DOCUMENT RESUME

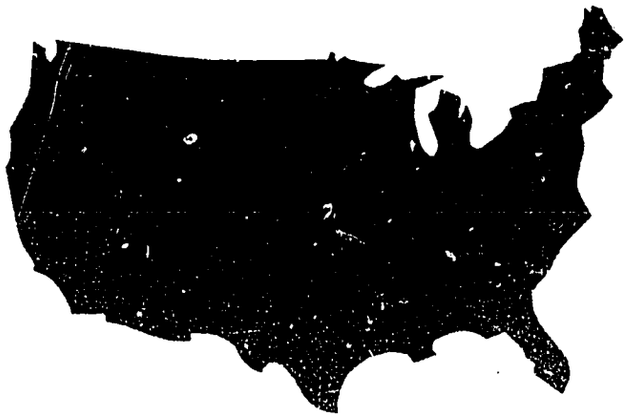
ED 070 172 EA 004 714

| | |
|---|---|
| TITLE | National Educational Assessment: Pro and Con. |
| INSTITUTION | American Association of School Administrators, Washington, D.C.; National Education Association, Washington, D.C. |
| PUB DATE | 66 |
| NOTE | 56p. |
| AVAILABLE FROM | Publications-Sales Section, National Education Association, 1201 Sixteenth Street, N.W., Washington, D.C. 20036. (Stock No. 382-11768, $1.00, Quantity Discounts) |
| EDRS PRICE | MF-$0.65 HC Not Available from EDRS. |
| DESCRIPTORS | Achievement Tests; Comparative Analysis; Comparative Testing; Curriculum Evaluation; *Evaluation; Evaluation Methods; National Competency Tests; National Norms; *Sampling; *Testing; *Test Interpretation; Test Validity; *Test Wiseness |
| IDENTIFIERS | *National Assessment of Educational Progress |

ABSTRACT

In this publication, designed to serve interested laymen as well as educators, various authors explore the viewpoints of the proponents and the opponents of the National Assessment Program. In their analysis of assessment and its related issues, these authors attempt to provide information that could serve as a basis for an objective consideration of the limitations, possibilities, and implications of the Program. Central to the discussions covered in the six papers is the question of the use of tests as a means of carrying out the assessment goals and the implications of having a pilot program run by a private concern such as the Carnegie Corporation. The presentation concludes with a transcript of the White House Conference assessment debate (July 20, 1965), the resolution of the Council of Chief State School Officers on measurement and reporting of educational results (November 1965), the ASCD Executive Committee statement of guidelines for national assessment of educational outcomes (January 1966), and the AASA Resolution concerning national testing and curriculum (February 1966). (Author/EA)

# NATIONAL EDUCATIONAL ASSESSMENT: PRO AND CON



Published by

American Association of School Administrators

and the

Organization Relations Division

and

Publications Division

of the

NATIONAL EDUCATION ASSOCIATION
1201 Sixteenth Street, N.W., Washington, D. C. 20036

Single copy, $1. Discounts on quantity orders: 2-9 copies, 10 percent; 10 or more copies, 20 percent. Orders accompanied by payment will be sent postpaid. Shipping and handling charges will be added to billed orders. All orders of $2 or less must be accompanied by payment. Order from Publications-Sales Section, National Education Association, 1201 Sixteenth Street, N.W., Washington, D.C. 20036.

# Contents

# Foreword

National assessment of educational performance is currently one of the most potentially significant and actively debated issues in American education. Unfortunately some of the recent debate has shed more heat than light on the subject. This publication is an attempt to provide enough information on and analyses of assessment and related issues to serve as a basis for objective consideration of the possibilities, implications, and problems posed by national educational assessment. It is designed to serve both educators and interested laymen. The views presented include those of proponents, opponents, and some in between.

Readers are urged to use this publication as a basis for applying reason rather than rhetoric to this important issue.

# National Educational Assessment: We Badly Need It

says FRANCIS KEPPEL

American education today is woefully short of the basic information needed to carry forward our many educational purposes, to set sound goals, and to work together to reach them.

The U.S. Office of Education, for example, can report on all sorts of things about education: how many teachers we have, how many school children, how many school buildings, and possibly whether the buildings are painted or not. But as yet we do not know how much our children really know, the subjects in which they are strong or weak, the relation between income levels and learning, or a host of other matters.

One of the topics for discussion last summer at the White House Conference on Education was a proposal to assess our nationwide educational performance in order to acquire the information we need. One might have predicted a calm discourse by educators on how, where, and by what means to proceed with this assessment. Instead, the discussion turned into controversy, and it is not over yet.

Thus far, educators have become embroiled not in the need for *assessment* but in the evils of *testing*. Some have said, and still say, that we have more tests right now than we know what to do with. Others have declared that national testing would ultimately force conformity, or worse, upon all our schools.

The mischievous word here is "testing," and it is this that has generated so much heat. A few comments may help lower the temperature a bit.

