TALKS TO TEACHERS

A FESTSCHRIFT FOR N. L. GAGE

EDITED BY

DAVID C. BERLINER
BARAK V. ROSENSHINE

RANDOM HOUSE NEW YORK

This book was developed for Random House by Lane Akers Incorporated

CHAPTER 1

DAVID C. BERLINER

KNOWLEDGE IS POWER

A Talk to Teachers About a Revolution in the Teaching Profession

It is almost 100 years since the philosopher, psychologist, and consultant to educators William James talked to teachers about "psychology and the teaching art" (1983). Beginning in 1891, he regularly delivered memorable lectures to the teachers in the Boston area. In a moment of great insight he uttered words that stood unchallenged for most of these past 100 years:

> You make a great, a very great mistake, if you think that psychology, being the science of mind's laws, is something from which you can deduce definite programmes and schemes and methods of instruction for immediate schoolroom use. Psychology is a science, and teaching is an art: and sciences never generate arts directly out of themselves. An intermediary inventive mind must make the application, by using its originality. (p. 15)

I think that in the last few years we have come closer than ever before to providing direct scientific underpinnings for the art of teaching. In some cases, the need for highly inventive, creative minds has been lessened, as research provides ideas and technology that are *almost* directly applicable to classroom life.

A hybrid form of research has developed that relies on more than just laboratory findings from psychological science. This new research is broad in scope, using concepts and methodologies from all the humanities, as well as from all the social and behavioral sciences. The focus of these diverse approaches is teaching and learning as they occur in schools and classrooms. This relatively new research has provided findings, concepts, and technology undreamed of by William James. I believe James would now look a little more kindly on the relationship between art and science and would see what we see—a genuine transformation in the profession of teaching as a scientific basis is developed for the art of teaching. No one I know denies the artistic component to teaching. I now think, however, that such artistry should be research based. I view medicine as an art, but I recognize that without its close ties to science it would be without success, status, or power in our society. Teaching, like medicine, is an art that also can be greatly enhanced by developing a close relationship to science. Let us explore these ideas about research and teaching by examining the knowledge base created in the past few years. I will start by examining one source of professional status in our society—professional knowledge.

KNOWLEDGE AND THE STATUS OF PROFESSIONS

Late in the last century, only a little over 100 years ago, there were three equally disreputable professions: medicine, law, and education. Doctors and hospitals, everyone thought, were the people and the places associated with pain and death. Only a few returned from medical treatment healthy, happy, and convinced that the medical professional was actually distinguishable from the local blacksmith or the visiting healer and snake-oil salesman. The true state of medical ignorance at that time has been well documented in a recent book entitled *Learning to Heal* (Ludmerer, 1985). For example, among Union troops during the Civil War, penetrating chest wounds incurred a 62 percent mortality rate, and abdominal wounds carried an 87 percent death rate. In contrast, during World War II, only about 3 percent of injured American soldiers failed to survive. The doctors of the Civil War period—120 years ago—believed that pus was part of the natural healing process, and thus infected wounds were rampant. Physicians themselves propagated much of the infection by probing wounds with fingers wiped on pus-stained aprons and towels. The importance of clean hands and sterilized equipment was simply not understood. Maggots were the most effective infection fighters of the war; they infested the wounds of the injured soldiers and ate only diseased tissue, leaving behind the healthy part of the wound. Although well known and often used in Europe, thermometers, stethoscopes, ophthalmoscopes, hypodermic syringes, and other simple instruments were rarely employed by American medical practitioners. The military physicians continued to give their patients dangerous doses of potent laxatives and emetics long after such substances had been proved absolutely ineffective or harmful. When a brave Union surgeon general, Dr. William Hammond, banned the use of two toxic drugs because he noticed they had severe side effects such as profuse salivation and development of putrid gangrene of the gums, mouth, and face, he was not only court-martialed by the military but condemned by the American Medical Association. The Union Army had developed a medical exam to weed out the most incompetent doctors. The exam, however, only served to reveal that "the typical physician expressed vague and confused ideas in barbarous English that was at defiance with all rules of grammar."

The standard course of instruction at the nation's medical schools when William James was founding experimental psychology in America consisted of two four-month terms of lectures during the winter season, with the second term usually identical to the first. So ignorant of contemporary developments in medical science in 1871 was one of James's colleagues, a Harvard professor of pathology and anatomy, that he confessed his inability to use the microscope. At no school was there any gradation of studies or sequencing of subjects. There was simply no logical order. Instruction was didactic. Rote learning was emphasized. Students endured between six and eight hours of lectures daily, supplemented by textbook readings. There were few if any written examinations. To receive a medical degree, students merely needed to pass a brief set of casual, perfunctory oral questions. The age of 21 tended to be an official requirement for graduation, but not too much attention was actually paid to this regulation. State licensing did not exist. A graduate with a degree from any kind of institution could practice anywhere in the country. The ability to pay the fee was the single entrance requirement for many medical schools. Literacy was definitely not needed. Indeed, when Charles Eliot, the eminent president of Harvard (1869–1909), proposed written examinations for graduation from the university's medical school, he was vigorously opposed by the professor of surgery, who argued that more than half of Harvard's medical students could barely write. In fact, medical schools were so disreputable and the profession was so bad that the best students chose careers in law, the clergy, or *teaching!*

Eliot pointed out that "an American physician or surgeon may be and often is a coarse and uncultivated person devoid of intellectual interest outside of his calling and quite unable to either speak or write his mother tongue with accuracy." Harvard's medical school was described at the end of the last century as a "money making institution, not much better than a diploma mill." Eliot wrote, "The ignorance and general incompetency of the average graduate of American medical schools, at the time when he receives the degree which turns him loose upon the community, is something horrible to contemplate." Eliot tried to put into effect a scientific M.D. degree, but he was voted down by his faculty. After all, the faculty argued, medicine was an art, not a science. The scientific approach was a fad, simply glorifying the European practice and showing disdain for the clinical wisdom of America's physicians. Experimental research was roundly scorned. Some of the most eminent medical professors of the time stated that they did not accept the germ theory of disease and were opposed to animal vivisection. The handful of people doing research in the field either had to be independently wealthy or had to steal equipment for their laboratories.

Lawyers in those days, everyone also seemed to think, were despicable. The profession appeared to be run by the greedy and the conservative, though the lawyer often came from a "better class" of people than did the medical practitioner. Legal education was really no more rigorous than medical education. It did not take much in the way of apprenticeship, and one had only to read a few law books in order to be able to practice law.

It was also true in our recent past that everyone seemed to think that teachers were simply persons who were unfit for the rigors for farm, factory, or commercial life. What was clear was that to the vast majority of the American people there were three equally disreputable professions—medicine, law, and education—and this was only a mere 100 or so years ago.

It is interesting to reflect on why two of these three disreputable professions gained stature over the past 100 or so years. Doctors have become wealthy and respected, assuming the qualities of a priesthood (and at times of a deity) because they have acquired knowledge, skills, concepts, and technology in the form of pharmaceuticals that have allowed them to treat and cure illnesses. Lawyers, although some retain a rather unsavory image, have managed to climb the occupational hierarchy in terms of finance and prestige. In our modern society they possess the knowledge and skill to respond to the ever more complex legal issues in the areas of commerce, finance, taxes, and social behavior.

