DOCUMENT RESUME

ED 124 599 TM 005 359

AUTHOR Reinstein, Barry J.
TITLE Public School Perspectives on the Uses of Large-Scale Testing Programs.
PUB DATE [Apr 76]
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; Academic Standards; *Educational Assessment; Elementary Secondary Education; *Information Utilization; Inservice Teacher Education; Instructional Improvement; Program Evaluation; Resource Allocations; *School Districts; Socioeconomic Status; *State Programs; Statewide Planning; Testing Problems; *Testing Programs; Test Results

ABSTRACT

The more commonly cited uses of state and local district assessment programs are addressed. The implications of seven proposed uses of state assessment data are reviewed point-by-point: (1) allocating state grants-in-aid to alleviate weaknesses in instructional programs; (2) designing instructional support programs for teachers; (3) developing state planning statements and priorities; (4) revising state minimum standards for schools; (5) reporting and making recommendations to the legislature; (6) determining if students are acquiring "survival level" skills or "minimum competencies;" and (7) determining the extent to which students in a state have attained the skills, knowledge, and attitudes reflected in the educational goals of that state. Next, the author comments on some uses of large-scale testing in school districts and some conditions that should be met if such uses are to be realized. Finally, several major obstacles inherent in developing measurement instruments and procedures in areas other than reading, language, and math are discussed. (RC)

Public School Perspectives on the Uses of Large-Scale Testing Programs

Barry J. Reinstein
Portland Public Schools

Large-scale testing programs are commonly defined as efforts to determine student achievement on a school, district, state, or national basis. Further, the recent initiation of state assessment programs has developed as a corollary of demands for accountability in public education. As educational expenditures have risen, coupled with increased competition for fewer available dollars, the demands for educational accountability have likewise increased; so much so, that accountability has become a preeminent concern of educational decision-makers. Typically, state departments of education and local school districts have initiated large-scale testing programs for the purposes of providing the necessary data for more informed decision-making and for use in judging the effectiveness of state and local schooling efforts. My comments today will address and be limited to the more commonly cited uses of state and local district assessment programs.

I would like to begin with a point-by-point review of the implications of some proposed uses of state assessment data.

1. Allocating state grants-in-aid to alleviate weaknesses in instructional programs.

   Use of state assessment data for this purpose assumes that accurate interpretations about instructional weaknesses can be made from statewide test data.

Prepared for a Symposium on Dissemination and Utilization of Large-Scale Test Results held at the Annual Meeting of the American Educational Research Association, San Francisco, California, April 1976.

Actually, although test data are a useful indicator of program strength and weakness, only a limited number of programs can be reliably and validly measured with present instruments and resources. Allocation of state funds only to programs in which measurement is possible carries the risk of diverting resources from programs less amenable to measurement. This could work particular hardship on upper grade and high school programs where diversity and specialization of offerings may be the key to quality.

Another problem is the need to take into account the social, economic and educational factors in the adult community that in large measure appear to determine levels of student achievement. To illustrate, in the Portland district where regression equations are used to predict mean achievement scores for schools from SES data, such as median family income, percent student attendance, median-grade completed for adults 25 years and older, percent free lunches, etc., multiple correlations of .80 to .90 are commonly found. The strikingly high correlation of group mean SES and achievement data suggests there is little value in reporting and comparing achievement scores of schools, school systems, or regions in the absence of such data. Thus, communities with especially low or high social and economic characteristics cannot be regarded as having especially poor or outstanding educational programs simply because the children score low and high, respectively, on achievement tests.

A third problem is the effect, both direct and indirect, that distributing funds on the basis of test scores might have. If

low achievement were used as a measure of socio-economic deprivation, it would overlap criteria for funding programs such as are found in Title I and state financed equalization programs. Used as a measure of financial need, it would also make it financially profitable for school systems to maintain weak programs.

2. Designing instructional support programs for teachers.

State curriculum development and in-service programs could be based upon statewide test information showing comparative strengths and weaknesses among areas of learning, on the basis of performance related to standards, or on the basis of achievement trends. An important caution that should be observed, however, is that statewide data may not be applicable to any one district, so such planning should be in the form of support for those districts in which other data confirm that a general weakness discovered through state assessment does in fact apply to them.