Personally I am vigorously opposed to any mammoth national testing of all children. I do not think it is either justified

---

Dr. Keppel, Assistant Secretary for Education, Department of Health, Education, and Welfare, served as U.S. Commissioner of Education from 1962 until January 1966; he was formerly dean of the Harvard University School of Education, 1948-62. This article is reprinted from the *NEA Journal*, February 1966.

or necessary. At the same time, we do need a far better reporting system than we have ever had on the quality and the progress of American schools.

"Assessment" does not necessarily imply national testing. Today, the science of statistical sampling is well advanced. It can produce answers accurate within 5 percent, even from very small samples. Although we are accustomed to statistical sampling in public-opinion polls, we have hardly begun to employ it in education.

Because I believe that the theory and technology of statistical sampling have developed to a point where they can facilitate educational assessment, as U.S. Commissioner of Education I encouraged the Carnegie Corporation of New York, with private funds, to explore the matter. The Corporation has set out to determine whether a good educational reporting system can be developed by taking samplings of the knowledge of school children—perhaps only 1 percent of the children, but certainly much less than 10 percent.

The Carnegie Corporation has already organized a committee headed by Ralph W. Tyler of the Center for Advanced Study in the Behavioral Sciences. One of its first tasks has been to set forth the purpose of a national assessment program. The committee has said:

> A well-conceived and well-executed assessment would, it is hoped, serve several important purposes.
>
> First, it would give the nation as a whole a better understanding of the strengths and weaknesses of the American educational system. Thus, it might contribute a more accurate guide than we currently possess for allocation of public and private funds, where they are needed, what they achieve, and decisions affecting education.
>
> Second, assessment results, especially if coupled with auxiliary information on characteristics of the various regions, would provide data necessary for research on educational problems and processes which cannot now be undertaken.
>
> Third, when sampling and testing procedures are adequately developed, international comparisons might be possible.

Several research groups using private funds have undertaken to develop such procedures. They are exploring promising

approaches; this spring, they will try some samplings experimentally.

These assessments efforts are not designed to test individual students, or individual schools, or individual teachers. They are designed to report on regional or nationwide educational levels over a period of time.

I believe that this matter of national assessment deserves the full attention of the education profession. We are not now reporting satisfactorily to ourselves, to our states, or to the nation on where we stand in education, where we are going, where we plan to go.

The American people today expect more of American education than ever before. At such a time, isn't it clear to all of us as educators that what we don't know *can* hurt us?



7

# National Educational Assessment: Will It Give Us a True Picture?

asks BANESH HOFFMANN

A governmental testing program involving large numbers of students would run the danger of ending as a flawed multiple-choice monstrosity. If it did, the results would tell educators little that current tests do not. Even worse, such a program would set a federal seal of approval on a purblind method of evaluation that exploits ambiguity, rewards superficiality, penalizes depth and originality, frustrates inspired teaching, and corrupts education.

For example, consider this sort of test item:

*Circle the one that does not belong:*
*duck, cat, dog, monkey*

A Vermont child circled *monkey*, reasoning logically that it was the one creature that did not live nearby. His teacher, however, corrupted by the basic fallacy of "objective" tests that there is only one "best" answer, counted the child categorically wrong for not having circled *duck*.

It is just this mind-stunting type of teaching that is fostered by mechanized evaluation. Moreover, even the best of the machine-gradable tests are by no means impeccably constructed. A 1964 sample College Board SAT question [eliminated in 1965] bears this out: In an item testing reading comprehension, the passage said that certain atmospheric layers are "transparent to" the longer radio waves; the wanted answer said that the layers "do not affect" the waves, *which is not at all the same as being transparent to them*. How could eye glasses function if they did not affect the light waves to which they are transparent?

---

Dr. Hoffmann, Department of Mathematics, Queens College, City University of New York, is author of *The Tyranny of Testing*, Crowell-Collier, 19[illegible] (Collier paperback, 1964). He is a theoretical physicist as well as a professor of mathematics. This article is reprinted from the *NEA Journal*, February 1966.

At best, educational evaluation is hazardous. To evaluate our national effort in education solely on the basis of the warped picture presented by mechanized tests would be calamitous. Gifted teachers and educational trailblazers would be rightly dismayed to have their work federally evaluated by such misleading methods, just as artists would be if their colorful paintings were to be judged by persons who were color-blind or worse.

The situation is not hopeless. While keeping in the background, officials of the U.S. Office of Education are encouraging the Carnegie Corporation to develop a new and ostensibly unofficial national evaluative program. In the purely political sense the program is brilliantly conceived. It will not tread on the toes of any individual, simply because no student, teacher, or school will be individually rated. Only a small sampling of students will be tested, and none of these students will be subjected to more than a small sampling of the total evaluative procedure.

Out of this sampling of a sampling could come a vivid picture of the state of American education, because the purpose is not so much to rate students or schools as to gather evaluational data. Furthermore, the sampling, by greatly reducing the pressure of numbers, would afford a superb opportunity to explore quality in depth.

Will the Carnegie Corporation seize the opportunity to develop a meaningful system of evaluation? I am a little skeptical, for Carnegie has already farmed out details of the task to leading test-making organizations that have been long calloused to the ambiguities and other defects of machine-gradable tests.

