DOCUMENT RESUME

ED 260 494 EA 017 921

AUTHOR Hansen, Kenneth H.
TITLE Statewide Assessment: Convergent Principles, Divergent Policies. Discussion Draft.
INSTITUTION Northwest Regional Educational Lab., Portland, OR. Northwest Center for State Educational Policy Studies.
SPONS AGENCY Chief State School Officers of the Northwest and Pacific.; National Inst. of Education (ED), Washington, DC.
PUB DATE Jun 85
CONTRACT 400-83-0005
NOTE 15p.
PUB TYPE Viewpoints (120) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Agency Cooperation; *Educational Assessment; Educational Philosophy; Educational Policy; Measurement Objectives; Policy Formation; Program Costs; *Program Development; State Departments of Education; *State Programs

ABSTRACT

State education agencies are in increasing agreement regarding the basic principles that should underlie state educational assessment programs, though some areas of divergent opinion remain. The agencies generally accept the ideas that comparisons between states will be made, that assessment programs should serve multiple purposes, that meaningful comparisons cannot be made without knowledge of the contextual factors affecting the provision of education, that a large number of indicators must be employed to fully display an educational system's status, and that assessment programs are bound to be costly. The agencies disagree over the uses to which assessments should be put, the educational outcomes that should be measured, the extent to which various contextual factors affect education, the specific indicators that should be used, and how the costs of assessment are to be controlled and allocated. Agencies need to develop policies for assessment that will resolve all of these questions in ways that are appropriate both to the individual characteristics of the state and to the demand for data that will allow meaningful assessment and comparison on a national level. (PGD)

STATEWIDE ASSESSMENT:

CONVERGENT PRINCIPLES, DIVERGENT POLICIES

Discussion Draft
Prepared for the
Chief State School Officers of the Northwest and Pacific

Kenneth H. Hansen
Principal Author

Northwest Center for State Educational Policy Studies
Northwest Regional Educational Laboratory
300 Southwest Sixth Avenue
Portland, Oregon 97204

June 1985

The work upon which this publication is based was performed pursuant to Contract 400-83-0005 of the National Institute of Education. It does not, however, necessarily reflect the views of that agency.

STATEWIDE ASSESSMENT:

CONVERGENT PRINCIPLES, DIVERGENT POLICIES

Introduction

The tremendous current interest in statewide assessment and evaluation programs may have been triggered, as some believe, by the "wall chart" controversy; it certainly has been given importance by the strong impetus coming from the Council of Chief State School Officers through their proposal to establish a national assessment center; but its greatest support, it may be maintained, has come from the interests and activities of the nation's state education authorities. It is the states themselves, over the past few years, with their concern for academic excellence, educational reform, and instructional improvement, that have been in the forefront of the drive to strengthen and refine statewide assessment programs.

Disagreements about the best way to get the job done still abound, but there is developing a discernible degree of consensus on certain basic principles--call them, perhaps, philosophical assumptions--that should underly the statewide assessment movement. Conversations with Chief State School Officers and their staffs, publications of the various state education agencies, and reports appearing in current education literature all suggest that there is at least some convergence of thinking on the topic. At the same time, and from the same sources, it is equally apparent that there remains a very considerable divergence of beliefs about the appropriate educational policies which should be adopted to give proper direction to the more-or-less-agreed-upon ends.

This paper addresses the resultant problem of convergence and divergence for very practical reasons. While basic principles and philosophies must undergird sound policies, they do not in themselves constitute educational policy. Agreeing upon what ought to be and setting concrete policies calculated to give organizational direction toward the desired end are, of course, two different things. Nevertheless, it is possible to consider some of the policy options that are available to carry out accepted and adopted principles, then we can not only support the measure of consensus that is developing within the educational community, but also give support to the uniqueness and independence of the several states through encouraging the analysis and development of the policy options specifically suitable for a given state.