With reference to the setting of performance standards in state assessment programs, the following should be considered. The attachment of standards to criterion or goal referenced tests, the mainstay of state assessment, is based upon the mastery concept which is most effectively applied to acquiring specific information or skills for specific purposes within finite time limits. The mastery concept does not apply nearly as appropriately, if at all, to long term developmental types of learning such as are represented in reading, writing, and math problem solving. If we attempt to set standards on this type of learning, they will almost invariably create pressures for some students and be too easily achieved by others.

Since the designation of arbitrary standards for total test or specific goal performance has many illogical and potentially damaging aspects, it is recommended that state support-service allocation not be based on performance in relation to standards. A much sounder criterion is longitudinal evidence of declining performance.

One further comment should be made about inferring need for support services from statewide test data. To deduce that a downward trend in achievement in itself implies less adequate instruction, or that increases reflect better instruction over time, is unwarranted without additional data. Changes in character of student populations can affect achievement levels in a state as well as in a locality. Achievement is also associated with general attitudes of youth. The social protests of the mid-60's and manifestations of this movement such as the drug culture, appeared to have a depressing effect on student achievement. It is doubtful that allocation of additional resources to in-service education could have prevented a decline during this period.

3. Developing state planning statements and priorities.

Planning statements of a state educational agency should reflect a support and monitoring posture rather than an instructional management intent. If uniform goals or standards for local districts are set in state educational agency planning statements, local needs and priorities may be set aside even though they are

more valid indicators of local needs. It may be appropriate to set statewide priorities, but monitoring of local performance in such priority areas should not be of such character as to force local resources and activity to be directed toward that general need unless there is clear evidence it is also a need of that local system.

4. Revising state minimum standards for schools.

Measurement procedures required for this proposed use are somewhat unclear. State minimum standards for schools should be concerned with whether or not local districts are achieving their own goals, given the assumption of local curriculum determination. No specific standards of achievement should be included in state minimum standards for schools, since a given level of performance may be excellent in one district and very commonplace in another, depending upon the social, economic, and educational condition of the community population. Even if these factors could be controlled, setting this type of standard could have adverse effects on educational programming, such as the inevitable diversion of resources toward the achievement of the standard, however difficult or even impossible it may be for some children.

5. Reporting and making recommendations to the Legislature.

This commonly cited use of large-scale test results is not helpful without defining the types of reports and recommendations envisioned. One possible type of recommendation, allocating grant-in-aid resources on the basis of test results, has already

been evaluated. Other types of reports could be purely informational.

Reports on comparative achievement among districts should be scrupulously avoided because of the ease with which such information can be misinterpreted and misrepresented for political purposes. Experience in several states reveals such misuse to be a predictable consequence of this type of analysis.

6. Determining if students are acquiring "survival level" skills or "minimum competencies".

Quite frankly, although attention to "survival skills" or "minimum competencies" is approaching epidemic proportions among state legislatures and educational agencies, I have serious misgivings about the movement and the purely political response it appears to represent. It is well to adopt a cautious if not wary attitude toward simple definitions of minimum or survival competencies when the skills, knowledge, attitudes, and values needed to survive are so dependent on individual differences. While one can sympathize with public frustration over basic skills, lack of consumer education, weaknesses in vocational training, student indifference to rights of others, and other concerns, it seems more appropriate that such concerns should be addressed by providing programs that respond to individual as well as group needs in these specific areas of concern rather than struggling to define minimum learnings that all must acquire.

7. Determining the extent to which students in a state have attained the skills, knowledge, and attitudes reflected in the educational goals of that state.

Achievement tests of the type usually found in state assessment testing cover such a limited area of state policy concerns that their value for this purpose needs to be placed in perspective. If one examines the goals of a number of states, it will be found that they vary widely in character. Some define broad areas of learning; some specify personal-social qualities that education should help citizens acquire; some refer to procedures and programs to be established; some to equity in allocating resources or providing opportunities; some to competencies that students should acquire. Even where state goals define broad areas of learning, achievement tests of the type found in state assessment programs provide such limited coverage that they have only limited usefulness in assessing attainment of such goals.