Now let us look at the third profession—education. I firmly believe that now it is our turn. We now have something that an ordinary person does not have—a knowledge base consisting of facts, concepts, and technology that can transform our profession as well. Although this is a time of intense criticism of education, it is also a time that gives us the greatest chance to reformulate our profession. The transformation we envision will put us on a scientific and technological track that will command the public's faith in and admiration of our competence. I believe we will get their full support when they believe that our services cannot be performed by ordinary, untrained, inexperienced members of society.

We in education are haunted by the public's erroneous belief that someone can walk in off the street and deliver a curriculum to 30 or so children. Raising any number of children and having gone to school for a number of years does not make an expert teacher. I believe we are on the threshold of creating a scientific basis for the art of teaching that will be acceptable to the general public as *truly specialized knowledge*. As the art of medicine has been enriched by a scientific underpinning, so is the art of teaching about to be enriched. I propose to subject my beliefs to a few direct questions about the issue:

1. How good is contemporary research in education in comparison to research in other fields?
2. What function does research in education serve with regard to the field of education?
3. Is any of the research that has been done of genuine practical value?
4. How should we interpret the findings of educational research?
5. How should we transmit the findings of educational research to teachers?

I will try to answer these questions in turn.

HOW GOOD IS RESEARCH IN EDUCATION IN COMPARISON TO RESEARCH IN OTHER FIELDS?

Educational research has suffered from criticism. Within and without the profession we hear people saying how awful it is and how much more insignificant it seems than in the hard sciences. Educational research suffers because education is not only a social science but an applied social science, meaning to some that the research work must necessarily be even "sloppier" than that conducted in the "purer" social sciences and certainly of less trustworthiness than that in the hard sciences.

A report by Larry Hedges (in press) of the University of Chicago debunks this myth. I would like to describe some of his data. It's very important that we put a quick end to this nonsense about the wonders of research in physics and chemistry and the abysmal state of research in the social sciences.

One source for examining the quality of data in the hard sciences is in the reviews of research that are done to see if the data collected are consistent over time, scientists, locales, and so on. A group of particle physicists at the University of California at Berkeley has been collecting experimental results on the properties of elementary particles for years. Their reviews are analyzed quantitatively as well as qualitatively. The field of particle physics engages some of the best physicists in the world. There is much current interest in this field, and literally billions of dollars are spent on the overall research enterprise. Millions of dollars may be spent on a single experiment. What I am going to show you next is the consistency of results for reviews of data of the so-called stable particles—that is, those particles that are judged stable against strong decay. This class of particles includes the familiar proton, neutron, and electron. The data presented are on the mass and lifetime of these stable particles. It is a well-understood research area. Let us now examine Table 1.

The 13 particles described in Table 1 include the muon, charged pion, lambda, omega, and so on. The property studied is either the lifetime or the mass of these particles. The number of studies included

Table 1. Summary of Homogeneity Statistics from 13 Particle Data Group Reviews.

		ALL STUDIES				STUDIES IN PDG REVIEW			
Particle	*Property*	*Number of Studies*	χ^2	*p*	*R*	*Number of Studies*	χ^2	*p*	*R*
Muon	Lifetime	10	29.496	0.000	3.28	9	11.602	0.170	1.45
Charged pion	Mass	10	20.034	0.018	2.23	7	2.046	0.915	0.34
Charged pion	Lifetime	11	34.139	0.000	3.41	10	14.280	0.113	1.59
Neutral pion	Lifetime	11	37.897	0.000	3.41	10	14.280	0.113	1.59
Charged kappa	Lifetime	13	17.633	0.127	1.47	7	14.937	0.021	2.49
Short-lived neutral kappa	Lifetime	13	18.524	0.101	1.54	10	11.415	0.248	1.27
Lambda	Mass	10	39.037	0.000	4.34	5	4.791	0.309	1.20
Lambda	Lifetime	27	70.676	0.000	2.72	3	4.929	0.085	2.46
Sigma +	Lifetime	21	9.834	0.971	0.49	19	8.101	0.977	0.45
Sigma −	Lifetime	16	24.252	0.061	1.62	14	16.808	0.208	1.29
Xi −	Mass	11	9.802	0.458	0.98	9	2.707	0.951	0.34
Xi −	Lifetime	17	11.058	0.806	0.69	11	7.724	0.656	0.77
Omega −	Mass	11	8.611	0.569	0.86	10	8.591	0.476	0.95
Mean		13.9		0.239	2.11	9.2		0.395	1.43
S.D.		5.1		0.343	1.28	4.1		0.363	1.05

NOTE: χ^2 is the chi-square statistic and R is Birge's ratio.

SOURCE: From Hedges (in press).

in a particular review is then given; for example, there were ten separate research studies of the lifetime of the muon. The next column presents a chi-square analysis of the results of these data. Chi-square is a statistic that can be used to see if data from different studies are homogeneous or heterogeneous. Chi-square gets very large when big differences occur among a set of findings and is very small when few differences occur among the findings. The probability of a chi-square can be estimated, and the next column in Table 1 gives that probability. We find that 6 of the 13 reviews showed statistically significant disagreement among the studies. That is, the findings of different studies were discrepant. The column R presents a statistic called the Birge ratio. That ratio is 1.00 when a set of studies yields consistent results. It departs from 1.00 as the studies get more divergent. The average Birge ratio across these 13 reviews, covering nearly 200 different studies, is 2.00, which is 100 percent larger than would be expected if the studies yielded consistent results. On the right of Table 1 are studies that the particle data group used for its special, most rigorous reviews. Thirty-four percent of the studies that were available were deleted because the particle data group decided they did not find the data trustworthy. What is noteworthy is that even with this more restricted set of studies, the Birge ratio is still nearly 1.5 and 2 of the 13 reviews showed, through chi-square analysis, that there were statistically significant disagreements among the studies. We are not knowledgeable physicists, and we can only admire what modern physics has accomplished. But it looks like it has accomplished its achievements with data sets that are not at all in agreement. Now we must ask how these data stack up against social science data. Let's turn to Table 2 and look at 13 reviews of social science data that Hedges has assembled. The data in reviews published in refereed journals tell us whether the data in some area are consistent or inconsistent across a series of replicated studies.

Five of these reviews come from the so-called hard areas of psychology—studies of sex differences and cognitive abilities, for example. Six of the reviews were selected from what are "soft" areas of educational psychology—studies of the effectiveness of open education. And two of the reviews examine the validity of student ratings of instruction and the effect of teacher expectancies on student IQ, which might be considered some middle ground between the hard and soft areas of educational research. What do we find in using the same statistics, either chi-square or the Birge ratio? For the social science reviews in the area of education and educational psychology, similar results were found as in the study of the physical science data sets. The chi-square analysis reveals that 6 of the 13 reviews showed statistically significant disagreement among the studies. In addition, the Birge ratio is more than 2.0. If some of the outliers in these studies making up these reviews are thrown out, as the particle data group did with its untrustworthy data, we find that none of the 13 reviews shows a significant chi-square. That is, the

Table 2. Summary of Homogeneity Statistics from 13 Social Science Reviews.