In attempting to develop an instrument which will truly evaluate the education of American students, will the test makers do some soul-searching? Will they for instance seek to find out what harm the current emphasis on machine-gradable tests may have done to students' powers of written expression? They could do this easily by giving questions, calling for written responses, and comparing them with those of Canadian students, who have not been brought up on a steady diet of mechanized tests.

If the Carnegie Corporation's sampling procedure is developed with boldness and imagination to probe educational excellences and evils by a broad range of methods, it will provide valuable information about American education—including current evaluative procedures.

Certainly national assessment will exercise coercion—let no one deceive himself about that. Indeed, the coercion will have federal overtones. If wisdom and luck prevail, however, it will be a flexible, enlightening coercion amenable to reason rather than a dogmatic coercion of the blindly numerical sort exercised by such instruments as the TV ratings and IQ scores.

If, however, the Carnegie program ends up as little more than a rehash of current objective tests and interest inventories, let us have none of it. It will do more harm than good.



10

RALPH W. TYLER

# A Program of National Assessment

Because a national assessment differs so markedly from kinds of educational evaluation in current use, it may be helpful to relate assessment to the range of evaluation purposes and procedures. One common use of evaluation is to appraise the achievement of individual students. This is usually done with several purposes in mind. It may furnish a further incentive for students to study because they know they will be tested. It may be used as one of the factors in promoting students. It provides information that can be used by the student and counselor in planning for further education, and it often furnishes one of the bases for awarding scholarships.

A second use of evaluation is to diagnose the learning difficulties of an individual student or an entire class to provide information helpful in planning subsequent teaching. A third use of evaluation is to appraise the educational effectiveness of a curriculum or part of a curriculum, of instructional materials and procedures, and of administrative and organizational arrangements.

Each of these kinds of evaluation is an essential part of the processes of teaching or administration. Teachers and administrators are using evaluation of one sort or another as one of their normal procedures. The information gained from these appraisals is focused upon individual student's efforts, class performance, or the effectiveness of the plans, materials, and procedures used by the teacher, the school, or the school system.

## THE NEED FOR ASSESSMENT OF EDUCATIONAL PROGRESS

There is a fourth use of evaluation which is to assess the educa-

Dr. Tyler is Director of the Center for Advanced Study in the Behavioral Sciences, Stanford, California. This paper was presented on February 12, 1966, at the Annual Meeting of the American Association of School Administrators in Atlantic City, New Jersey.



11

tional progress of larger populations in order to provide the public with dependable information to help in the understanding of educational problems and needs and to guide in efforts to develop sound public policy regarding education. This type of assessment is not focused upon individual students, classrooms, schools, or school systems, but furnishes overall information about the educational attainments of large numbers of people.

The distinction may be illuminated somewhat by comparing the situation in education and in the field of health. The public has information about the incidence of heart diseases, cancer, and other diseases for different age and occupational groups, and for different geographic regions. This information is useful in developing public understanding of the progress and problems in the field of health where greatest effort and support may be needed. At the same time, physicians have evaluative procedures to diagnose diseases, to appraise the progress patients are making, and to evaluate the effectiveness of treatments. The physician's evaluative techniques are devised to serve his purposes and the public health assessments are designed to provide the public with helpful information. One type does not take the place of the other.

This is a rough parallel to the difference in education between the tools needed and used by teachers and administrators and those needed to gain information helpful for the guidance of responsible citizens. Heretofore, little attention has been given in education to the assessment problem because the need for wide public understanding of educational progress and problems was not widely recognized. Now it is.

Because education has become the servant of all our purposes, its effectiveness is of general public concern. The educational tasks now faced require many more resources than have thus far been available, and they must be wisely used to produce maximum results. To make these decisions, dependable information about the progress of education is essential; otherwise we scatter our efforts too widely and fail to achieve our goals. Yet we do not now have the necessary comprehensive and dependable data. We have reports on numbers of schools, buildings, teachers, and pupils, and about the monies expended, but we do not have sound and adequate information on educational results. Because dependable data are not available, personal views, distorted reports, and journalistic impressions are the sources of public opinion, and the schools are frequently attacked

and frequently defended without having necessary evidence to support either claim. This situation will be corrected only by a careful, consistent effort to obtain valid data to provide sound evidence about the progress of American education.

## STEPS IN THE DIRECTION OF VALID ASSESSMENT

In recognition of this need, Carnegie Corporation of New York, a private foundation, in 1964 appointed an Exploratory Committee on Assessing the Progress of Education. I was asked to serve as Chairman. The Committee's assignment is to confer with teachers, administrators, school board members, and others concerned with education to get advice on the way in which such a project may be constructively helpful to the schools and avoid possible injuries. The Committee is also charged with the development and tryout of instruments and procedures for assessing the progress of education. The Committee has been working on these assignments for nearly two years.