Of greater importance, however, is the inherent conflict between local curriculum determination and the assumption of a common curriculum that is necessarily embodied in statewide tests. Although this lack of congruence is probably within acceptable limits in convention-based studies such as language and mathematics, it is a great problem in such important curricular areas as science and social studies.

Uses of Local District Testing Programs

At this point I would like to comment on some uses of large-scale testing in school districts and some conditions that should be met if such uses are to be realized. It should be observed that since local school districts have the delegated authority to establish specific curricula, valid measurement of outcomes is at least theoretically possible, and this eliminates one of the major problems faced by state testing programs, i.e., the inability to collect data that adequately represents the curriculum of any particular school district. Also, since local districts are the basic unit of educational management, and evaluation is an essential function of management, the obligation clearly rests upon the local district to determine if the learning outcomes of the system are being realized. This should be the purpose of city-wide testing. Where this capability exists, it will be possible to conduct evaluation of ongoing programs, evaluation of specially funded programs, and research and experimentation. Based upon the information produced by testing for these basic purposes, management decisions can be made about program operation and resource allocation, and the public can be informed about the effectiveness of regular, special, and experimental programs.

It is necessary at this juncture to discuss some realities or conditions that must be addressed if school district testing programs are effectively to serve the uses just described. While the need for local district information about goal attainment is self-evident, school districts generally provide for it only in those areas where testing has traditionally been used: elementary reading, language

and mathematics, and a smattering of coverage of other subjects at the elementary and secondary level. This has primarily been due to several major obstacles inherent in developing measurement in other areas of learning. I should like to comment briefly on these obstacles, for in a way it is unprofitable to speak of uses of tests when so many of those uses cannot be realized because of the inability of school systems and test publishers to produce the tests necessary to provide total curriculum coverage.

A first obstacle is the difficulty of producing well-defined outcome statements in local districts. Efforts I have observed to define behavioral objectives have been particularly disappointing in quality and utility for instructional planning and evaluation. In the Portland area we have spent four years developing clearly stated learning outcomes (called course goals) in twelve major areas of instruction (Art, Biological & Physical Sciences, Business Education, Health Education, Home Economics, Industrial Education, Language Arts, Mathematics, Music, Physical Education, Second Language, and Social Science). This is a comprehensive and carefully classified set of learning outcomes and its purpose is to enable teachers to select rather than create such statements. This certainly does not solve the problem of making the use of goals operational in instructional planning and measurement, but it is an important first step.

A second obstacle is the diversity of philosophies and instructional approaches encountered among teachers and instructional specialists. This is a special problem in the sciences and social studies, but is something of a problem in almost every field of learning.

One aspect of this complex problem in the science area is failure to distinguish between processes of acquiring, organizing, and interpreting existing scientific information and processes of inquiry employed by scientists to discover and validate information. Another is the senseless argument between advocates of process and product learning. Another, in social studies, is the failure to acknowledge that concept learning must be defined by the informational loading given the concepts. Still another is the failure to acknowledge that outcomes are just as clearly needed and useful in interdisciplinary planning as for planning within a structured field of learning.

A third major obstacle to extending measurement to all fields in which it is needed is the rigorous and resource-consuming requirements of local test development; yet local test development is essential if validity is to be achieved within the framework of local curriculum autonomy.

A fourth obstacle is the failure of many teachers to distinguish between means and ends of instruction; a problem that has been ingrained by traditional dependence on texts and other support materials. This deters teacher ability and willingness to define measurable outcomes of learning.

This catalog of obstacles is intended to convey the seriousness of problems to be faced if valid and reliable local measurement is to be achieved in enough areas of learning to enable testing programs to be seriously regarded as a tool for local management decisions. The problem obviously compounds for state measurement programs based on

assumptions of common goals when in reality common goals do not, and according to principles of local control, should not exist.

Finally, in Oregon, the new state minimum standards call for school systems to define goals for all courses offered, and this, I believe, is a big step toward enabling valid local measurement to be developed. But it will not itself carry us past the obstacles I have outlined; for almost infinite variation of course goals are still permitted within these guidelines. Answers, if they are to be found, will probably lie in devising more effective processes for making informed and defensible judgments about what specific learning the courses and programs of school systems should be held accountable to produce.

jns