Review		ALL STUDIES				REVIEWS DELETING SOME STUDIES			
		Number of Studies	*H*	*p*	*R*	*Number of Studies*	*H*	*p*	*R*
Linn & Peterson	Spatial perception[a]	62	99.88	0.001	1.64	56	43.24	0.828	0.82
	Spatial visualization[a]	81	98.81	0.056	1.27	81	98.81	0.056	1.27
	Mental rotation[b]	29	43.34	0.024	1.61	29	43.34	0.024	1.61
Becker & Hedges	Verbal ability[a]	11	32.69	0.000	4.09	9	11.36	0.078	1.89
	Field articulation[a]	14	19.29	0.056	1.75	11	19.29	0.056	1.75
Hedges, Giaconia, & Gage	Reading achievement	19	105.60	0.000	5.87	16	25.68	0.054	1.65
	Math achievement	17	43.65	0.000	2.73	14	17.40	0.182	1.34
	Attitude to school	11	21.62	0.017	2.16	9	9.00	0.342	1.13
	Self-concept	18	23.66	0.129	1.39	18	23.66	0.129	1.39
Crain & Menard	Randomized	13	12.22	0.428	1.02	13	12.22	0.428	1.02
	Longitudinal	57	56.15	0.469	1.00	57	56.15	0.469	1.00
Cohen	Validity of student rating	67	104.62	0.002	1.59	65	79.48	0.092	1.24
Raudenbush	Teacher expectancy[b]	19	17.61	0.414	1.04	19	17.61	0.414	1.04
Mean		32.1		0.123	2.09	30.8		0.243	1.32
S.D.		25.0		0.183	1.42	24.8		0.239	0.33

NOTE: H is the chi-square statistic and R is Birge's ratio or a generalized Birge's ratio.
[a] A three-parameter fit is used.
[b] A two-parameter fit is used.

SOURCE: From Hedges (in press).

findings tend to agree among themselves, and the average Birge ratio is 1.3. As Hedges (in press) says:

> The data reported in Tables 1 and 2 are strikingly similar. When all studies actually conducted are included, reviews in both the physical science and social science domains suggest statistical inconsistency among research results. . . . When studies with [questionable characteristics] are deleted, research results in both domains are much more consistent.
>
> What is surprising is that the research results in the physical sciences are not markedly more consistent than those in the social sciences. The notion that experiments in physics produce strikingly consistent (empirically cumulative) results is simply not supported by these data. Similarly the notion that experiments in the social sciences produce relatively inconsistent (empirically noncumulative) results is not supported by these data either.

It is clear from these findings that research in the hard sciences is no better or no worse than in the soft sciences, if the criterion is consistency of results when studying the same area. That is, replication in the hard sciences is no easier to achieve than it is in the soft sciences.

Hedges goes on to point out that the social sciences also have a bad reputation because their measurement systems are so inaccurate—at least as compared to, say, physics or chemistry. The social sciences rarely have ratio scales to work with, while the physical sciences often do. But it looks like measurement problems plague the hard sciences, too. Table 3 shows some thermal conductivities of chemical elements and presents the values given in the *Metals Handbook* of 1961 and those given in the 1970 review of metals by the Thermochemical Properties Research Center. In a single decade, between 1960 and 1970, we find differences in measurement of carbon ranging as high as 8,000 percent, of indium ranging to 242 percent, of iridium to 151 percent, and so on. Once again, we see that the problems of the social sciences are the problems of other sciences as well.

My goal in presenting these data is not to pick at thermochemistry or physics. Surely the achievements of the physical sciences are well known and respected. But the fact of the matter is that social science data do not seem to be any poorer than those in the hard sciences and, furthermore, our measurement problems are no different than those in the physical sciences. We have unique problems in social science research and particularly in educational research, but the inherent quality of our studies and our measurement systems should no longer automatically be taken to be inferior to those in the physical sciences.

A second point that I must make about educational research data is that these data have the potential for being as compelling as research data in other fields. I would like to discuss this issue by using an argument made by the man we honor in this volume, our colleague N. L. Gage (1985). In 1982 Gage noted that the results of a large-scale experiment with a drug called propranolol were released to the public.

Table 3. Thermal Conductivities of Chemical Elements for Which There Are Relatively Large Differences Between the Data in the 1961 *Metals Handbook* and the 1970 Compilation by the Thermochemical Properties Research Center.

Element	*TPRC (cut-off date December 1970)*	*Metals Handbook (1961)*	*Difference (%)*
Antimony	0.247	0.19	30
Beryllium	2.04	1.46	40
Calcium	2.02[a]	1.26	60
Carbon (graphite)	0.057–19.6	0.24	−76–8000
Chlorine (gas)	0.000086	0.000072	19
Chromium	0.944	0.67	41
Cobalt	1.01	0.69	46
Erbium	0.143	0.096	49
Gadolinium	0.105	0.088	19
Indium	0.821	0.24	242
Iridium	1.48	0.59	151
Lithium	0.850	0.71	20
Neodymium	0.165[a]	0.130	27
Plutonium	0.067	0.084	−20
Rhenium	0.481	0.71	−32
Rhodium	1.50	0.88	70
Silicon	1.52	0.84	82
Tellurium	0.0343[b]	0.059	−58
	0.0199[c]		
Thallium	0.463	0.39	19
Thorium	0.543	0.377(373.2 K)	44
Titanium	0.220	0.167	32
Yttrium	0.172	0.146	18

[a] = Estimated value.
[b] = Parallel to c axis.
[c] = Perpendicular to c axis.
SOURCE: This table was adapted from Touloukian (1975) by Hedges (in press).

Propranolol is one of the family of so-called beta-blockers, a family of drugs intended to increase the rate of survival of people who have had at least one heart attack. Random assignment of about 3,800 people to either the drug or a placebo treatment took place. The investigators were quite thorough; they used double-blind procedures and spent about $20 million on this study. The results of this experiment are shown in Table 4.

After 30 months, 9.5 percent of the men who had received the placebo had died, while only 7 percent of those who had received the beta-blocker had died. The drug reduced the percentage of fatalities by only 2.5 percent. Such a result would certainly be regarded as trivial in research on teaching. If 2.5 percent more children score above grade

Table 4. Beta-Blocker Study.

After 30 Months	*Received Propranolol (N = 1,900)*	*Received Placebo (N = 1,900)*
Percentage dead	7.0	9.5
Percentage alive	93.0	90.5
	100.0	100.0

Correlation obtained from these data = .045.
Percentage of variance explained = .2.

level or remain in regular classrooms due to an educational program, we do not run out and shout with glee. But in the medical arena, the results of the propranolol study were regarded as so important that the experiment was discontinued on ethical grounds; the researchers felt that their data were so strong that it was unethical to continue denying treatment to the control group. Their results led to the recommendation that propranolol be used, with a potential saving of about 21,000 lives per year in the United States alone.

In 1984 news of a similar experiment was released. This seven- to ten-year-long experiment cost $150 million and dealt with the lowering of cholesterol levels through diet and the use of a drug. These results are shown in Table 5.

Of the men who received the placebo, 9.8 percent had a definite fatal or nonfatal heart attack. Of the men who received the experimental treatment, only 8.1 percent had a definite heart attack. The result was front-page news in the *New York Times* and the *Boston Globe* and led to a cover story in *Time* and an article in *Science*, which held that the results would "affect profoundly the practice of medicine in this country." But the treatment produced only a 1.7 percent difference.

