educational policy.

Manuscript received June 4, 2002

Revisions received July 8, 2002

Accepted July 11, 2002