### Broad Scope of Assessment

The discussions with administrators, curriculum specialists, teachers, and school board members clearly recommended that the initial assessment include more than the 3 R's and that it ultimately cover the range of important educational tasks of the modern school. In harmony with this suggestion, instruments are now being constructed by four leading test development agencies in the fields of reading and the language arts, science, mathematics, social studies, citizenship, fine arts, and vocational education. In subsequent years, other important areas will be included.

### Educational Objectives Identified

Because the purpose of the assessment is to provide helpful information about the progress of education that can be understood and accepted by public-spirited lay citizens, some new procedures are being developed. In each field, scholars, teachers, and curriculum specialists have formulated statements of the objectives which they believe faithfully reflect the contributions of that field and which the schools are seriously seeking to attain. For each of these major objectives, prototype exercises have been constructed which, in the opinion of scholars and teachers, give students an opportunity to demonstrate the behavior implied by the objective. These lists of objectives and

prototype exercises which help to define them have been reviewed by a series of panels of public-spirited citizens living in various parts of the country in cities, towns, and villages. Each panel spent two days reviewing the material and making a judgment about each objective in terms of the questions: "Is this something important for people to learn today? Is it something I would like to have my children learn?" This process resulted in some revisions of the original listing of objectives and some eliminations. However, the procedure was designed to ensure that every objective being assessed is (1) considered important by scholars, (2) accepted as an educational task by the school, and (3) deemed desirable by leading lay citizens. This should help to eliminate the criticism frequently encountered with current tests in which some item is attacked by the scholar as representing shoddy scholarship or criticized by school people as something not in the curriculum or by prominent laymen as being unimportant or technical trivia.

### Populations, Not Individuals, To Be Assessed

A national assessment to identify kinds of progress being made in education, and problems and difficulties arising, will not be very meaningful unless separate measures are obtained for populations within the total country which vary among themselves and thus present different degrees and kinds of progress and different problems to be solved. The particular populations that need to be treated separately may change over the years ahead, but for some time, age, sex, socioeconomic status, geographic location, and rural-urban-suburban differences will probably be significant. Hence, the present plan is to assess a probability sample for each of 192 populations defined by the following subdivisions: boys and girls, four geographic regions, four age groups (nine, thirteen, seventeen, and adult), three divisions by urban, suburban, rural classifications, and two socioeconomic levels.

The fact that *populations* are to be assessed and not individuals makes it possible to extend the sampling of exercises far beyond that of an individual test in which each person takes it all. It may be that a comprehensive assessment would require so many exercises that if it were to be taken by one person he would need ten hours or more to complete them. With a population sample, 20 persons, each spending 30 minutes, would together take all the exercises. In this case, a population of

10,000 persons would furnish a sample of 500 for each of the assessment exercises, and no one would have given more than 30 minutes of his time. Assuming that an assessment would be made every 3 to 5 years in order to ascertain the kinds of progress taking place, it is very unlikely that many of those individuals who participated in the earlier assessments would be involved in any of the subsequent ones. Hence, from the point of view of the child or adult, no serious demand would be made on his time. Furthermore, it is unlikely that the children taking exercises in later years would be drawn from the same classrooms as the earlier ones. Therefore, the demands made upon a teacher in releasing a child for half-an-hour will be minimal. The assessment, though costly, should be feasible and involve little or no inconvenience to individuals or to schools.

Since the assessment does not require that all participants be in classes, the exercises to be used are not limited to the usual test items. Interviews and observational procedures are also to be employed to furnish information about interests, habits, and practices that have been learned. Because school objectives commonly include these areas, it is necessary to see that some assessment is made of the levels of attainment.

### Characteristic Achievements To Be Measured

The assessment exercises will differ from current achievement tests in another important respect. An achievement test seeks to measure individual differences among pupils taking the test. Hence, the items of the test are concentrated on those which differentiate among the children. Exercises which all or nearly all can do, as well as those which only a very few can do, are eliminated because these do not give much discrimination. But, for the purposes of assessing the progress of education, we need to know what all or almost all of the children are learning, and what the most advanced are learning, as well as what is being learned by the middle or "average" children. To get exercises of this sort will be a new venture for test constructors. They are required to develop exercises at each age level in which approximately one-third represent achievements characteristic of most of those at that age level, one-third represent achievements characteristic of about half of those at that age level, and one-third represent the achievements characteristic of the most advanced, that is, the top 10 percent, of that age level.


15

To summarize the educational attainments of these several populations it is not necessary to compute test scores. Instead, the following sorts of things would be reported:

> For the sample of seventeen-year-old boys of higher socio-economic status from rural and small town areas of the Midwest region, it was found that:
>
> 93% could read a typical newspaper paragraph like the following. . . .
> 76% could write an acceptable letter ordering several items from a store like the following. . . .
> 52% took a responsible part in working with other youth in playground and community activities like the following. . . .
> 24% had occupational skills required for initial employment.

### Periodic Reports Planned

It is anticipated that the assessment would be in charge of a commission of highly respected citizens. They and the commission staff would prepare reports of the findings of the assessment, much as we now obtain reports of the findings every 10 years of the decennial census. These reports would be available to all people interested in education, providing them in this way with significant and helpful information on what has been learned by each of the 192 populations. In subsequent years, the progress made by each of these populations since the preceding assessment would also be reported.