Now, let us ask the questions that are most often asked in educational research. That is, what is the correlation? and how much variance is

Table 5. Effects of Cholesterol Lowering.

Results After 9 Years	*Experimental (N = 1,906)*	*Placebo (N = 1,900)*
Percentage definite fatal or full heart attack	8.1	9.8
Percentage no fatal or definite heart attack	91.9	90.2
	100.0	100.0

Correlation obtained from these data = .03.
Percentage of variance explained = .1.

accounted for by the treatment effects? If we go back to the study of the miracle drug propranolol we find out that the correlation between longevity and treatment is .045. The percentage of variance explained is .2. If we look at the effects of lowering cholesterol levels, we find that the correlation is a minuscule .03 between heart-attack rate and the type of treatment received. The percentage of variance explained is a trivial .1. These data lead to statements about medical miracles, but data of greater magnitude in the field of education are often simply disregarded. As Gage pointed out, correlations or differences do not need to be large in order to be significant. In education we may not be influencing life or death, but we are influencing dropout rate, literacy, placement in special classes, love of learning, self-esteem, and the ability to integrate many facts and concepts in complex ways. The implications of research for practice depend *not* on the size of the effects but on whether we see any benefits, whether change in practice results in a change we value. When we deal with the many millions of students in our classrooms, even small differences result in significant amounts of overall change in the lives and productivity of our students.

With this as background for the interpretation of the research literature, what have we discovered over the last dozen or so years? Let us take a look at the second question.

WHAT FUNCTION DOES RESEARCH IN EDUCATION SERVE?

I believe that five characteristics are indicative of a productive scientific community. First of all, a productive scientific community should verify ideas and practices believed to be effective by most people. Second, a productive scientific community should discover new ideas and practices. Third, there should be an expression of ideas that actually complicate everyone's lives. This is because I believe a scientific community should not flinch from ideas that make our tasks more complicated than we would like. Fourth, there ought to be an elucidation of ideas that simplify everyone's lives. And finally, there ought to be the discovery of ideas and practices that are counterintuitive. It seems to me that if these characteristics exist in educational research, we can talk of a very productive scientific community. Let me try to give you some examples under each of these headings.

Verifying Existing Ideas

First of all, our research community has verified ideas and practices known to be effective by most people. The president and the Department of Education recently released a report called *What Works*, a description of 41 research-based findings about how to help students and schools achieve better (U.S. Department of Education, 1986). In this little book are a number of ideas that verify things many people already know. For example, the report states that well-designed homework assignments that relate directly to classwork extend student learning beyond the classroom. It goes on to say that homework is most useful when teachers carefully prepare the assignment, thoroughly explain it, and give prompt comments and criticisms when the work is completed. Although the report does not tell us how teachers are supposed to fit time for the correction and feedback into an already busy day, the findings are quite clear: homework of this type improves achievement. Another verification of ideas known to be effective is the research in the last decade or two on how time spent on a task is a strong predictor of classroom achievement. I was responsible for some of that research, and I well remember people asking us why we would study the obvious. Everyone knows that time is related to learning, we were told. What we found, of course, was that time was not spent equally in classrooms. That is, there was incredible variation across classrooms in the way time spent on a task occurred, and it was this incredible variation that was the predictor of achievement. Still another recent finding that verifies something we all know is that new learning is best remembered when it is related to old learning. This goes back at least as far as the Greek philosophers and became a "modern" theory during Herbart's time in the mid-nineteenth century. New research on "schema theory" and learning verifies the ideas and practices that educators have held for years. Research will, I am sure, continue to provide scientific underpinnings to ideas that are now called "common sense." That is good. Common practices need such justifications, and we have learned, time and time again, that common sense is not always common practice.

Finding Out New Things

Is our scientific community coming up with any new ideas and practices? I think the answer here is yes. For example, Annmarie Palincsar and Ann Brown (1984) have recently written about reciprocal teaching, a new method of teaching whereby students who have deficits in comprehension skills are taught ways to ask questions of text so that their comprehension improves. With the teacher and other students, they take turns creating different kinds of questions until they have developed metacognitive skills—that is, the habit of reflecting upon what they are reading. Reciprocal teaching works. It holds much promise for helping students who are able to decode but have trouble comprehending reading.

Another new idea brought forth by Walter Doyle (1983) deals with the negotiation of low-risk tasks in classrooms. While many teachers are criticized for keeping the intellectual level low, Doyle finds that

that children learn when they are "stretched," that some intermediate level of difficulty is needed. Perhaps this is because adults seem to learn as often from their errors as from their successes. But the research seems reasonably clear that such is not the case for young children. In approximately the first through fifth grades, a number of researchers are estimating that children's homework assignments and workbook assignments ought to yield a success rate of 90 percent or better and that the questions asked in classroom discourse and recitation should probably yield 80 percent or more successful responses.

One of the oldest examples of counterintuitive ideas in the research literature has to do with the status of "eggheads." Common folklore says that the brightest students are somehow "nerdy" in their appearance and in their social relationships, that they have maladjustments through life, and so on. There is a continuing pervasive belief that the most brilliant children are the least well adjusted. The evidence against this was produced by Terman in his monumental longitudinal study of the gifted that began in the 1920s (Terman, 1925). These children of genius were, in general, more physically attractive than the general population and in better health than the general population. They needed glasses less often, engaged in sports more often, had better social relationships, and were class stars more often and class isolates less often.

It is my belief that research in education has already passed the test of being a productive endeavor. It verifies ideas and practices known to many, it provides new ideas and practices, it promotes ideas that complicate life, it promotes ideas that simplify life, and it provides research that is counterintuitive. Let us move on, then, to the third question.

IS ANY OF THIS EDUCATIONAL RESEARCH OF PRACTICAL VALUE?

I would like to discuss the practical value of educational research by citing findings, concepts, and technology that educational researchers have developed in the last few years. A brief but representative set of findings that show the very practical nature of contemporary educational research includes studies on success rate, structuring, academic feedback, monitoring, motivating, expectancy, and wait-time.

Success Rate

I have already discussed the high success rate that young children need and how the research findings in this area have been counterintuitive. I might also add that the data supporting the value of a high success rate for young children seem to be more impressive with lower-class children, and the data also appear to be more impressive when the nature of the curriculum is hierarchical. That is, unless a child has been successful in learning addition, he or she will have trouble in learning multiplication, and unless the child has been successful at subtraction, he or she will have difficulty in learning long division. Moreover, the necessity for high success rates for young students, where curriculum has been carefully matched to the student so that the student can succeed at it, seems to be the precursor for the development of a positive academic self-concept (Shavelson & Bolus, 1982). The evidence from educational research appears to be quite clear: positive self-concept as a learner *follows* successful experiences as a learner.

How do effective teachers obtain these high success rates? This topic is discussed in Chapter 4. Some of the answers that have been learned from studying effective teachers include: presenting new material in small steps followed by student practice; directing initial student practice through problems and questions; continuing practice until students are confident; and providing for additional, spaced practice to the point where student responses are automatic.

