

NO CHILD LEFT BEHIND

NO CHILD LEFT BEHIND; Does It Work

By James Traub

- Nov. 10, 2002

See the article in its original context from November 10, 2002, Section A, Page 4

JOURNALISTS using the most exacting method available to social science -- that is, counting -- have determined that the phrase "scientifically based research" occurs more than 100 times in the Bush administration's No Child Left Behind Act of 2001. The government, for example, will award \$5 billion over six years to states and districts where reading is taught using "scientifically based" methods. Though No Child Left Behind is perhaps best known for requiring every state to test annually in English and math, its passage may ultimately be recalled as the moment when education came to be treated more like medicine -- a science that advances according to the findings of impartial research -- than moral philosophy or folk wisdom.

The idea that pedagogy ought to aspire to the condition of science, or even social science, is quite novel, and it runs against the grain of mainstream educational culture. As Grover J. Whitehurst, assistant secretary for research and improvement at the Department of Education, says, "Education has not relied very much on evidence, whether in regard to how to train teachers, what sort of curriculum to use or what sort of teaching methods to use. The decisions have been based on professional wisdom or the spirit of the moment rather than on research."

Last month, Congress passed legislation replacing the Office of Education, Research and Improvement, which has been widely criticized for being too easily sold on fads, with a more independent institute intended to foster a new culture of rigorous research. Dr. Whitehurst, a psychologist who has researched the effectiveness of the Head Start program, will lead the new Institute of Education Services. He is also in the process of setting up the What Works Clearinghouse, a body that will establish standards for research and then determine which of thousands of studies on class-size reduction, peer tutoring, reading instruction and so on meet those standards. There is giddy talk in the research world of some day establishing the equivalent of the Food and Drug Administration, declaring educational doctrines safe and effective, or not.

The history of educational research is not necessarily encouraging to those who foresee a golden age of scientific clarity. Historians like Diane Ravitch have shown that American public schools have been battlefields of conflicting doctrine since their founding in the mid-19th century, but until recently those doctrines were too inchoate and varied too much from place to place for rigorous comparison. All that began to change in the

1960's, when the Johnson administration made the education of the poor one of the great social experiments of the War on Poverty. Both Head Start and the so-called Title I program were intended, among other things, to improve the academic performance of disadvantaged children. The federal government sponsored extensive research into the effectiveness of both, but the findings were painfully disappointing.

"The evidence came back that nothing worked," says Thomas D. Cook, a professor of sociology at Northwestern University and a veteran of educational research. And if nothing schools did helped failing children, he adds, there was not much point in further study. The moral drawn by people in the field, especially by professors of education, was that schools were such complex and singular institutions that reforms in the mass were unlikely to work. And so scholars turned away from such top-down programs in favor of addressing the individual school, as if, Dr. Cook says, they were so many management consultants.

But the problem was not only that nothing worked -- the wrong thing worked. In 1968, the federal Office of Education commissioned a multiyear, \$500 million study to compare competing approaches to teaching basic skills in the early grades -- the first attempt to see what light "scientifically based research" could shed on different teaching methods. The Follow Through study, as it was called, ultimately involved nine approaches ranging from a highly progressive Open School model to an extremely structured design called Direct Instruction. The results, published in 1977, were stunning: only Direct Instruction significantly raised scores of third graders on a series of achievement tests. Children exposed to more progressive models did far worse than children at "control" schools. Direct Instruction was thus the first research-proven pedagogy.

But Direct Instruction, which involved breaking skills down to their smallest cognitive units and then teaching each subskill explicitly and repetitively, was wildly unpopular among educators, who found it almost robotic; rather than promote the one method and discourage the others, the Office of Education, absurdly, decided to certify most of the approaches as effective. At the time, scholars said that the entire study was methodologically flawed, though subsequent research tended to confirm the findings. "Apparently," the developer of Direct Instruction, Siegfried Engelmann, later bitterly wrote, "decision makers had a greater investment in romantic notions about children than in the gritty details of actual practice or the fact that some things work well."

Here one comes to a crucial distinction between education and medicine: in education, a priori beliefs about the way children ought to learn or about the relative value of different kinds of knowledge seem to have tremendous force in shaping judgments about effectiveness. Direct Instruction could not be deemed uniquely effective because, according to the progressive model then widely embraced, it wasn't supposed to be effective.