## FEARS ALLAYED

This project is encountering some difficulties in getting itself understood. It is being confused with a nationwide, individual testing program, and several common fears are expressed by those who make this confusion. They note that tests used in a school influence the direction and amount of effort of pupils and teachers. In this way, if national tests do not reflect the local educational objectives, pupils and teachers are deflected from their work. This criticism does not apply to the assessment project because no individual student or teacher can make a showing. No student will take more than a small fraction of the exercises. No scores will be obtained on his performance. He

will not be assessed at any later time and can gain no desired end, like admission to college or a scholarship.

A second fear is that such an assessment enables the federal government to control the curriculum. This is also a misunderstanding. The objectives to be assessed are those which are accepted by teachers and curriculum specialists as goals toward which they work. They have been reviewed by lay leaders throughout the country so as to include only aims deemed important by public-spirited citizens. This project will report on the extent to which children, youth, and adults are learning things considered important by both professional school people and the informed public.

A third fear is sometimes raised that this project would stultify the curriculum by not allowing changes over the years in instructional methods and educational goals. It should be made clear that the project will assess what children, youth, and adults have learned, not how they have learned it. Hence, the assessment is not dependent upon any particular instructional methods. For example, we shall report the percentage of 13-year-olds who can comprehend the plain sense of a typical newspaper paragraph. We will not be reporting the methods of reading instruction that are used in various schools. Or, as another illustration, we shall report on the percentage of adults who participate regularly in civic affairs but not on the methods used in teaching high school civics.

The matter of changing educational goals is a relevant question because the objectives determine what will be assessed. Our plan calls for a review one year in advance of each assessment of the objectives of each field in order to identify changes and to include the new objectives in the next assessment.

## A CONSTRUCTIVE CONTRIBUTION

Through the various conferences with school people and interested laymen, the Committee has been able to identify concerns and problems that such an assessment must deal with. As the plans are shaping up, it appears to be possible to conduct the project in a way that will not injure our schools but will provide greatly needed information.

The need for data on progress has been recognized in other spheres of American life. During the depression, the lack of dependable information about the progress of the economy was

a serious handicap in focusing efforts and in assessing them. Out of this need grew an index of production, the Gross National Product, which has been of great value in guiding economic development. Correspondingly, the Consumer Price Index was constructed as a useful measure of the changes in cost of living and inflation. Mortality and morbidity indices are important bases for indicating needed public health measures. Facing the need for massive efforts to extend and improve education, the demand for valid information to support the requests and to guide the allocation of resources must be met. The assessment of the progress of education should make an important and constructive contribution to this purpose.



18

# Panel Discussion on a Program of National Assessment

This discussion followed Ralph Tyler's address at the AASA Annual Meeting on February 12, 1966. The panel* was moderated by Martin Essex, Superintendent of Schools, Akron, Ohio. George B. Brain, President of AASA, presided at the general session and made brief remarks at its conclusion.

BROWNELL: There is no aspect of public service in the United States which is assessed more critically and frequently than American education.

The pupils assess their schooling every day, and the assessment may not be sophisticated, but it is candid, and often revealing and effective in pointing out the needs and deficiencies of the particular school situation; and the parents and the citizens generally and the press and the writers make their assessments daily, so that those of us who are in the operating end of the school systems are quite aware that there is a national assessment of education, and that the ones who are most critical are often the ones that get the most notoriety.

One of the problems is that those are usually assessments of small pieces of the school program. It seems to me that the idea of an operation which reports to the public where we are nationally in the attainment of the educational objectives is a highly desirable operation, but I sense what is of great concern is whether the procedures that are proposed and under way with the commendable objective of national assessment of education *are* going to accomplish the genuine national assessment of education.

---

* Other panel members were S. M. Brownell, Superintendent of Schools, Detroit, Michigan; Leon P. Minear, State Superintendent of Public Instruction, Salem, Oregon; L. B. Newcomer, Superintendent, Clark County Schools, Las Vegas, Nevada; and Harold Spears, Superintendent of Schools, San Francisco, California.

Darius Green and others had the laudable objective of soaring in space. That is what they wanted to do. Their procedures didn't get the results they expected, and were disastrous for them in getting support for what they were trying to do. And many educators are concerned, I think, today, lest the proposed national assessment program result in injuries to education rather than accomplishing the stated purpose. It is not that they are opposed to the objectives. I am glad, however, that some persons persist in their efforts to fly, despite the tragic results of Darius Green's experiments; and I hope the careful and critical examination and present plans for national education assessment will so direct and modify the plans of those who are pressing for a trail that unfortunate results can be minimized and constructive results maximized.

Now, let me turn to some of the questions which I assume and hope are being considered by those who are working on what is called the national assessment program. One of the persistent handicaps in the local school assessment of the status and effectiveness of the school program is the difficulty of separating out and measuring what the school has contributed to educational results.