Structuring

Another important research finding is that structuring behavior (providing clear directions, objectives, reviews, and advance organizers for material to be presented) improves classroom achievement. Research shows this to be true. When the distinguished psychologist Jerome Bruner looked at American schools after having spent about a decade in England (Bruner, 1981), he thought that one of the key factors for improving education in America would be to tell children what is expected of them. He saw many children unable to figure out what to do at what time, or where to be, or whom to study with. My own experience of reading protocols of lessons led me to the same conclusion. I spent a good deal of time reading transcripts or protocols of the lessons of teachers who had been classified as either highly effective or less effective. While reading the transcripts, I tried to put myself into the mind of a second- or a fifth-grader. I asked myself whether I could simply understand from the record of classroom instruction what was going on. If the answer was yes, I had almost always identified an effective teacher. If the answer was no—because there were too many interruptions, or the lesson was disjointed, or I could not see where the lesson structure was—the teacher almost always turned out to be less effective. Thus I learned about the practical importance of structuring.

Academic Feedback

We have also learned that effective teachers provide academic feedback. It can be positive or negative feedback, as long as it is academic in its focus. Negative feedback sometimes gets confused with criticism, and criticism *is* harmful to students. Academic feedback that is negative and task-oriented, however, is a positive predictor of achievement. There is a big difference between saying to a child "You write badly" and "The essay you wrote is not good, for the following reasons." Teachers who give a higher rate of academic feedback produce higher levels of achievement than their colleagues who give a lower rate of academic feedback (Fisher, Berliner, Filby, Marliave, Cahen, & Dishaw, 1980). In order to provide such feedback at high rates, there is a need for creative time management by teachers. But creative interpretations of the research are not as necessary as William James had thought. Some of these findings are much less in need of translation to practice than were the findings James had to offer the teachers in his day.

Monitoring

Effective teachers also engage in monitoring during seatwork. An effective teacher does not sit in a corner working with one student in a tutorial. An effective teacher does not grade papers or perform other relevant or irrelevant activities while seatwork is occurring. Effective teachers walk the classroom answering questions privately, working with students privately, and maintaining attention. Effective teachers also provide contingent reinforcement. They do not say "Good" to everything. They do not give out gold stars easily. They make sure behavior and reinforcement are related.

Motivating Students

Effective teachers introduce people to tasks in ways that are positively motivating (see Chapter 9). However, Brophy, Rohrkemper, Rashid, and Goldberger (1983) observed six intermediate grade classrooms 8 to 15 times during reading and mathematics instruction and found that not one teacher ever mentioned that students might derive any personal satisfaction from a learning task. Only a third of the introductory statements included comments likely to have positive effects on student motivation, and most of these were brief general predictions that the students would enjoy the task or would do well on it. One hundred hours of classroom observation yielded only nine task introductions that included substantive information about motivation. None of these went into enough detail to be very meaningful or memorable for most students.

Expectancy

Another set of findings from research on teaching suggests that expectations for performance do actually affect performance. In comparison to students for whom teachers hold high performance expectations, the students perceived to be low performers are more often seated farther away from the teacher, are treated as a group rather than as individuals, are smiled at less, are engaged in eye contact less, are called on less to answer questions, are given less time to answer those questions, have their answers followed up less frequently, are praised more often for marginal and inadequate answers, are praised less frequently for successful public responses, are interrupted in their work more often, and so forth. This kind of differential treatment between students for whom teachers hold high and low expectations appears common enough to worry about and is discussed at length in Chapter 8.

Wait-Time

This finding is particularly interesting. Wait-time is usually defined as the amount of time a teacher waits between asking a good, meaty, higher-order question and calling on students. Researchers have completed at least 16 studies on the topic of wait-time (Tobin, in press). The results clearly indicate that when teachers learn to wait 3 seconds before calling on a student, instead of their customary 0.8 second, some wondrous things happen. The following seven events have been found to be connected with increased wait-time when a good, strong, higher-order question is asked:

1. The length of student response increases.
2. The number of unsolicited but appropriate responses increases.
3. The failures to respond decrease.
4. The complexity and cognitive level of student responses increase.
5. The number of alternative explanations increases.
6. The incidence of child-to-child interactions increases.
7. Student achievement in science and mathematics increases.

No negative side effects are known. The positive effects seem to be many. Yet the behavior is not in the repertoire of most teachers.

It is important to remember that the findings of these 16 studies have the power of the medical research we discussed earlier. The causal connections with achievement are persuasive. The amounts of variance accounted for are persuasive. In medicine, such findings would make their way into accepted practice in just a few years. But in education we do not implement our findings. This does not mean to say that every time a teacher asks a question he or she should wait. What I do mean is that teachers ought to have this behavior in their repertoire, and know how to justify the use and nonuse of wait-time as they work through their lessons. Thoughtful application of our findings in educational research is what is needed, not blanket, unthinking application.

Now let us look at another kind of useful information derived from research. These are the concepts that help us organize our world better—

by helping us to name things, by giving us a technical vocabulary that we can then use with colleagues as we describe classroom phenomena. Concepts need to be easily accessed by teachers if they are to help teachers organize the myriad phenomena present in classrooms. A representative set of practical concepts is discussed, and these include a number of time concepts, which are discussed further in Chapter 5.

Allocated Time

The first of these concepts is a simple one—allocated time. It turns out that allocated time is a positive predictor of achievement because there is enormous variation in the amount of time allocated for instruction in each classroom. We have studied classrooms in which allocated time ranges from 16 to 50 minutes a day in mathematics instruction and from about 45 to 137 minutes a day in reading instruction. These are variations of 2:1 and 3:1 in the amount of time allocated for instruction. This is the kind of concept we must have easy access to in our minds when we think about instruction. When the public criticizes us—when they say children do not know how to write, or they do not comprehend well, or they are not getting enough science—the public is asking whether we as educators have allocated enough time for writing, comprehension, and science activities. Whenever we hear someone criticize the schools in this way, we should ask ourselves whether we are allocating sufficient time. If we are not allocating sufficient time, then students do not have the opportunity to learn the skills the public is asking for. This is a very useful concept for thinking about what instruction is actually offered to students.

Engaged Time

A second time concept that contributes to our technical vocabulary is engaged time. Engaged time, or time-on-task, or attending time, is also a positive predictor of achievement, again because enormous variation has occurred. We find 40–50 percent time-on-task in one classroom, while next door, perhaps with children of the same socioeconomic group, we may see 90 percent time-on-task. These are enormous differences in the amount of curriculum the children are exposed to. We have learned by organizing our observations around these kinds of time concepts that some elementary classrooms cannot even deliver 100 hours of time-on-task over the course of the school year in reading and mathematics combined! This happens, for example, because the school year is not really a year but is in fact nine months. The nine months are not a full nine months; they are really 180 days of instruction. However, the 180 days of instruction must be reduced by the number of days on which there are plays, field trips, strikes, bus breakdowns, snow days that are not made up, mental-health days for teachers and students,

and genuine illness of teachers and students. We might end up, if we are lucky, with 140 *real* days of instruction. If mathematics is taught 30 minutes a day—which is very common in, let us say, second grade—we end up with a total allocated time for mathematics of 70 hours for the year. With a low engaged time rate, say 50 percent, the curriculum delivered is 35 hours. All the mathematics for the year is 35 hours! Reading may be double that, about 70 hours, and the whole year may total about 100 hours of engaged time. It should be obvious, therefore, that engaged time is an important concept for thinking about instructional activities in the classroom.

