



HENRY PAYNE

Why Johnny Can't Teach

by Jonathan Marshall

**Without knowing what works in the classroom,
American educators have become fad followers.**

It's a giant industry, one that most Americans consider vital to their country's future competitiveness. Yet it spends less than 1 percent of its \$375-billion annual gross revenue on research and development—and much of that is squandered by reliance on shoddy, unscientific methods. No wonder this industry is deeply troubled and in need of fundamental reform.

Public education—the industry in question—still uses much the same methods it did a century ago. One reason is the failure of educators to learn better ways of doing things. Bad research, underfunded research, and lack of official interest in good research have all crippled the ability of this industry to show improved results over time.

In a rare public acknowledgment of the research gap in education, the state of California's nonpartisan legislative analyst earlier this year declared that because of "severe methodological problems" afflicting nearly all evaluations of state-sponsored instructional programs, "educators simply do not know how well most programs address the problem for which they were created."

Last year the prestigious National Research Council of the National Academy of Sciences issued a scathing report condemning the field's penchant for "methodologically weak research, trivial studies, an infatuation with

jargon, and a tendency toward fads." Without "high-quality and credible evaluations," it warned, "school districts will never be able to choose wisely among available innovations." And even when scientifically valid research is available, the report added, teachers, administrators, policy makers, and parents often ignore it.

No other industry would last long with such a haphazard approach to self-improvement. But in education, notes Diane Ravitch, a visiting fellow at the Brookings Institution and former assistant secretary of education for research, "there are no consequences for failure. It is a public monopoly like the Post Office. Whether you are good or bad, it will be funded. There is no bottom line."

Flying blind without good research guidance has produced many an educational crash, charges Donald Orlich, a researcher at Washington State University in Pullman, Washington. "This nation has wasted billions of dollars on poorly conceived but politically popular reform movements that have sapped the energies of school people," he says.

Billions of dollars more stand to be wasted unless things change. As Albert Shanker, president of the American Federation of Teachers, warned in 1988, "Without good research, we will continue on an endless cycle of mistakes and the loss of successful insights and

discoveries. Without good research, there will continue to be an endless invention of mousetraps, the same rehashing of controversies, and, in the end, the same faltering school system."

Robert Slavin, who runs a distinguished educational research facility at Johns Hopkins University, has observed that the typical educational innovation starts with a burst of enthusiasm, followed by "widespread dissemination, subsequent disappointment, and eventual decline—the classic swing of the pendulum." He says education resembles a progressive science less than it does the fashion and design industries, which gyrate according to fads and changing tastes. Gullible principals, school boards, and even state legislatures too often jump on the latest educational bandwagons, led by charismatic proselytizers who promote their programs with unsupported or anecdotal claims.

Slavin cites as one example the highly popular Instructional Theory into Practice, or Madeline Hunter model, which emphasizes the need for clear objectives, careful control of classroom time, and frequent assessment of student understanding—all sensible, but hardly revolutionary, steps. The technique did not receive a large-scale evaluation until 19 years after it started sweeping the nation. The study, based in South Carolina, found no significant improvement in performance by students of teachers trained in the model. Studies in New Jersey and in Napa, California, produced the same result. "You could get the same effect by throwing a pizza party," says Orlich, another critic. Yet the model continues to enjoy favor in many states.

Another example is the "open classroom," which was all the rage two decades ago. Schools were built (or redesigned) without walls, and students moved flexibly within classes between "learning stations," supposedly guided in part by their natural curiosity and sense of direction. "This was the hottest thing of its time in the early 1970s," Slavin says, "but research found consistently it didn't have the effect claimed."

The "whole-language movement" in reading is one of the latest fads to sweep the nation. Deriding the phonics method, the movement rests on an intuition that children will learn to read naturally when exposed to books and other reading materials. "It is not only not validated by any research whatsoever, but carries with it a philosophy opposed to evaluation," Slavin complains. "It may or may not be a good idea, but the extraordinary diffusion of this method from coast to coast without a shred of evidence is terrifying."

Besides cheating students, the cycle of high promises and dashed hopes often burns teachers out. When a really good pro-

The "whole-language movement," the latest educational fad, assumes children will learn to read naturally when exposed to books. "It is not only not validated by any research whatsoever, but carries with it a philosophy opposed to evaluation," warns one researcher. "It may or may not be a good idea, but the extraordinary diffusion of this method from coast to coast without a shred of evidence is terrifying."

gram comes along, they may be reluctant to give it a try. "They have learned that the present innovation will be gone in a year," notes Thomas Guskey, an educational researcher at the University of Kentucky. "In fact, it is not unusual to hear teachers refer to the staff development program topic of the moment as TYNT, for This Year's New Thing. And cynics know, of course, that TYNT is bound to be different from LYNT, which was Last Year's New Thing."

