

Evaluation of Educational Effectiveness

by Julie S. Vargas

During the past two administrations, the federal government has spent billions of dollars on education. The results of this infusion of money have been disappointing. Two general approaches to improving education lie before the new administration: to increase levels of support, and to put existing resources into programs that are more effective than current ones. While most educators would argue for the former (myself included), political realities may make substantial increases in educational budgets unfeasible. Policies adopted by the new administration could, however, make a real impact on the educational performance of youth across this country with the second strategy. To do this, policy makers must have a clear notion of effectiveness.

Hierarchy of Measures of Educational Effectiveness

When comparing the performance of America's youth to that of other countries, or when documenting trends in educational performance, authors usually cite test results. Admittedly, completing test items correctly is not the end-all of education. Nevertheless, a well-educated youth should have no difficulty in solving the kinds of problems found on standardized tests. Though quality education must go beyond enabling students to pass tests, an effective program should, at the very least, improve standardized test performance over current gains.

Using, for the moment, performance on standardized tests as a measure, one can judge effectiveness on several levels. Table I shows a hierarchy of educational effectiveness measures, from the least rigorous (level 1) to the most thorough (level 5).

Copyright 1988 Julie Vargas

Level One

Evaluation at level one is casual at best. By looking only at a postmeasure, one cannot tell whether, in fact, any improvement was made. It would be possible for a program to decrease student performance and still show satisfactory postmeasures!

Level Two

Level two at least shows whether students gained in the skills measured. But it provides no way to tell whether they would have gained as much or more from alternative experiences.

Level Three

Level three begins to provide evidence for or against a particular educational procedure. By comparing gains under a particular program with trends before, or with gains under other programs with similar populations, one can see whether or not that program made an impact.

Some evaluation at level three can be done with statistics that are routinely gathered by most school districts. The following graph, for example, plots scores

readily available: the mean scores made by students in the Sacajawea School in Great Falls, Montana against the mean scores for the whole school district (including Sacajawea). In 1974, Sacajawea School began a Precision Teaching project, the heart of which was individualized progress and daily measurement. Each day, in each academic subject, students took short, timed quizzes according to their individual progress through the curriculum. They scored and graphed their own performance, progressing to the next quiz when they met the required criterion. Students who started the program as first graders originally scored, like their peers, at about the 70th percentile in reading and the 65th in math, but increasingly gained in their achievement until, as fourth graders, they scored *19 to 44 percentile points higher* than students who had not benefited from Precision Teaching. While the increase in scores by itself is impressive, the comparison data are needed to show that not all scores in Great Falls were improving over those same years, that is, that something different was happening in the Sacajawea School (Pre-

world of difference. The behavioral science approach measures that difference, and tells you how to get it. (Having now read the feedback, what do you think of your answer?)

Role of the Computer

The classroom teacher can implement many behavioral methods, as illustrated by some of the articles in this issue. Some require special apparatus that may be needed to obtain full benefits, especially when the teaching program moves out of the laboratory and is extended to many learners. Of these special kinds of equipment, the computer plays a significant and complex role.

The computer is significant because it can implement the heart of the behavioral arrangement. It is the consequences of our actions that determine what we learn and repeat. In a teaching game presented on the computer, if the learner (using a hand-held computer mouse) puts the record on the phonograph whose color matches the sample, then the song will play another verse. This arrangement, used properly, can teach color matching. In this way the computer is acting as a teaching machine.

But the computer can do more than this. It can also give students initial placement tests, and it can keep track of what the student is learning. Such information helps the lesson be adapted to the particular learner, and is also used in records consulted by administrators and researchers.

These roles overlap, and it's easy to get them confused with one another, or even to argue about which is "really the right use of the computer." They are all important, but it should be remembered: the teaching is done by the way in which the consequences of the learner's actions are arranged by the computer program.