I think that time is the single most important resource over which schools have control. Time concepts, therefore, are important concepts for thinking about curriculum and instruction. They are the yardstick by which we can measure and judge if we are doing what we are trying to do as educators. Research informs us that these concepts are of major importance. Thus they need to be learned well by teachers. I would argue that teachers need to have easy access to these concepts when thinking about their instructional activities.

Curriculum Alignment

Another concept of great utility is variously called curriculum alignment, or overlap, or congruence. This is a very powerful concept in a test-oriented (perhaps test-crazy) society such as ours. The concept refers to the alignment between what is taught and what is tested. Schools, we think, are vastly underestimating their students' performance by using tests mismatched to the curriculum (Freeman et al., 1980). Research informs us, for example, that if a school district is unfortunate enough to have chosen the Stanford Achievement Test as a criterion and uses the Addison-Wesley fourth-grade Mathematics Curriculum Series, there will be an estimated 47 percent overlap. That is, items on the test will have been found to correspond to items in the text about 47 percent of the time. That means, of course, that 53 percent of the items on that test will have been unfamiliar to the students unless the teachers have taken extraordinary action. Even in the best case, with the Scott-Foresman series and the Metropolitan Achievement Tests, the overlap is only about 70 percent. If you ever want to underestimate what your students learn, simply teach one thing and test another! This concept holds for classes as well as districts. If you ever want to get students angry, teach A, test B. In college we get rebellions; in public schools you get anger, hostility, and high dropout rates. This is a very important concept. It helps remind teachers each day that what they do in class should match the outcome measures they use or, *more important*, that the outcome measures they use should match what they do in class.

Concepts to Guide Small-Group Work

Some concepts serve primarily descriptive purposes. They mostly help us organize our world better by rendering that world more comprehensible. Conductance and resistance in electricity; social class in sociology; ego, id, and superego in psychology; and homeostasis in biology are concepts of this kind. These kinds of concepts enrich our lives by helping us describe phenomena. In pedagogy we are also developing a wonderfully descriptive conceptual language from our research endeavors. For example, let us consider the teacher who might choose to work in small groups on a lesson in, say, social studies. He or she may have to use many different techniques simultaneously in order to have successful instruction. The techniques of modern cognitive science have been used to explore teachers' decision making during such interactive group work (Marland, 1977). It was found that teachers who engage in small-group instruction seem to be simultaneously attending to five principles of teaching:

1. *Compensation*, or favoring the shy, the quiet, the dull, or the culturally different.
2. *Strategic leniency*, or ignoring some of the inappropriate behavior of special children.
3. *Power sharing*, whereby the teacher selectively reinforces certain students in order to enlist their aid in sharing responsibility.
4. *Progressive checking*, in which the teacher makes a special effort to check the problems and progress of low-ability students.
5. *Suppressing emotions*, which many teachers feel is appropriate during certain kinds of teaching. Their reasoning is that showing emotion could lead to higher levels of emotionality among the students, which creates management problems in some of the curriculum areas.

Thus the apparently simple task of running a small group, when examined from a cognitive science perspective, requires complex decision making about the application of many principles, and this kind of decision making makes considerable cognitive demands of a teacher. Many rich concepts are used to perform the task.

We have only recently been able to describe exactly how complex the job of teaching really is: researchers have found that teachers make about 30 nontrivial decisions per hour. They are decisions about whether Johnny should stop fractions and go on to decimals, or whether Jane should be moved into the fast mathematics group. These complex, professional, nontrivial decisions take place in environments where teachers can have 1,500 distinct interactions per day with different children on different issues, in different classes where aggregates of students need to be supervised all the time. We are just beginning to uncover the concepts that help us to describe how experienced teachers make

decisions in the face of this kind of enormous complexity. We are expanding our practical knowledge base very rapidly.

To conclude these comments on practical concepts, let me remind you that these concepts are the names of things, and that by naming and defining things we thereby organize our world and render it more manageable. When we have a precise concept like "allocated time," or descriptive concepts like "gifted" or "learning-disabled," or even that wonderful concept given to us by Jacob Kounin (1970) and called "withitness," we are developing our technical vocabulary and our language to communicate with each other. The development of a technical language goes hand in glove with the development of our science and the ability to earn more respect from the general public. We in teaching need our unique pedagogical concepts to develop our technical language.

Now let us discuss briefly the new technology we have developed that was totally unforeseen by William James in the 1890s.

Cooperative Learning

In the last decade we have created the technology for producing cooperative behavior among students and between the students and teachers (see Slavin, 1983). These techniques have also given rise to higher self-esteem for the learners, as well as higher rates of proacademic behavior and, even more astounding, higher academic achievement. These recent experiments in cooperative learning environments have involved thousands of students in thousands of classrooms. In the process we have learned how certain teaching techniques can help to integrate handicapped children into the mainstream of the classroom, help to integrate minority students into the majority culture, and help to produce more nearly equal performance of students of different social classes. Such technology has been developed and field-tested only in the last ten years. This is the kind of technology to be taught to preservice and practicing teachers throughout the country. It should influence teacher behavior directly.

Classroom Management

In classroom management, we also have new technology. This is the area that the press and the public love to criticize teachers about, and it is the area that teachers themselves are most afraid of when they begin to teach. We have made unbelievable strides in the last decade. We have identified many of the teacher behaviors that lead to smooth-running, on-task, cheerful classrooms. The original research was first reported only in 1970. Others verified the findings through the 1970s and transformed them into teacher-training materials during the early 1980s (see Chapter 3). These materials, based on empirical research, have been field-tested recently by the American Federation of Teachers. In New York and elsewhere, the results have been amazing. Teachers who had failed to meet the criteria of good management were in complete control of their classes for the first time in their professional lives. One twenty-year veteran of the New York City schools said that the training produced nothing short of a miracle.

Outcome-Based Learning

Let us look at one last example of recently created technology. This technology has to do with outcome-based or mastery strategies (Spady, 1984). We think first of the California Achievement Test data in reading for Johnson City, New York, a small blue-collar community, that has used Bloom's mastery concepts for a number of years. The data we have seen are for 1978, 1979, and 1980. Over time the students do better and better in relation to grade-level norms. By the time students finished eighth grade in 1980, they were two and half years above grade level. On the California Achievement Test in mathematics, Johnson City students gained greater amounts relative to national averages the longer they stayed in the system. In fact, in 1979 and in 1980 students finished about three years above grade level. Similar data are available from another town.

In Red Bank, New Jersey, in 1978, the average eighth grader scored 1.5 years below grade level in reading on the Metropolitan Achievement Tests. In 1983, the eighth graders were 1.5 years above grade level. In mathematics, from 1978 to 1983, the eighth-grade class went from a mean of 1.5 years below grade level to a mean of almost 2.5 years above grade level. In five years of a special program in Red Bank, the mean achievement level in mathematics went up four years on the Metropolitan Achievement Tests. A similar program was found in an elementary school in New Canaan, Connecticut. The overall results showed that about 15 percent of the graduating sixth graders in that school finished half of Algebra 1. About 60 percent of the sixth graders scored at the 99th percentile on the mathematics section of the Metropolitan tests. About 30 percent of the fifth graders scored at the 99th percentile. In the entire school in the last seven years, only about two students per year scored under grade level on this standardized achievement test. The programs in use in these very different schools are outcome-based mastery-type programs. This technology must be understood in order to decide whether to use it or to reject it. Since there are impressive data to support its utility, we all should take this kind of technology very seriously.