Therefore, in the sections of this paper which follow, a number of significant issues respecting the establishment, operation, and strengthening of statewide assessment systems will be explored, each with two subsections: first, a statement of the convergence of opinions that seems to be developing around the issue; and second, a brief look at the divergent options which are being proposed as appropriate to move the state system in the direction desired.

Convergence and Divergence: Key Issues

1. Comparisions Are Inevitable

Convergence. After the initial shock at seeing the bald display of certain data about state educational systems in the controversial "wall chart," and after disputing many of the alleged "facts" or criticizing them as inadequate or misleading, educators have generally come to a somewhat calmer mood of acceptance or at least resignation. There seems

to be a growing body of opinion that state-by-state comparison (and similar comparisons within the state at school-district and school-site levels) are really inevitable, and might not be a wholly bad thing at that. Thus far, then, some degree of consensus. But what policy direction should be established with respect to these comparisons?

Divergence. Some educational decision makers would still support a policy in direct opposition to any state-by-state comparison. Although their position may be both reasonable and tenable, granting them the validity of their own prior assumptions, from a purely pragmatic standpoint proponents of this resist-it-all view are most likely to find themselves in a minority, and a relatively ineffective one at that.

Thus, a majority of state-level educational policy developers--state boards, Chiefs, and appropriate staff members--would appear to be supporting a policy of accepting the seemingly inescapable fact that we're going to have comparisons, so let's improve the data! But even "good" data are of little value in themselves. First, from a public accountability point of view, they have to be clearly understood and correctly interpreted. Second, from an internal management view, they are good only if put to good use. Thus, acceptance of a policy which supports state-by-state comparisons leads logically to support of a policy which requires a conscious program to improve the entire statewide assessment and evaluation program for state purposes but from a correlative national perspective.

It becomes fairly clear that, short of outright rejection of all between-state comparisons, the policy options which emerge all tend to call for an increasingly complex commitment of time, effort, and money to develop a comprehensive state/local program which will make the comparisons fair and meaningful.

2. Purposes Are Multiple

Convergence. Whatever value may be attached to the employment of statewide evaluation and assessment programs for comparison purposes, there is growing commonality of agreement that such use is only one of many, and perhaps a minor one at that. The data made available from these programs serve to inform the many publics how well or poorly the schools are doing; to help the schools monitor their own programs and their students' progress; and above all, to provide better data for more informed decision making, which will in turn improve the schooling progress. A large order, to be sure, but one that an adequate evaluation and assessment program ought to satisfy. This is the general theme of the covergent thinking on multiple usage of the data.

Divergence. Policies designed to carry out the basic principle of multiple-use represent a very wide range of policy options. One option reflects the view that education is best improved by raising test scores and upping average grade-level achievement scores of all students. Strong academic emphasis on specific factual learnings is encouraged, and scores probably will go up. Another option for priority emphasis in educational improvement might quite logically focus on certain kinds of intellectual skills, such as reasoning ability, ability to see relationships and draw inferences, skill in applying what has been learned, for example, in language arts or English courses, to actual speaking, writing, and listening; basic factual knowledge would, in such a case, be deemed of less importance.

If an educational policy option chosen by the appropriate decision makers is one which emphasizes the overall personal and social development of children and young, still different data will be needed, and formal testing programs may be generously supplemented by other carefully designed but less rigidly structured assessment schemes.

Not only are there varying beliefs about how the evaluation and assessment data can best be used for purposes chosen from among a multiplicity of possible uses, but there are policy options which must be exercised in determining who will be the primary user of the data. If they are for use primarily by the local school district or individual schools and teachers therein, different kinds of test instruments and different methods of aggregating and reporting data are needed than would be necessary for primarily state-level use.

When reference is made to the "use" of the data, the assumption is generally that they will be used by decision makers for making decisions--a statement that would appear somewhere between self-evident and redundant were it not for the embarrassing fact that many acquired data are often not so used. They are just collected, and nothing much happens.