One can still find fierce rebuttals of the Follow Through results in contemporary defenses of progressive education like Alfie Kohn's "Schools Our Children Deserve." Nor is it simply a matter of disputing results. Many progressive educators, including

influential figures like Howard Gardner of Harvard University, argue that education should lead to forms of deep understanding that cannot be properly measured by standardized tests. (This is what Mr. Engelmann meant by "romantic notions about children.") In other words, they don't accept the very premise of the What Works Clearinghouse, or for that matter of No Child Left Behind. This is a problem that the F.D.A. does not have to contend with.

And so it is probably safe to say that ideology has done as much to retard the rise of scientifically based research as the skittishness of researchers themselves. The pattern of traditional teaching methods faring better in rigorous comparisons than more open-ended ones, and then of the open-ended ones flourishing nevertheless, has repeated itself many times over. The best-known instance of the phenomenon is probably the endless battle between the proponents of phonics and of whole language instruction. Virtually every impartial effort to analyze the hundreds of studies on the subject, most recently by the Reading Panel of the National Institutes of Health, has found that the step-by-step approach of phonics is more effective, especially with poor children. But phonics, which the Bush administration unabashedly promotes, is still a four-letter word in progressive reaches of the educational world, where it is widely held that children can learn to read through immersion in language rather than through memorizing letters and sounds.

No Child Left Behind may mark a turning point in this battle between educational folk wisdom and social science. From now on, classroom practices will have to "work" to gain wide acceptance (and federal grants), and the criteria for "work" will be explicitly defined. Education researchers are increasingly turning to what is known in medicine as the randomized-control trial, in which large populations of patients are randomly assigned to receive either the treatment being tested or whatever they would have received otherwise. The F.D.A. will not normally approve new drugs or therapies without evidence from such large-scale experiments. Randomly assigned trials are still unusual in education, a field where, for perfectly good reasons, the treatment comes first and the study later. Even a study as sophisticated as Follow Through did not involve random assignment, and for that reason has been dismissed by some social scientists. But such experiments are now carried out on a wide range of educational practices, including teacher development, peer tutoring and voucher programs.

Probably the most highly regarded of all educational studies is the experiment in class-size reduction begun in 1985 by the state of Tennessee, which offered schools extra funds if they agreed to randomly assign teachers and students to classes of 15 or 22 children. Students in smaller classes enjoyed significant gains in reading and math scores, though gains faded over time. And because the experiment was carried out with such care, the merits of small classes are now widely accepted. Frederick Mosteller, a renowned statistician at Harvard, compared the study to the Salk vaccine trials as a seminal moment in the history of research.

And yet while the polio vaccine trials led to a universal treatment, the class-size trials certainly have not. California was moved by Tennessee's experience to spend billions of dollars hiring new teachers, but the results were much less impressive than Tennessee's. Partisans of reducing class size blame California's implementation -- too many incompetent teachers hired too fast -- but it may also be that small classes on their own are a less effective treatment than the polio vaccine. Here, too, ideology intervenes, because liberals point to the benefits of reducing class size as proof that more money should be spent on schools, while conservatives who think that money is not the answer point to California's experience. Eric A. Hanushek, a senior fellow at the Hoover Institution at Stanford University and a leading critic of additional school spending, recently wrote, "Despite the political popularity of overall class-size reduction, the scientific support for such policies is weak to nonexistent."

In an essay in January for the Hoover Institution, E.D. Hirsch Jr., founder of Core Knowledge, a whole-school reform model that re-envisioned curriculum and structure, wrote that the continuing debate over class-size reduction as well as over whole-school reforms showed that even the most rigorously designed experiments would never cut the Gordian knots. Virtually no study, he wrote, offers a plausible account of why a particular practice does or doesn't raise student achievement, so scholars cannot draw a firm line from specific findings to the reform. And so many variables go into learning, he suggested, that classroom data is inherently ungeneralizable.

It may be that the wish for an objective answer to the question of what works in education is a will-o'-the-wisp. It may not be possible to remove the elements of subjectivity, of values, of differing conceptions of the good, which make education different from medicine (itself not value-free, of course). Why, in fact, would anyone wish to? Education is an ineluctably moral act. At the same time, we could attain a great deal more clarity than we have now about the effectiveness of the torrent of practices and theories washing all around us. Perhaps we will simply have to accept the fact that research will help us decide what is best but will never make those decisions for us.