Jennifer Schindler, vice principal and teacher at El Vista Elementary School in Modesto, a city with large numbers of poor Hispanic and Asian immigrants in California's Central Valley, is all too familiar with this pattern. She remembers wearily how schools moved from a "touchy-feely, let's-talk-about-our-problems" ap-

proach in the 1960s, to open classrooms in the mid-'70s, to back-to-the-basics in the early '80s, and then, in step with the rest of the nation, to the whole-language movement.

"We did the whole-language approach for a couple of years and didn't see any results," Schindler says. "The whole-language people say you shouldn't put structure in teaching, but our kids don't have a lot of structure at home.... They flounder and wonder what to do next."

After years of floundering themselves, teachers at El Vista finally agreed to try a new approach developed by Slavin's team at Johns Hopkins for teaching reading and writing to grade-school children from disadvantaged backgrounds. Based on experimental research, the program, called "Success for All," combines several proven approaches, including "mastery learning," a method of using frequent assessments and individual tutoring to prevent slippage by slower students in the class, and "cooperative learning," which makes small groups responsible for individual mastery of subjects.

Teachers at the school get lots of flack from ideologues who teach the latest educational fads at local colleges, Schindler says, so "we have to defend ourselves against the trends." Fortunately, their results are defense enough. "Our program is probably not the answer to all the world's problems, but all of our children read, and our first-graders are writing really well. Teachers with long experience say they are seeing a big improvement." Jerry Fry, program director for the entire district, says "this is the first time in my career that I've been in something that prevents failure. It's magic. It takes work, but it works."

Slavin devised his program based on years of careful evaluation of teaching programs. A stickler for rigorous research design, he won plaudits from the National Research Council last year for using systematic methods "not un-

common in the natural sciences, but...rare in education research and development."

The best evaluations of teaching methods, Slavin argues, have much in common with the experimental design used by medical researchers in testing a new drug. Teachers and students are selected randomly for an "experimental group," which uses the new approach, and a "control group." Both groups are tested before and after implementation of the new method. Ideally, testing continues some years out to see whether any improvements stick or fade with time.

Howard Bloom, an economist at New York University, calls such experimental design "the most powerful existing methodology for measuring the impacts of social programs." Random assignment ensures that factors such as age, education, and race do not bias the results. Of even greater importance, he says, it ensures that experimental and control groups are "comparable in terms of unmeasured factors such as motivation, intelligence, and emotional stability. Therefore, any subsequent differences between outcomes for these groups can be attributed to differences in the treatments to which they were exposed."

Randomized tests are not a panacea. As in all evaluations, there must be some objective criteria of success and a sufficiently long testing period to give the program a real tryout. Political problems sometimes stand in the way. If a new program seems especially promising, says Ricky Takai, a senior Department of Education official, parents and teachers often resent being left behind in the control group, at least until administrators convince them that selection by lottery is fairer than any other method when experimental slots are limited. Another limitation is that the test may indicate only that a program works, but not which of its components count the most or why. Close field observation of actual classrooms is needed to supplement and interpret the results.

Slavin says much educational research does not even come close to these standards. Evaluations are often conducted by the program developer, who tests under optimal conditions by selecting the most enthusiastic teachers and best-motivated students for the program. No comparison groups are used, much less randomly selected control groups; instead classes are simply tested at the beginning and end of the year, and gains in grades or test scores are attributed to the program. Research assumptions and limitations are often poorly documented or ig-



Teacher Jennifer Schindler uses the "Success for All" learning program, developed by Johns Hopkins University education researcher Robert Slavin, as she teaches her second-grade students to recognize new words with the letter "d" in them.

nored altogether. And results are not replicated in other settings before the developers begin beating the drum for the latest fad.

The National Diffusion Network, a government-funded clearinghouse that informs states and local districts of promising new educational methods, seeks to weed out the worst of these research claims. Even so, "their standards are still very low," Slavin charges. "Few of the programs or reports [they endorse] had control groups. There are some 500 projects in the book that are listed as

being effective; let me assure you there aren't 500 methods that really are effective."

If anything, interest in the use of scientific research methods in education is waning, not growing. Instead of investing in large-scale, long-term evaluations of classroom teaching methods, most research today favors impressionistic studies of individual classrooms and teachers. "You have an absurd movement to anecdotal, anthropological studies of classrooms," says Herbert Walberg, a renowned educational researcher at the Chicago campus of the University of Illinois. "In my view it's almost anti-science. But two-thirds of the members of the American Educational Research Association would disagree with me."