So, multiple kinds of data for multiple uses for multiple groups of decision makers for making multiple decisions for effecting multiple improvements in education would seem to introduce so many combinations and permutations as to be hopelessly baffling. The apparent confusion, however, can be partially straightened out by exercising policy options quite readily available.

First, for an example, if it is firmly established policy to determine first who needs what data for making what specific decision, much wheel-spinning can be avoided. Certain kinds of very specific data from each student are needed for diagnostic and teaching purposes by the classroom teacher; other kinds of data for groups of students or for specific program elements are needed by the principal for effective instructional supervision; still other data are needed by the central administration for monitoring and program review, and especially for program planning and improvement. Yet, only part of these data, and much of that only in aggregated form, is needed by the SEA. The other side of the coin: the SEA needs kinds of data for which the individual school may have little use. The key to the policy formulation? Matching data, user, and purpose.

One further illustration. If the primary policy concern at the state level is collecting and disseminating data on educational achievement which will report <u>what is</u>, one range of data-based information is needed. If the primary use of these data, however, is intended to reveal <u>trends</u>, rather than just <u>present status</u>, so that these trends can be used for making judgments about program adequacy and needed program change, then other or additional kinds of data may be needed.

In brief summary of this look at the paradox of seeming agreement that a multiplicity of educational purposes is to be served by state evaluation and assessment programs, and the divergent opinions--and hence divergent policies--that stem from this basic agreement, an important point is again illustrated: it's the <u>policy</u> that makes the <u>operating difference</u>.

3. Background (Input) Factors Are Crucial

Convergence. No more common--or more justified--rejoinder is heard from educators who have been "burned" by unfavorable comparisons with other schools (local-to-local or state-to-state) than this: "They didn't take into account our special situation! We have more families from below the poverty line; more students from disturbed home environments; greater concentration of urban problems and urban overburden (or conversely, greater rural isolation and population sparsity); we have less state/local support; our state (district) is in an economic downswing; and you've no idea of the magnitude of our bilingual problems!"

All of these are appropriate and legitimate responses, for knowledge of background factors is absolutely necessary for making intelligent and fair comparisons between any different school entities--specific attendance centers, local districts, or states. So, there is remarkably convergent thinking on this point; background factors must be considered. But then the agreement begins to fall apart.

Divergence. There is painfully little agreement among educational decision makers on either the definition of the various factors or the actual significance they have in determining the success or failure of either individual students or of particular programs. What is meant by "dropout rate"? What factors constitute "school environment"? And how do these background factors relate directly and unequivocally to student performance and successful instructional programs?

It seems likely that no one would be so optimistic or so bold as to hope that common agreement could be reached among educational decision makers, professionals and concerned laypersons alike, on either the

definit ons or the specific effects of these background factors, but carefully-thought-out and clearly-articulated policies can help bring some order to the present state of confusion.

Specificity is extremely hard to come by at this point, because of the infinite variety of differences, exceptions, applications, and interpretations which revolve around any one of the background elements. Certainly, however, a reasonable place to start would be to have any policy-making body--local or state school board, legislature, Congress, or whatever--adopt as part of any relevant policy a clear definition of each factor as it is used with their own constituency. Rather than seeking total agreement on terms, that is, seek only clear definition. Then, when comparisons are made, at least the language used in these comparisons will reflect known disparities and disagreements.

Similarly, with the problem of the specific educational effect that these background factors may have on educational programs, or on individual student success or failure. Since we don't "know," in any absolute sense, policy statements which reflect educational decisions made on the basis of thoughtful assumptions about the relationship of, say, the percentage of children from poverty-level families to academic achievement might well contain clear wording about the assumptions that are being made and the relationships that are believed to exist. In a word, the solution lies not in certainty, but in candor.

4. Multiple "Indicators" Are Needed

Convergence. Consensus is rapidly coalescing around the principle that no significant assessment of educational success or failure can be fairly made using only a limited number of indicators such as SAT scores, grade-level achievement in terms of established norms, dropout rates or the like. A fairly large number of indicators--variously categorized under such labels as "input," "process," and "outcome" measures--must be employed. But beyond that general principle, agreement begins to disintegrate.