Let me illustrate. Schools, by law in most states, are required to teach the harmful effects of tobacco and alcohol. You can figure out the amount of time that schools are able to devote to this requirement, and the money that they are able to spend on it, compared with the exposure that pupils get to "education" by other forms, such as advertising, to offset the school education in this one objective. I think many people are wondering how, in this assessment program, it is proposed to assess the effectiveness of schools on an objective which is clearly set forth as an objective of the schools, but which, at one and the same time, may be offset or crippled by the other "educational agencies" that are at work.

I would use another illustration which has to do with good human relationships. I know of no school system that doesn't consider as one of its major objectives, or no group of citizens that doesn't say one of the major objectives of the schools is, good human relationships; and I suspect that most of those who are concerned about the assessment program are wondering how it is going to be possible to evaluate school education effectiveness in this area. What is to be the basis: by the relationship of pupils to each other in school or their human relationships in the com-

munity, by their knowledge of good human relationships or their performance at school age or as adults?

My point is this: I think that many who have been struggling to assess the effectiveness of schools locally in terms of the goals of education that are subscribed to by citizens and scholars wonder whether the national assessment program has or will have criteria which can identify the contributions of schools to what they report as a national assessment of education.

And many educators are so much aware of the limitations of paper-and-pencil tests, as Ralph pointed out, in evaluating the total school output without relating these results to input, that they question the value of placing the national investment and pressure approval behind a procedure that they assume will be merely an extension of present questionable testing practices.

Secondly, I would make one or two other points. Allusions were made in your report to the usefulness of the development of a gross national product in guiding economic development and the consumer price index as a useful measure of the changes in cost of living and inflation.

Now, personally I consider this quite different from the value judgments that are implied in an assessment process. If the objective of the assessment group is to develop an index of gross national educational product, rather than a national assessment of education, it would certainly clarify to me the objective or purpose of the effort in this field. I think that most educators would be glad to have developed a gross national education product index as a tool to emphasize national growth in educational productivity.

If emphasis were placed upon the relationship of expenditure for schools, on personnel-per-thousand pupils, on length of total time spent in school or frequency of change from school to school, on the increase in school holding power, the job success of pupils, and so on, as well as the test scores, then I think some of those who are worried about a national assessment would have their fears reduced.

Then there are many thoughtful persons, among whom are some of the most outspoken critics of the present conditions of schools and colleges, who wonder what are the safeguards that are being planned or erected to reduce the educational ties if the national assessment program misfires in its intended interpretations.

In any social experiment which could affect importantly a



21

major social process for good or for ill, I assume that those in charge must make careful plans to safeguard the social process against harm should their proposals fail, and assure those who are responsible for the operations of social progress as to those procedures that will protect workers on the project or innocent bystanders against a misfiring or a miscalculation of the results.

These are some of the questions I find that are in the minds of persons who are interested in this program of national assessment.

ESSEX: As I listened to Sam and these very pertinent questions, he is asking: one, can we filter out in this assessment process the differences that the school makes and other educational forces in our environment make; secondly, is an analogy fair between the measurable item such as the gross national product or cost of living, as related to value judgments to be made here; and thirdly, he wants to know whether you can retrieve this instrument at a certain point before blasting, in case it seems to be too damaging?

TYLER: I think these points or questions that Dr. Brownell has made are still failing to distinguish between identifying the educational problems we have in our society and the responsibility of different groups for doing something about them. When we find out that there are X percent of the people who have heart disease in the 40-to-50-year age group, we don't automatically say how poor or how good are the doctors in our community. What we do not now have—and I realize there are lots of other kinds of information needed—we do not know whether it is true, as often maintained, that in the rural areas their people don't read as well, or in the rural areas that you have less adequate or more adequate development of citizenship or other things.

The purpose of our assessment, the purpose of using age groups, not grade groups, and not schools, in Detroit, for example, is that in taking all the children in a block, before getting down to the question of what we can do about problems, we need to know what our problems are. The children that would be sampled for the whole Midwest, and not simply for Detroit, will be people in public schools, parochial schools, private schools of other sorts, and perhaps, because we found this in some samples—a larger percentage than we realized—some children who weren't in school at all.

Let me then briefly summarize by saying that the purpose, first, is to get information about what our national problems are, and we are getting them from various sources that are inadequate. There was a book that came out 15 years ago, called *Why Johnny Can't Read.* There was no evidence as to how many Johnnies there were who couldn't read. We need information of a general sort about how well educated our people are and where our problems lie. I realize that at some point in Detroit, for example, there may need to be a further analysis in different school districts as to what their problems may be and what they can do about it; but right now, we have a great deal of money being spent—public money—on the ground that there are certain groups that are not well educated. We don't really have any adequate data on it. We are talking about a prior step, which asks just how are our young people coming on and does not try at this point to make a diagnosis of where and how it might be improved.

Again, let's make the medical analogy. We are trying to find out just where the incidence of disease is. I want to know whether we have diseases and where they are.

MINEAR: Ralph, I believe that many of the people here have heard for the first time the actual project which you and your friends are undertaking, and I believe you have converted many of us to some new ways of thinking. But probably you have aroused some opposition in some other quarters.