Resistance to Effective Instruction

All these possibilities are expanding rapidly as the costs

of hardware and software continue to drop, and smaller machines become able to do incredibly complex things. Powerful teaching technology is becoming cost-effective. But many problems still need to be faced; we do not want this technology to merely enrich the advantaged.

Serious resistance to powerful instructional methods is to be found within the educational establishment itself—quite apart from whether machines are employed or not. This resistance seems not to be altered even when objective scientific data is presented which documents the superiority of a particular method. (Deborah Shanley's article documents this sad fact.) Interests are threatened, or are perceived to be in danger. It must be acknowledged that new methods of instruction will probably require adjustments in the current career structure of the teaching profession, and make unfamiliar demands upon those who would stay in this field. Can the diverse interest groups with a stake in the action construct a way to work together?

The Federal Challenge

The federal government faces a particular challenge for the very reason that its role in our pluralistic society is indirect. As a partner with state and local agencies, the federal presence has acted to provide models, to demonstrate, to compensate and to encourage. A strategy is needed which will promote cooperation at all levels in undertaking the tryout and evaluation of promising instructional methods and systems, and the adoption of those which work. This process should remain open enough to permit continuing evolution as results are recorded, data collected, and new facts about learning emerge from the laboratory.

To address this challenge successfully would be to take a historic step in improving the capacity of our educational system, so that equal access to opportunity would reach those who today are still isolated and alienated in America.

New Organizations In Educational Technology

NSPI

Name: National Society for Performance and Instruction. Formed: 1962.

Purpose or emphasis: Dedicated to increasing productivity in the workplace, through the application of performance and instructional technologies. Instruction, job design, or motivation, shown to be effective by scientific measurement.

Membership: 3,500

Structure: Annual Meeting; over 40 Chapters

Publications: *Performance and Instruction Quarterly*; *Performance Improvement Quarterly*

Address: 1126 Sixteenth Street NW, Suite 102
Washington, D.C. 20036

Phone: (202) 861-0777

AERA

Name: American Educational Research Association. Formed: 1916.

Purpose or emphasis: Improving the educational process through the encouragement of inquiring into education, the dissemination of research results and their practical application.

Membership: 1,400

Structure: Annual meetings

Publications:

American Education Research Journal; *Review of Educational Research*; *Educational Evaluation and Policy Analysis*; *Journal of Educational Statistics*; *Review of Research in Education*

Address: 1230 Seventeenth Street, NW
Washington DC, 20036

Phone: (202) 223-9485

ABA

Name: Association for Behavior Analysis
Formed: 1974

Purpose or emphasis: The experimental, theoretical, and applied analysis of behavior.

Membership: 1,800

Structure: Annual convention

National, Regional, State and Local chapters; Interest Groups: Education, Gerontology, Social Action, Clinical, Computers, Direct Instruction, Experimental, Organizational, Verbal Behavior

Publications:

The Behavior Analyst; *Journal of Applied Behavior Analysis*

Address: Department of Psychology
Western Michigan University
Kalamazoo, Michigan 49008-5052
Phone: (616) 383-1629

cision Teaching Project).

Level three evaluation does not necessarily require a comparison group; a comparison track record will do. In West Virginia, for example, a court awarded the costs of private schooling in Virginia to the parents of a retarded boy, based upon evidence that his education in the Cavell County West Virginia School System was inadequate. No "comparison student" existed. Instead, the student's lack of progress over three years in the West Virginia school was compared against the gains made subsequently in the private school. (*Board of Education of the County of Cabell v. Paul Dienelt, et. al.*)

Level Four

In addition to student gain, a program should be cost effective. A program may produce gains, but at a cost per student which does not justify the expense. Take, for example, Title I funding (now Chapter 1) under the Elementary and Secondary Education Act of 1965. Most school districts have a limited budget for compensatory education. For maximum impact, those dollars must go to programs which provide the most benefit to the largest numbers of students.