In summarizing this discussion of the usefulness of research for practitioners, we can conclude that we have a strong research tradition and a strong research base with which to build bridges to practice. Our scientific base grows richer every month. Our findings, our concepts,

and our technology are becoming more advanced every year. Our research is every bit as good as, or better than, that in the medical profession. Our kind of research, however, is much tougher to do. Our colleague N. L. Gage was dining not long ago with a Nobel laureate in biology. The Nobel laureate was very interested in the kind of research that we do in education. Gage asked him, "Why don't you join us? Why don't you become an educational researcher? We need your talent!" The Nobel laureate looked Gage in the eye and said, "God, what an awful thought! That kind of research is much too difficult to do!" This brings us to the fourth question that needs to be asked about research in education.

HOW SHOULD WE INTERPRET EDUCATIONAL RESEARCH FINDINGS?

We have a number of problems in interpreting educational research. These problems limit the generalizability of our findings in a way that does not occur in other professions. Classrooms are complex, dynamic environments. They are kept together by a delicate balance of forces. Sometimes these forces favor us; sometimes they conspire to do us in. Our classrooms are constantly changing: the time of year, current events, special needs of our students—these and other factors sometimes all plot to throw up barriers to implementing educational research in ways that will prove its usefulness. We also have the problem of new brooms in education. That is, we are subject to fads, new curricula, new textbooks, new demands for accountability, new state tests, new superintendents, principals, or curriculum consultants. Findings that have held up in one place, therefore, may not always hold in another place when new events absolutely must be dealt with.

It is also true that we have size problems in education. Technology that might work in tutorial sessions or in small groups of 8 students may not work in a class of 27. Concepts that are useful to describe a group of 8 students may be useless to describe a heterogeneous class of 28. As William James noted only too wisely, educational research needs to be implemented by a thoughtful, reflective practitioner. Our practitioners are very human, and this means that a teacher's personal relationships and personal status sometimes get in the way of implementing programs of research. For example, to implement the mastery program requires hours and hours of preparation time. A teacher who is also a single parent may not be able to spare that time. Teachers who do other kinds of work in the evenings or on weekends may simply not be able to provide the feedback necessary to allow homework assignments to improve achievement in the ways that research says it might. My point is that because of problems such as the delicate balance of classrooms, the way new brooms sweep, the size of the enterprise, and the personal status and relationships of the people involved, the generalizability of educational research is problematical. I believe that in physics and chemistry the relationship of the investigator to what is being studied affects the findings only slightly. In medicine we find more problems when we discover how individual differences affect the course of treatment. We learn, for example, that penicillin works, but not for everybody. In educational research, we are at the other end of the continuum. We work in settings where the investigator, the subjects, and the context all interact in confusing ways. In comparison to some other fields, educational research is much more fragile.

But note: fragility does not imply uselessness. In our culture, fragility means that something should be handled with care. And that is the point. Educational research needs to be taken seriously but handled with care. I am not sure whether educational research will generalize as well as physical and biological findings often seem to. Nevertheless, when we find some conditions under which these findings hold, we are obligated as professionals to take them very seriously. The job of teachers, as William James noted, is to use their minds creatively to apply those findings.

Now, if we are to take the body of findings, concepts, and technology seriously, how should we think about communicating that body of knowledge to practitioners? This brings us to the final question.

HOW SHOULD WE COMMUNICATE RESEARCH KNOWLEDGE TO PRACTITIONERS?

The educational philosopher Gary Fenstermacher (1980) has talked about two ways to put research findings, concepts, and technology into practice: through rules and through discussion of evidence. One of these ways enhances the lives of teachers and one does not.

Communicating Research by Rules

Rules may serve as the means for bridging educational research and practice. This happens when the results of research are converted to imperatives for teachers to follow. For example, from the findings of one study, we learned that "more substantive interaction between the student and an instructor is associated with higher percentages of student engagement." A person engaged in bridging research and practice could use this finding as a rule to govern teacher practice. A principal I know, for example, having read this study, asked his staff to devote not less than half the time available in a given instruction period to teacher-led small-group instruction. In taking this action, the principal bridged research and practice by converting a finding to a rule and requesting compliance with that rule. The principal's school board thought well of him. Nevertheless, a request by someone in authority is often viewed as a command by those who have less authority. That is, there is always a veiled suggestion that doing X will be a factor in a person's evaluation; thus the request to do X can turn into a most demanding requirement. In every case, there is an expectation that one will modify one's behavior according to some other person's interpretation of the research finding. When using rules to communicate, the recipient of the command is not asked to consider the research finding itself but rather is asked to behave in ways suggested by the finding. Communicating research findings by means of rules brings little, if any, advantage to practitioners and makes most researchers angry. It takes research findings out of context, and it also ignores the limits to generalization that concern researchers deeply. Perhaps the most debilitating aspect of communication by rules is its effect on a practitioner's perception of his or her stature and competence. Persons expected to change their behavior on the basis of rules imposed by others are denied a portion of their freedom to think and act independently. If practitioners are ever to have the opportunity to grow as professionals, other means of communicating research to practitioners may be far more productive. Dewey (1929) recognized the temptation to communicate by rules and cautioned that "no conclusion of scientific research can be converted into an immediate rule of educational art." He was, and is, right. William James, as I noted earlier, said that teachers would make a great mistake if they thought that psychological science would provide them with prescriptions for how to act in classrooms. Developing rules is not a style of communicating knowledge that we can approve.

Communicating Research Through Discussion of Evidence

There is another way to communicate knowledge, however, and that is through serious discussion of evidence. A good use of evidence occurs when the results of research are used to test the beliefs that teachers hold about their work. For example, a teacher might argue that it is perfectly acceptable for students to make mistakes. "After all, how can you learn if you don't make mistakes?" A teacher who believes this may be quite casual about preparing classroom or homework assignments for students, thinking that they can always ask questions if they are confused or that their confusion will make them think harder. Findings from one of the studies discussed earlier stated that "tasks which produce low error rate result in higher achievement as measured by standardized tests." This finding casts doubt on the adequacy of our hypothetical teacher's belief about the acceptability of errors. It may not always be possible, or always advisable, to provide error-free tasks, but the findings suggest that the teachers should be more careful in developing and assigning learning tasks, homework, and seatwork to young children. What the research findings do is call into question the reasonableness of the practitioner's beliefs. This is good. This is how we grow. The point made by Fenstermacher, with which I agree, is that a confrontation with evidence (such as having teachers thoughtfully discuss the meaning of findings, concepts, and technologies that researchers have generated) does not require a teacher to modify his or her beliefs automatically every time research findings are presented. It requires only that the teacher weigh the results of the research seriously. To insist on more would be to place greater confidence in educational research than it legitimately commands. In this respect, William James was right. The translation of research findings into practice will always take a creative and inventive mind. For example, consider again the finding that "tasks which produce low error rate result in higher achievement." Now suppose the teacher is teaching long division (or fractions, or the skill of identifying the topic sentence in a paragraph). Simply stating this finding, or even giving two examples, does not tell the teacher how to present the lessons in long division. Even though the research of today is closer to the world of practice, we are still faced with the fact that research rarely can be applied automatically and directly to solve classroom problems.

