

The Black Hole of Education Research

By D.W. Miller

Why do academic studies play such a minimal role in efforts to improve the schools?

A couple of years ago, the education-policy analyst Diane Ravitch checked into a Manhattan hospital with leg pains and shortness of breath, symptoms of a pulmonary embolism. As doctors gathered around to discuss her ailment, recalls Ms. Ravitch, a fellow at the Manhattan Institute for Policy Research who headed federal education-research efforts in the Bush Administration, “I had this vision: ‘Oh my God — what if, instead of medical researchers, I were being treated by education researchers?’

“I had a fantasy of people disagreeing about how you make a diagnosis. Everyone was saying the tests for my illness were ineffective and they didn’t trust any tests at all. And they were arguing endlessly about whether I was even sick, and I was going to die on the table.”

Her judgment of education researchers may sound harsh, but it is shared by many in the field. There is no shortage of ideas for reforming classrooms and boosting achievement: Reduce class size, raise teacher pay, group students of differing abilities together, end automatic “social” promotion of failing students, and so on. And teachers face a noisy bazaar of curricula, methods, and materials built upon various pedagogic theories. Why don’t we know more about what works in the classroom?

“When policy is made, do people reflexively ask, ‘What does the research say?’” asks Kenji Hakuta, an education professor at Stanford University who heads the U.S. Department of Education’s advisory board on research priorities. “It’s not working the way it should.”

Research on the effectiveness of reforms is often weak, inconclusive, or missing altogether. And even

in areas illuminated by good scholarship, it often has little influence on what happens in the classroom.

All disciplines produce lax or ineffective research, but some academics say that education scholarship is especially lacking in rigor and a practical focus on achievement. “Vast resources going into education research are wasted,” says James W. Guthrie, an education professor at Vanderbilt University who advises state legislators on education policy. “When you look at how the money is channeled, it’s not along productive lines at all.”

To all appearances, the field of education research is robust. The American Educational Research Association has more than 23,000 members; more than 6,000 researchers presented their findings and opinions at its annual meeting last April. Furthermore, in the last 30 years, American universities, foundations, and governments have spent hundreds of millions of dollars for research on K-12 education. Yet test scores on the National Assessment of Educational Progress have been virtually flat since 1970, and a wide gap in scholastic achievement continues to separate rich from poor, and some ethnic and racial groups from others.

The research-to-practice pipeline, according to scholars and educators, has sprung many leaks. Governments and foundations don’t spend enough on research, or they support the wrong kind. Scholars eschew research that shows what works in most schools in favor of studies that observe student behavior and teaching techniques in a classroom or two. They employ weak research methods, write turgid prose, and issue contradictory findings. Educators and policy makers are not trained to separate good research from bad, or they resist findings that challenge cherished beliefs about learning. As a

result, education reform is often shaped by political whim and pedagogic fashion.

Patricia Ann Baltz, a grade-school teacher in Arcadia, Cal., and a member of Mr. Hakuta's panel, says her colleagues have grown cynical about reform. "California is a fad state," she says. "For almost every teacher in my school, their basic perception is 'Okay, we're back to phonics or phonemic awareness or word walls, and in three years, we'll be doing something else.'"

In some cases, no research exists to back up teaching fads. Consider so-called whole-school or comprehensive reforms, the off-the-shelf programs that provide principals and teachers in low-performing schools with a ready-made plan for teaching methods and curricula. They are very popular, in part because they can be purchased with federal Title I subsidies, and they all claim to be based on research.

In 1998, at the behest of national groups representing teachers, principals, and school administrators, the American Institutes for Research took a closer look at the research behind those programs. A.I.R. rated the evidence for the effectiveness of 24 programs on a four-point scale, from "mixed or weak" to "strong."

Only three of the 24 received the highest rating: Direct Instruction, Success for All, and High Schools That Work. Of the eight programs whose research base was "weak" or "marginal," according to the report, three had been adopted by 1,000 or more schools across the country. And seven well-established programs had never been subjected to any rigorous studies.

Robert E. Slavin, a research professor at the Johns Hopkins University and the developer of Success for All, laments the "declining spiral" of research quality. "At the policy level, there is very little respect for research. If there is little value placed on research, then it doesn't get funded, then there's very little research."

In some cases, popular policies thrive despite strong evidence of their harm. A case in point is the widespread move to eliminate automatic promotion of failing students to the next grade. According to a report this year from the National Research Council, empirical evidence on balance shows that failing stu-

dents who are held back perform worse in later years and are more likely to drop out before graduation than are similar students who are promoted. That holds true even when schools offer retained students remedial instruction. Yet legislators across the nation are calling for an end to automatic promotion.