Kansas City is one district for which

data are available on costs and student gains in Title I programs. In the late 70's and early 80's two teaching approaches were used with Title I students: programmed instruction learning centers and tutorial groups of seven students per teacher. Students in both programs made gains. But the 47 learning centers handled close to 200 students each and still produced over twice the gain of Title I programs nationwide (1.6 months average increase per student for each month in the program compared with .7 nationwide). At the same time, the costs per student in the center were less than half those of other Title I programs in the state, making a fourfold increase in the amount of learning purchased per dollar spent (Weinstock, 1984).

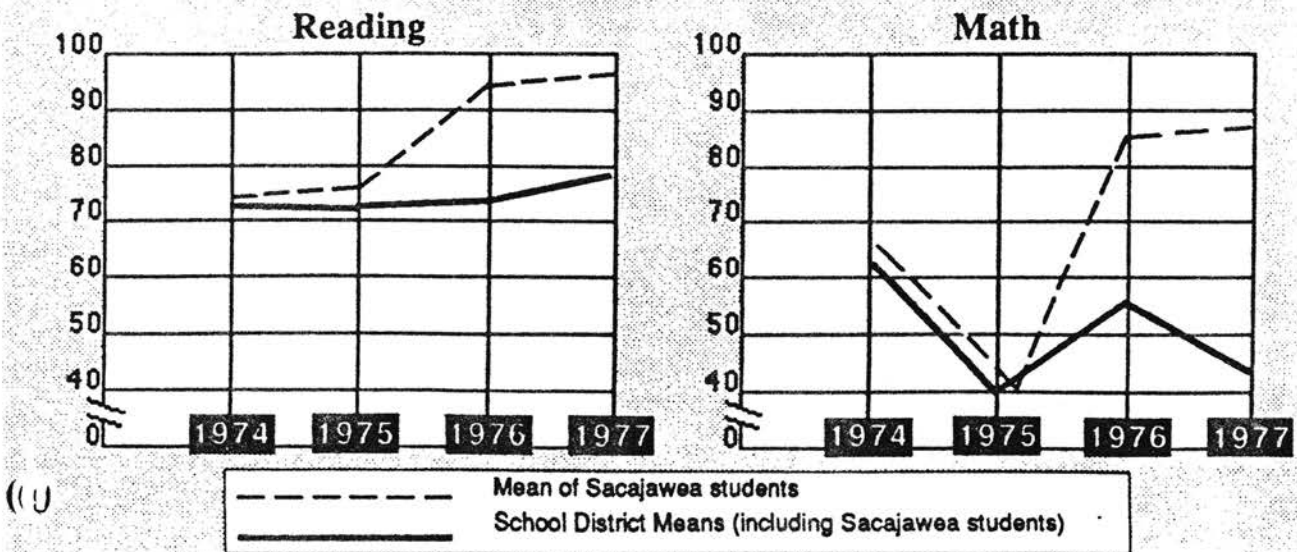
Level Five

Level five considers the long-term national impact of educational practices. The Children's Defense Fund recently prepared a table of "Successful Preventive Investments in Children." According to the organization's data, "each dollar invested in quality preschool education returns \$4.75 because of lower costs of special education, public assistance, and crime" (Children's Defense Fund, 1988). The Children's Defense Fund does not specify which preschool programs

produce those results, but the kind of data they cite are the level five data needed to realize a high return on educational investments.

So much for achievement test results. The levels of measurement apply equally well to other kinds of behavior, including those classed under the titles of "appreciation," "attitudes," or "affective behaviors." Consider attitudes toward science. At level one, researchers would collect information on how well students liked a program—perhaps with a questionnaire or with a more convincing measure such as how many stayed after school to continue working on science, or the number that voluntarily watched science programs on TV or bought science magazines. At level two, researchers would present evidence that students in the experimental program exhibited more of these pro-science behaviors than they had before their participation in it. At level three, evaluators would track students' science appreciation for several years. A jump in science activities the year the project started would be evidence for program effectiveness. Alternately, at level three, evaluators would show that peers in other programs did not gain as much in appreciation as those in the target program. At level four, researchers

Sacajawea Students' Performance Compared with the School District Means



would calculate costs of the achieved gains in "science appreciation" and compare the gains per dollar from other approaches. And finally, at level five, they would assess societal benefits, not so much by dollars saved but by such indices as the gains in numbers of students majoring in science or choosing science careers.