Funding is also scarce for really good field research. Most federal research money goes to regional research centers that disseminate information rather than oversee careful experiments.

And yet it is sheer folly not to invest the money to find out what works. The federal government pumps more than \$6 billion a year into so-called Chapter I funds, which aid school districts with significant numbers of "disadvantaged" students. One of the chief ways local districts use the money is to reduce class sizes—just about the most expensive possible intervention given the cost of hiring extra teachers and building more facilities. Yet strong teacher lobbies with a stake in new hiring promote class-size reduction as the answer to America's educational needs. As Keith Geiger, president of the National Education Association, declared, "If we're serious about improving learning in America, there's no more important place to begin."

But does it work? For years nearly everyone had an opinion, but nobody really knew until recently because hardly any systematic tests had ever been done to answer the question. In-

**Even the best research on
teaching methods often fails
to decide the most critical issue
facing school districts: cost-
effectiveness. When funds are tight,
schools cannot necessarily afford
to opt for programs that offer
the most performance gains for
students. When it comes to picking
from a menu of worthy options,
school districts have little
to go on.**

tuitively, it seems obvious that smaller classes should help, yet Japanese students manage to excel in mathematics despite class sizes in the low 40s.

In the mid-'80s, the Tennessee legislature decided it needed a definitive answer. With help from researchers in the state university system, it appropriated \$12 million to carry out a bullet-proof test. (The actual research cost less than \$1 million; the rest paid for smaller classes.) Known as Project STAR, the study took 7,500 students in grades K-3 and assigned them randomly to three types of classes: normal ones with 23 kids and a single teacher; normal-size classes that included a teacher's aide; and classes with only 15 children per teacher. Teachers were also assigned randomly to avoid bias. Careful, consistent testing tracked the children through these classes and into later years.

The results of Project STAR were fascinating and instructive: Students who attended smaller classes made significant cognitive gains in all subjects, proving for the first time that smaller classes really do aid learning. (In contrast, teacher's aides did not help academic performance at all.) At the same time, however, the performance gains were modest—well below those achieved by several proven teaching methods, such as mastery learning and cooperative learning, that work well in normal-size classes.

As John Folger, professor emeritus at Vanderbilt University, concluded in a review of Project STAR, "the high cost of reducing class size across the board makes it unfeasible. There are other interventions which produce much larger improvements in student achievement for the same or lower costs than would be involved in a substantial reduction in class size." Folger cited Slavin's Success for All as a reading program that produces three to five times the gains of reducing class size.

Praise for the careful experimental design of Project STAR has been almost universal. Orlich calls it "the most significant educational research done in the U.S. during the past 25 years." Slavin, who calls it "an extraordinary experiment that makes most previous research on class size obsolete," notes the irony that "we are willing to spend massive sums on educational services of unknown effect but find it shocking to spend \$12 million to find answers to questions at the top of people's list of what we need to know."

Such investments in high-quality research have time and again proven their ability to change the terms of social debates. In education, for example, few studies have been more influential than one of about 50 poor, black children who attended the Perry Preschool in Ypsilanti, Michigan, in the early '60s. Using a randomized design, the experiment

showed that disadvantaged kids who attended a high-quality preschool achieved more later in school and in adulthood than similar kids who did not attend the preschool.

Experimental tests of social interventions have been even more common and influential in the area of job training and welfare-to-work programs, such as California's GAIN initiative and the national Job Training Partnership Act. Title II of the 1988 Family Support Act, which promotes job search and training programs for welfare recipients, passed with bipartisan support thanks to randomized, experimental studies of successful welfare-to-work programs conducted by the private, nonprofit Manpower Demonstration Research Corporation. "MDRC's studies really changed the tenor of the debate," says Erica

Baum, former legislative aide to Sen. Daniel Moynihan, the New York Democrat who authored the bill. "Everybody accepted its data as gospel because it was the most rigorous social science to date and showed that some things did work."

Norton Grubb, an educational economist at the University of California, Berkeley, observes that millions of dollars of government and foundation dollars have supported randomized experiments in AFDC and job training without trickling down to education, a much larger field. "Education has had very little sophisticated research going on in the last 15 years," he says. "There is really an imbalance that we need to rectify."

California's legislative analyst agrees. In a recent report the office complained that "while the Legislature has spent millions of dollars on program evaluations in education, the state has little to show for these expenditures." As a "critical ingredient" of ensuring better evaluations, the report said, the state should "require use of randomly selected control groups. Almost no education programs use this evaluation design. Without the use of randomized control groups, it is very difficult to accurately measure the impact of services. This design also permits measurement of longterm impacts at a relatively low cost."