I want to add and say to you that the Council of Chief State School Officers, the state superintendents, the commissioners of education of the 50 states, went on record in Honolulu a few months ago substantially approving a form of assessment; and, of course, they went on and said, if and but and so forth. I think you have covered those angles.

I am certain you realize that AASA has a resolution that is before the various committees at the present time, opposing a form of national testing, and you have assured us on that particular point.

In our state of Oregon, I surveyed the superintendents and find that 60 percent of the superintendents who responded to a survey that I made were mildly in favor of some type of assessment, but 40 percent were vehemently opposed.

Obviously, then, it becomes a highly debatable issue, and I have two questions for you. The first is this. On page 147 of

the special analysis of the Budget of the President of the United States for this coming session of Congress appears a statement. and I quote:

> The sum of $2.9 million is included in this budget for developing a new program of collecting educational achievement data on a uniform, nationwide basis for the purpose of assessing the quality of education, or the purchase of data to be collected as a supplement to the current population survey and to initiate a survey of adult education and training for employment.

Basically, what I am saying here is that of the $2.9 million in the President's budget, $2.5 million of this is apparently planned for a new program of collecting educational achievement data on a nationwide basis. My first question is this: Is there a relationship between the assessment program and the President's budget in this particular statement in so far as you are aware?

My second statement and question: I have been told that there are three ways such an assessment program could be mounted. You mentioned the fourth way. A panel of distinguished citizens selected by foundations, and so forth. One additional way would be a Presidential commission of citizens. The direct allocation to the United States Office of Education of this responsibility would be a third way. I can see a fourth way, and that is to leave it where it starts, in the private sector of the educational enterprise.

Now, we are all aware of the fact that the 50 states of our nation have the assignment in the 50 constitutions for the general development of public education. Only under the general welfare clause and possibly portions of the Tenth Amendment of the Constitution do we bring education into the federal sector of government.

So my second question is this: Would you deem it appropriate that the Carnegie program, which is established to get this assessment under way, would you deem it appropriate that this might be, in turn, transferred as one of the first responsibilities of the Compact of States? It is a state right. Once this Compact is inaugurated, there will be demand for a responsibility, or for some job for the Compact of States to undertake. Do you deem this would be an appropriate transfer of responsibility to that agency?

TYLER: The two questions that Superintendent Minear has raised are both questions that are highly hypothetical; namely, the possible connection later that the federal government might have with this. Actually, I don't see any necessary connection. Our work has been entirely in the private sector. We were established by the Carnegie Corporation. Our membership includes people who are interested primarily in this question: how can we develop a process of assessment that will give the public information that will not have any harm on the schools and will not be an assessment of the schools.

This emphasis, for example, on no smaller geographical region than the regions represented by the four in the United States—Northeast, Southeast, West and Far West—is one means of ensuring that we are not talking about comparing one state with another. We are not talking about comparing one kind of community with another. My own belief is that whatever may be the need and the desires of the Congress in trying to assess their responsibilities, our concern is with the assessment of our educational development as a whole, which includes children who may have been educated in parochial schools, private schools, as well as public schools, who may have been at home or out of school altogether.

The question is often raised now: Is it true that such a large percentage of our youngsters are unable to pass the test involved in getting into the Armed Forces? I am sure that if we took a comprehensive sample, we would discover that young people are, as a whole, much better able to read and carry on these operations than the data we are getting from the draft. We are interested in getting balanced, adequate pictures of how well educated our people are and where our problems lie, rather than the questions which are of concern to you as administrators: what can we do in our own school systems to continue and improve the quality and the quantity of education.

ESSEX: Ralph, seated out in front of you are some three or four thousand school superintendents who function in that peculiar organism known as the American community. Now, you state that you propose to do this in four sections, with two elements of the population, various ages, sex, and so on.

What if Madame X of the PTA, League of Women Voters, or a hundred and one other organizations, wants to know of the Board of Education or the Superintendent of Schools how the

youngsters in Ox-Bow are doing on these—will this be a public instrument? Is it available?

TYLER: It will not be a public instrument because the whole basis of the sampling depends on having large enough populations, so that no one city would have any value in using this instrument. It will not be a public instrument. It is copyrighted by the Carnegie Corporation, so as to retain it entirely for use of this sort.

NEWCOMER: Let me put my questions to Dr. Tyler this way. There have been many concerns expressed about the national assessment program, and usually you hear people talk about the national testing program—the ills of national testing. It seems to me that in your comments today you have very well attempted to show that this is not a testing program, and that assessment is different from the testing. However, I seem to detect a gradual change in the communications regarding the assessment program over the past year or several years. We don't see the word "comparison" any more. We did see the word "comparison," I think, when we first read about it, and I suspect—I want to ask the question, whether the controversy that has been going on somewhat along this line has had a tendency to narrow the scope or to sharpen the process. That is one question.