The levels of educational effectiveness are cumulative. Each requires at least the data at the lower levels. For example, level five requires some measure of how, years later, those students would have performed had they not entered the evaluated program, in addition to documentation of dollars spent per student gain at level four. This in turn requires the evidence from levels three through level one.

Lack of Evaluative Rigor by the Department of Education

The Department of Education seems to be interested in program effectiveness. After all, during Secretary Bennett's tenure, the department published four booklets in a "What Works" series. One would expect that educational practices endorsed by the federal government would have documented effectiveness at least at level three. Not so. Documentation of effectiveness is sparse at best. For example, the booklet called *Schools that Work* highlights 23 "profiles" to serve as exemplars of schools and approaches that "work." While all of the profiles may describe worthwhile endeavors, the booklet fails to give convincing evidence of that fact. Of the 23 descriptions, only *four* provide both pre-and post-measures. Of the remaining 19 profiles, nine give some kind of post measures, and *ten offer no data whatsoever*. None of the 23 profiles gives evidence of effectiveness at levels four or five. Far from showing "what works," the book gives the impression that enthusiasm, rather than results, is what the government endorses.

Proposals for the Future

Particularly now, with the increasing educational skills demanded by the workplace, America cannot afford to

fail in educating all its youth. With every student who fails to obtain a job, there is one taxpayer less and one dependent more. Policy-makers at all levels of the educational system could make a major impact on our youth by shifting resources from ineffective educational practices into programs which "work" at least at level four of this hierarchy. In the past, millions of dollars have gone into programs which have not produced results, and funding has been cut off from programs that work (see the article by Watkins p. 7). While research on effective educational practices should continue, the government could make a dramatic improvement in education with the existing principles and programs which are known to make a difference— an impressive difference. Many are documented in this issue of *Youth Policy*. For the education of this country to dramatically improve, programs which receive continued funding and administrative support must be cost-effective (level four) and moreover, in the long run (level five), they must provide a good return for the educational dollars spent.

Dr. Julie S. Vargas is a professor at West Virginia University. She is the author of two books in education and of several

programmed instruction series in reading and math, including 30 lessons of Computer-Assisted-Instruction in reading comprehension.

References

Board of Education of the County of Cabell Versus Paul Dienelt, Gary Dienelt and Cynthia Dienelt, U.S. District Court South District of West Virginia, Civil Action # 3:860287 (to be published in the U.S. Court of Appeals Fourth Circuit # 87-1083.)

Children's Defense Fund, A Call for Action to Make our Nation Safe for Children: A Briefing Book on the Status Of American Children in 1988, Children's Defense Fund, Washington, D.C., p. 39.

Great Falls Public Schools, Great Falls, Montana. Precision Teaching Project, Sacajawea Plan.

Weinstock, Ruth, "A Title One Tale: High Reading/Math Gains at Low Cost in Kansas City, Kansas," Phi Delta Kappan, May 1984.

Table I

Hierarchy of Measures of Educational Effectiveness

Level One

Postmeasure only.

Level Two

Documentation of gain. [Premeasure and postmeasure provided.]

Level Three

Evidence of superiority over usual methods. [Not only must gains be documented, the gains must surpass those of students not in the program.]

Level Four

Evidence of cost effectiveness. [Not only must students learn more effectively than those not in the program, but the increased gain per student must be worth the money expended.]

Level Five

Evidence of the long term total impact on society, usually in dollar costs or savings. [Educational programs should ultimately contribute to the gross national product by producing wage earners and cutting down on the costs associated with unemployment, poor personal health behaviors, and with crime and delinquency.]