Divergence. The divergence of belief--and hence of policy--is quite wide. Some education decision makers would maintain that it is only clear-cut measures of academic achievement which are of fundamental importance, and the only ones which will be understood anyway by the general public; therefore, as a matter of policy, these are the ones which should be used. Others will insist on using an extremely large number of indicators of every sort, such as are embodied in the lists which are beginning to appear in print both at state and national levels. Their policies, likewise, naturally reflect this belief in using almost innumerable indicators.

Actual policy options available here are very difficult to formulate precisely. Rather, some policy considerations may be offered.

First, the indicators employed will probably be most useful to the extent that they bear a close relationship to the adopted goals and objectives of the state/local system. We will want to be looking for measures of the things we deem most important as outcomes of the educational system.

Second, the most useful input indicators will likely be those which most particularly describe the actual conditions which exist and which have the clearest observable or demonstrable effect on school programs and student performance. Both out-of-school and in-school indicators are needed: not only such factors as parental occupation and socioeconomic status, language backgrounds, special education and advanced placement populations and the like, but internal school factors such as instructional time available, courses offered and completed, extracurricular participations, and other data which reflect what the instructional program is actually like, are of greatest importance.

Finally, although every factor imaginable could be calculated to have some degree of importance, the indicators must be manageable in number, not so overwhelming that energy and attention are dissipated, time wasted, and money better devoted to other aspects of the educational program unnecessarily expended. Data can inform, but they can also--unfortunately--be used to confuse or dissemble.

5. Investment Is Needed

Convergence. While agreement on many of the points discussed in this paper may be elusive, and what consensus reached relatively fragile, there is very clear unanimity on one point: state evaluation and assessment programs are bound to be costly. Limiting these costs--both in money and time--and dividing them fairly between the state and local education authorities will require some clear policy decisions and some painful priority-setting.

Divergence. As opinions diverge and policies differ, the questions that come to the forefront do not appear to be wholly the old state/local controversies and tensions over "turf" or "control" or "power" or "who's going to pay?" Perhaps we are beyond that stage; although these remain issues, the essential concerns are true policy questions. For example, if the state education authorities should decide to give a single statewide test for a subject or group of subjects or for a grade-level or group of grade-levels, on the grounds of needing uniformity for reporting purposes, they would commit themselves to a great expense in time and money and lose out on reaping the values of different kinds of local tests given for different local purposes.

Alternatively, the state may choose to encourage the local districts to develop and carry out their own individual testing programs, hoping that there will be enough common data emerge to allow for statewide aggregations and interpretations of these data. This option, however, overlooks or blurs the distinction between local purposes and state purposes. Local districts need data on each student for purposes of student diagnosis, grouping decisions, guidance and counseling, and selecting students for inclusion in special programs, as well as for formative and summative program evaluation. The state has primary need for aggregated, not individual, student performance results, again for program evaluation but also for survey assessments, and in some states, for setting state standards for student performance.

One way around this difficulty is for the state to offer technical assistance to local districts in test selection (or construction) and use in order to increase the degree of conformity without imposing regimentation. The state would then restrict the actual state-required testing

program to types of tests limited to multiple matrix, cyclical, sampling testing which cover a lot of ground with technical adequacy, but which do not require every-year testing or every-student coverage, since individual student performance is not really the state's concern.

There are many technical questions regarding the kind of state tests required, the frequency of testing, the sampling techniques to be used, and the time that can be legitimately devoted to testing purposes. The trick of employing such technology, of course, is not to allow the technology to become the policy, but to inform the policy decisions.

In Conclusion

There is developing a commendable solidarity of support for state evaluation and assessment programs, with a quite discernible convergence of opinion on some basic philosophies and principles, but with wholly understandable divergencies in specific approaches. What actually happens next will depend on directions chosen--policy options to be clearly articulated, carefully chosen, and vigorously carried out.

Doc. 1422s