A fundamental problem facing the field, says Ellen Condliffe Lagemann, a professor of education at New York University, is its fragmentation among many disparate disciplines, including statistics, anthropology, and cognitive psychology. In education, she says, "there are no common patterns of training. If you don't have common patterns of training, it's hard to reach agreement on what research is, much less what good research is."

That weakness is evident in the growing debate over rigor and methodology. Until the 1970s, education research was dominated by cognitive psychologists performing laboratory studies and conducting surveys on how students learned. Since then, more and more researchers have favored the qualitative techniques of anthropology, writing descriptive case studies and narratives about classroom activity based on observations. At the same time, scholars with a more statistical bent -- many of them social scientists from outside of education schools -- have sifted reams of test data to identify the factors that correlate with students' success.

In general, most scholars agree that quantitative and qualitative inquiries are complementary. Quantitative researchers create experiments or look for statistical correlations in large data sets to see which factors and policies boost achievement. Qualitative observation can both unearth useful theories for quantitative inquiry and help explain things that statistics cannot, such as why a particular reform works, and in which circumstances.

In practice, some scholars believe, education has not been well served by such varied methods of research. "We've made enormous advances in the cognitive sciences," says Ronald Gallimore, an education professor at the University of California at Los Angeles. Yet "very little work has been done on the implementation issue" -- translating knowledge about learning into effective methods and materials.

Qualitative research has been criticized for yielding little that can be generalized beyond the classrooms in which it is conducted. Even Mr. Gallimore, a leading defender, says that too much useless work is done under the banner of qualitative research. At the April meeting of the education association, he urged his fellow researchers to set high standards for rigor, lest the qualitative enterprise lose all credibility with congressional appropriators.

A growing number of quantitative scholars want to revive a research tradition, prevalent in the 1960s and 1970s, of evaluating classroom practices through randomized tests. As Ms. Ravitch's hospital musings suggest, many of them see the clinical trials of medical research as a model for education scholarship. When a drug company wants to test the benefits of treating some new pill, they reason, it finds a group of volunteer patients and compares them to an identical control group. "Why is education research any different?" asks Paul E. Peterson, a political scientist at Harvard's Kennedy School of Government.

Yet such trials, routine in the sciences, are relatively rare in education research, according to Thomas D. Cook, a sociologist at Northwestern University. He recently found that such experiments have hardly been used to examine the effectiveness of common reforms, including vouchers, charter schools, whole-school reform, and continual teacher training.

A task force sponsored by the American Academy of Arts and Sciences is urging researchers to increase the use of random sampling and control groups. Without them, advocates say, no researcher will ever be able to attribute a successful outcome to a particular policy. Any other method leaves doubt about alternative explanations, such as selection bias.

The debate over school vouchers is one example. In recent years, several thousand students in failing public schools have used government-financed vouchers or private scholarships to attend private schools. But skeptics have pointed out that it is hard to assess whether vouchers actually improve achievement, because those self-selecting students who use them may be smarter or more motivated than those who do not.

Mr. Peterson was able to use random sampling in a recent study of privately financed vouchers in New York because the program was oversubscribed, and vouchers were awarded by lottery. He found that, after a year or two in private schools, voucher recipients did score better on achievement tests than the unsuccessful applicants who stayed in public schools.

Some scholars say that the experimental approach is resisted for ideological reasons. "The whole field of education research is dominated by an orthodoxy that the best kind of learning is the kind that spontaneously emerges," says John E. Stone, an education professor at East Tennessee State University. The so-called learner-centered approach, he says, assumes that learning will occur when "conditions are right," so qualitative researchers account for differences in achievement by examining the differences between one classroom and the next.

Many scholars, financing agencies, and journal editors, he says, would prefer to ignore experimental research, which tends to show that traditional, directed instruction, with drills, lectures, and step-by-step lesson plans, works better than more-spontaneous approaches. By example, he points to a \$1-billion federal evaluation in the 1970s of various Great Society programs designed to help low-income students. After most approaches flunked and a method called Direct Instruction came out on top, he says, the education field essentially ignored the results.

Even quantitative researchers differ over the relative worth of randomized experiments. The prime exhibit for the method is a Tennessee study from the late 1980s. That research appeared to show that reducing classes to 15 to 17 students in the early grades raises achievement, particularly among low-income students.

Yet when Eric A. Hanushek, an economist at the University of Rochester, testified before Congress last June about research on class size, he gave the Tennessee experiment alone less weight than his statistical analysis of 277 (mostly non-experimental) studies on the subject. His conclusion: There is no consistent connection between class size and performance.