California's legislature seems to have gotten the message. Last year it authorized just such a study of a promising but unproven high-school program of "career academies" that aims to integrate academic and vocational education and keep potential dropouts in school. MDRC has been selected to run the study at school sites in California and several other states, using its expertise in random-assignment experiments. The organization has to raise its own funding, mostly from foundations.

But even the best research on teaching methods often fails to decide the most critical issue facing school districts: cost-effectiveness. When funds are tight, schools cannot necessar-

ily afford to opt for programs that offer the most performance gains. When it comes to picking from a menu of worthy options, districts have little to go on. "Most evaluations neglect to consider the costs of potential interventions," noted an important 1984 paper on the issue by Stanford University's Henry Levin, Gene Glass, and Gail Meister. "The result of these gaps in information is that there is little to guide policy-makers or school districts in choosing among school reforms that will account for both the costs and effects of educational interventions."

Their nine-year-old study of the "Cost Effectiveness of Four Educational Interventions" appears to be the only one of its kind in the entire field, according to many researchers. Its results were eye-opening, although not definitive: "Peer tutoring" (tutoring of young children by older ones) was the most cost-effective approach to improving math and reading. Computer-aided instruction came next, followed by reduced class size and, at the very bottom, longer school days.

Why are the most promising research approaches so often ignored or abandoned, given the stakes? One fundamental reason is that offered by Ravitch, that the public-school monopoly is ultimately not accountable to consumers for performance. Sheltered from competition by the high cost of private schooling, protected from parental wrath by layers of bureaucracy, and answerable mainly to organized interest groups such as teacher and custodial unions, public schools simply have no systematic stake in perfecting themselves. On the contrary, they have a strong incentive to avoid evaluations that might help parents make informed choices, pinpoint flaws, and demand institutional changes.

Yet this is far from the only reason. Most teachers, public as well as private, would like to do better but don't know how. And many private schools are as backward as their public counterparts in teaching methods.

John Goodlad, who runs a research and development center at the University of Washington specializing in teacher education, notes that most teacher-education programs are simply too short to expose future practitioners to the latest research, much less to the methodological principles needed to evaluate research claims. "For the most part, teacher-education programs have been too short or limited to build research into them, unlike four-year medical programs." Although Goodlad is



Two second graders at El Vista Elementary School engage in "partner reading."
In the activity, part of the "Success for All" program, students help each other practice reading by reading aloud to each other one page at a time.

too polite to say so, many teacher trainees are among the least capable of college graduates and are thus less equipped to grapple with the intellectual challenges of scientific research. Teachers who emerge ignorant of research strategies may, years later, become administrators ignorant of what good research has to offer and which bad research to avoid.

Compounding the problem is the limited interest of many politicians, whose lives are divided into two-year chunks, in supporting long-term studies. They would rather cut ribbons

and enact bold new programs than fund studies whose results may not show up during their terms in office. Their short time horizon renders much of the research they do fund worthless. Audrey Pendleton, who analyzes dropout prevention programs for the U.S. Department of Education, complains that most research grants last only three years and thus track students only through graduation—not long enough to find out whether the programs have any measurable effect on future earnings or job prospects.

Finally, educational research is inevitably crimped by the lack of any fundamental agreement on educational goals. Should schools be measured by student performance on standardized tests, by more subjective tests that aim to capture "critical thinking" skills, by their ability to turn young people into good citizens, or by how well they harmonize students of diverse ethnic and social backgrounds? The goals of education are a rapidly moving target. If members of the public cannot even agree on what constitutes a good school, researchers will be hard pressed to say what produces good teaching.

Yet taxpayers who spend nearly \$400 billion a year on education deserve more hard answers to questions that have for decades sparked only ideological debates over educational reform. Surely they deserve more than one solitary, dated study on the relative cost-effectiveness of alternative teaching methods. If nothing else, they deserve some assurance that educators and policy makers are capable of learning. No one says finding and implementing better educational methods will be easy. But the nation's failure to make a more serious effort is surely a grave indictment of its commitment to quality education. ■

Contributing Editor Jonathan Marshall is economics editor of the San Francisco Chronicle.



Know Thy Enemy

In past book issues, REASON has asked people to recommend edifying books—books that are significant because of the good they can help accomplish. But a well-rounded person should be familiar with pernicious books as well. So this year we asked each contributor to suggest a book published in the last 50 years that is significant because it has helped promote wrongheaded ideas with serious consequences. We suggested that the contributors might want to recommend antidote books as well. Here are their responses.

ARTHUR MATSON