Our research findings may often conflict with the beliefs held by teachers—beliefs that are supported on the basis of other reasonable grounds. Thus we are well aware that a teacher may at any time be quite justified in choosing to modify or ignore certain findings. They are required, however, as professionals, to reject or adapt research findings thoughtfully. This is what we hope new professionals in teaching will be able to do. They should have a great deal of knowledge about our findings and our concepts. They should have had practice in using our technology. And they should also know when not to use such findings, when such concepts are inadequate, and when such technology is inappropriate.

Given this discussion, we can compare and contrast the two forms of communicating research: rules versus serious discussion. Rules are imposed with the expectation of obedience, while discussion of findings, concepts, and technologies derived from research leads to serious consideration. Rules are imprecise representations of research findings because their construction requires the rule maker to interpret the findings; discussion conveys to the practitioner precisely what has been learned from research studies and lets *them* interpret the data. To adapt or ignore a rule is frequently regarded as an act of subversion, whereas evidence may be freely and openly accepted, rejected, or modified. The imposition of rules can leave a teacher's beliefs, whether sound or unsound, unaffected, while a consideration of evidence encourages the clarification and assessment of one's prior beliefs. Discussing evidence accords the practitioner the status of a thinking, reasoning person; communicating rules treats the practitioner as if he or she were little

more than an automaton. The remainder of this book discusses research findings by people who are engaged in the process of creating knowledge. They are talking directly to practitioners. They also hope that the arguments they make will be checked by practitioners who will seek out and read the original research. Through analysis of the original research and discussion of the conclusions of that research, teachers will confront evidence that may or may not support their beliefs about teaching. The result of that confrontation is bound to be good. If teachers do not weigh the evidence brought forth by this research and critically analyze it, someone else is likely to take the findings, concepts, and technologies reported on in this volume and develop rules for teachers to follow. No one engaged in research wants that to happen.

SUMMARY

I have talked about three professions—medicine, law, and education—that were equally disreputable 100 or so years ago. In my analysis, knowledge and skill are what gave the first two the high level of status they now hold. Knowledge is clearly power, a kind of social power. It commands respect and confers status in our technologically oriented society.

Educational research is fully prepared to bring that kind of power to the teaching profession. I hope I have made clear how our research has power, like medical research, if we simply choose to treat it seriously. I hope I have communicated that we have a productive scientific community that provides practical findings for teachers to think about. I hope I have convinced you that despite our problems with generalizability, the educational research produced thus far has to be taken very seriously. And finally, I hope I have convinced you that the communication of research findings to practitioners has to be done in a way that benefits and strengthens the practitioners.

The research reported in this volume is only a part of the broader research that is being developed to transform the teaching profession as it prepares for the twenty-first century. William James would not have recognized the current productivity and creativity in educational research. He probably would still be considered right when he commented that no research findings will tell a teacher what to do with Johnny on Monday morning. But at the same time, he could also be considered to be out of date. The educational research community is providing practitioners with more and more findings, concepts, and technologies that are closely related to their performance as teachers in the classroom. I believe that it no longer requires a great leap of faith to do what James said was impossible—"to deduce definite programmes and schemes and methods of instruction for immediate schoolroom use." The research community of today, so heavily influenced by N. L. Gage, is working with creative teachers everywhere to transform the teaching profession into one in which artists interpret science in ways that allow rapid and successful implementation of research into the schoolroom.

REFERENCES

Berliner, D. C. (1985). Effective classroom teaching: The necessary but not sufficient condition for developing exemplary schools. In G. R. Austin & H. Garber (Eds.), *Research on exemplary schools*. Orlando, FL: Academic.

Brophy, J. Rohrkemper, M., Rashid, H. & Goldberger, M. (1983). Relationships between teachers' presentations of classroom tasks and student engagement in those tasks. *Journal of Educational Psychology*, 74, 544–552.

Bruner, J. (1981, August). *On instructability*. Paper presented at the annual meeting of the American Psychological Association, Los Angeles.

Dewey, J. (1929). *The sources of a science of education*. New York: Liveright.

Doyle, W. J. (1983). Academic work. *Review of Educational Research, 53*, 159–199.

Fenstermacher, G. D. (1980). On learning to teach effectively from research on teacher effectiveness. In C. Denham & A. Lieberman (Eds.), *Time to learn*. Washington, DC: National Institute of Education, U.S. Department of Education.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. A., & Dishaw, M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), *Time to learn*. Washington, DC: National Institute of Education, U.S. Department of Education.

Freeman, D., Kuhs, T., Porter, A., Knappen, L., Floden, R., Schmidt, W., & Schwille, J. (1980). *The fourth-grade mathematics curriculum as inferred from textbooks and tests*. IRT Report No. 82. East Lansing, MI: Michigan State University, Institute for Research on Teaching.

Gage, N. L. (1985). *Hard gains in the soft sciences: The case of pedagogy*. Bloomington, IN: Phi Delta Kappan.

Hedges, L. V. (in press). How hard is hard science, how soft is soft science: The empirical cumulativity of research. *American Psychologist*.

Holmes, C. T., & Mathews, K. M. (1984). The effects of non promotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research, 54*, 225–237.

James, W. (1983). *Talks to teachers on psychology and to students on some of life's ideals*. Cambridge, MA: Harvard University Press.

Kounin, J. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.

Ludmerer, K. M. (1985). *Learning to heal*. New York: Basic Books.

Marland, P. W. (1977). *A study of teachers' interactive thoughts*. Doctoral dissertation, University of Alberta, Alberta, Canada.

Palincsar, A. M., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117–175.

Powell, A. G., & Farrar, E. (1985). *The shopping mall high school*. Boston: Houghton Mifflin.

Rosenshine, B. (1979). Content, time, and direct instruction. In P. L. Peterson & H. J. Walberg (Eds.), *Research on teaching*. Berkeley, CA: McCutchan.

Rosenshine, B., & Berliner, D. C. (1978). Academic engaged time. *British Journal of Education, 4*, 3–16.

Shavelson, R. J., & Bolus, R. (1982). Self concept: The interplay of theory and methods. *Journal of Educational Psychology, 74*, 3–17.

Slavin, R. (1983). *Cooperative learning*. New York: Longman.

Spady, W. G. (1984). Organizing and delivering curriculum for maximum input. In *Proceedings of three state conferences: Making our schools more effective*. San Francisco, CA: Far West Laboratory for Educational Research and Development.

Terman, L. M. (1925). Mental and physical traits of a thousand gifted children. In L. M. Terman et al., *Genetic studies of genius*, Vol. 1. Stanford, CA: Stanford University Press.

Tobin, K. (in press). The role of wait time in higher cognitive level learning. *Review of Educational Research*.

Touloukian, Y. S. (1975). Reference data on thermodynamics. In H. A. Skinner (Ed.), *International review of science, physical chemistry: Vol. 10, Thermochemistry and thermodynamics*. London: Butterworth.

U.S. Department of Education. (1986). *What works*. Washington, DC.