Critics say that such studies must be fairly large to be useful, and hence are quite expensive. The four-year Tennessee program, which was paid for by the

state legislature, was indeed pricey — about \$3-million a year for conducting research and hiring more teachers. But, say advocates of randomized trials, untested reforms will cost states billions of dollars in new classrooms and teachers.

Mr. Gallimore, the U.C.L.A. professor, is skeptical of relying on the clinical model for research. “Even in the biggest clinical survey, you get unexpected results,” he says. “If you don’t do qualitative research, you won’t understand the classroom.” And, he says, “finding out why it doesn’t work in one school is sometimes more important” than confirming a program’s general effectiveness. “That’s why the clinical model breaks down.”

Some scholars contend that education research can boast plenty of solid, useful findings about learning and reform. The problem is not that the research isn’t good, they say, but that it doesn’t find its way into the classroom.

“There is such a chasm that separates research from the field. My colleagues don’t have a clue what’s being done in the research field. They don’t know what an education researcher is,” says Ms. Baltz, the California teacher. “Teachers never get the benefit of research to pass on to their kids.”

At the nexus of most of those complaints lies the sprawling network of U.S. professional schools that train teachers. Those institutions and departments, of which there are more than 1,700, are responsible for cultivating many of the nation’s education researchers as well as for training most of the public-school educators for whom research is intended. Officials at education schools readily acknowledge that they do a poor job of training educators to distinguish good research from bad, and also of training their scholars to make their findings accessible to practitioners.

“Researchers need to be bilingual,” says Arthur Levine, the president of Teachers College at Columbia University. “They need to speak both to their colleagues and the public.”

But that language barrier may be less daunting than disharmony over basic principles. “There is no agreement on what the purpose is of educational research, or on the purpose of ed schools,” says Mr. Levine. In his view, his field’s whole scholarly enter-

prise should be focused on what works in the classroom. “Rome is burning, and this is no time for good violin research.”

Critics who say they want higher standards of rigor and a new focus on what works in the classroom point to encouraging signs. As the president of the National Academy of Education, a group of about 120 prominent education scholars, Ms. Lagemann, of N.Y.U., is co-chairing a project to create common goals and standards for research. And, recognizing the need for synthesis in a chaotic discipline, the National Research Council has recently published books summarizing research findings on pressing policy issues such as early reading instruction, bilingual education, and appropriate uses of achievement tests.

The council’s panel of scholars wrote *Preventing Reading Difficulties in Young Children* (1998), for example, to end the battle between educators who prefer traditional, phonetic reading instruction and “whole-language” advocates who believe that children should be taught to recognize whole words in the context of engaging, meaningful literature. The panel concluded that the best approach to teaching literacy combines instruction in phonemic awareness with language-rich materials.

The federal government also seems to be demanding more of the research community, using its financial leverage to funnel rigorous research to the classroom.

In 1997, in a law known as Obey-Porter for its sponsors, Congress enacted the Comprehensive School Reform Demonstration Program. It offers poor schools \$145-million for the adoption of whole-school programs that are based on empirical, replicable, peer-reviewed research. A similar law, the Reading Excellence Act of 1998, provides \$260-million for early-childhood literacy programs shown to be effective by rigorous scholarship.

The advisory panel headed by Stanford’s Mr. Hakuta recently urged the federal government to focus its research budget on improving achievement, and to cede its role in approving grants to review panels of top scholars. Peer review was adopted long ago by science agencies such as the National Science

Foundation and would, say scholars, improve the quality and relevance of such projects.

Other scholars suggest that the demand for rigorous and relevant research will rise only when educators actually face consequences for classroom failure. They are heartened by the recent moves by policy makers to set curricular standards, test students' mastery of those standards, and hold educators accountable.

Gary Galluzzo, the dean of George Mason University's education school, wonders, however, whether researchers will ever be able to provide classrooms with a general blueprint. "We haven't agreed on the key issue of what the purpose of schools is. Schools are trying to be all things to all people at all times." The new push for accountability, he says, rests on a disputed definition of achievement: performance on standardized tests of basic skills. But public

schools are also expected to mold citizens, teach practical skills and habits for adulthood, instill a capacity for critical thinking, overcome the opportunity gap for poor students, and more.

Eventually, he believes, the ideal of a common public-school pedagogy will give way to a marketplace catering to every preference. "I don't think that's in the best interest of the country, but that's where we are. I just hope we don't lose our focus on the least advantaged."

<http://chronicle.com>

Section: Research & Publishing Page: A17

Copyright © 1999 by The Chronicle of Higher Education