The other one is rather dear to me in that it has been expressed in several ways before. I think that one of the problems we face today is the tendency to equate education with public schooling. When you talk about the assessment of education—if this is going to be used to judge the public school—I don't know how you are going to delineate this process because I think that there are other educational institutions in our society that play an even more important role than the public school. If the national assessment program is successful, there may be—and I ask this in the form of a question: Could there be a tendency for the people to quit the great debate in their local communities, quit trying to determine what the unique task of the school in our communities is and consider that somebody up there is determining what it is, and that it is the same for everybody?

I have a sneaking suspicion that the task of the school in Las Vegas is very different than the task of the school in Newport Mesa. I wonder if this national assessment program might not have a negative effect on local people going into the con-

troversy and asking, as they should, what are the tasks *here* as opposed to other educational institutions.

TYLER: It is always difficult to deal with things that have not happened. What we can do is consider what we know about our own communities and what we have noticed in other fields. I have noticed that, for example, having national data about the problems of executives in getting heart disease serves in our area to make for a special study of the incidence of heart disease in our county and helps to focus the emphasis of the Heart Association upon it.

I do not understand why having knowledge (if this should turn out) that children from low-economic-level groups keep getting wider apart from others between 9 and 17, if this turns out to be true in each of the regions, why this doesn't cause us in Las Vegas or Palo Alto or anywhere else to begin to say: if this is a national problem, let's see what our own problems are and whether we ought to give attention to it. The business of getting more dependable information is not something we have to be afraid of. The role of those of us who are administrators is to make use of dependable information, to help remind our citizens of problems we have, of things we need to do. I don't fear more information; I fear either lack of information or the dependence upon very biased material. There are many organizations, as Sam Brownell has pointed out, that are continually making an assessment, trying to assess education, and I believe that the information about regions, about these divisions that we talk about, 192, may not be the things that will be important 10 years hence. We may find some other factors that we will want to look at in relation to these, either because we discover we don't have some of these problems or we discover we are doing very well with them. By having information of that sort of background, to indicate where our regional or national problems lie, gives us as school administrators a much better basis for saying where we need to put attention in our own communities.

ESSEX: We seem to be saying here that we are going to keep this tool separate from the teacher. We are not going to let the teacher get to this instrument because we fear that this would direct his teaching toward certain things or cause him to eliminate other elements of his teaching that were not deemed to be important here.

TYLER: We didn't really say that.

ESSEX: Is that an assumption we will have to make? If we don't influence the teacher by this instrument, how do we get the teacher with the materials we need to teach?

TYLER: Let me remind you of the medical analogy. The way we get incidence of heart disease does not determine the kind of instrument the physician finds helpful in making his own diagnosis. They are two different things. It is not that we are trying to keep things away from the teacher. A teacher needs a much sharper instrument to diagnose reading readiness, for example, and things of that sort, than the kind of thing we are providing.

ESSEX: Let's look at it from another point of view. Say in a county where one school system spends $300 per child, another spends $600, they are thrown into this polyglot of the aggregate, and the community and the superintendent make judgments. How can these judgments be made in these two distinct school systems in the same aggregate?

TYLER: Again, let me remind you that we are not getting to the level of counties. The people of the country need to know, for example, is it true that one of our critical areas is in providing youngsters of low economic level with marketable skills. We can have information upon what these critical areas are, and that will be useful in general. Then the question of your county is to find out what they are doing in different schools; and the responsibility for pinpointing decisions that need to be made is yours, just as in the Gross National Product, the Ford Corporation or some other particular corporation has to make up its own mind and make up its own study of how well it is producing but against the background of the national concern: are we increasing our productivity or are we not?

The school superintendent, the principal, the teacher, need instruments. They are not these instruments. They need instruments to find out their situation. But I am worried because the public has so many misconceptions about what our educational problems are, because we haven't provided the additional thing that the Public Health Service provides, that is not available just from the physicians' reports.

SPEARS: I noticed that every time, Ralph, that the word "assessment" came in, there was a rumble. Whether you said it or whether Sam said it. Also, I noticed when the word "national" came in, there was a bigger rumble.

I am not going to ask Ralph any questions, because superintendents are not accustomed to ask questions. We are on the receiving end of so many of them, we are out of practice. But you can imagine why we are skeptical, if we are, in this area, and why we need to have people such as you analyze it and give us the facts as to what was intended, because there is so much national management coming in that when anything else comes on a national basis today, we are bound to be a little skeptical until it is proven that it is something that we need and can use.

Even though you say that it is a major change in testing programs, I analyze common features, and what we have here seems to me to be common with what we have been doing with testing since the Army test came out.

This program has standardized tests, as the other programs we have of measurement, and assessment, and it is bound to cover subject areas in which you can make tests; and unless these test-makers are smarter than the same firms that have been making tests in the past, I doubt if they come up with good tests in the other areas outside of the academic and skilled areas.

Again they are testing individuals. We are not going to measure individuals, but we will measure school systems; but the only way you can get to the school system is to get to the individuals and compile scores of numbers of individuals, even though each takes a partial part of this test. Then you combine them and then we get an assessment on different areas of the country.

I am a little sensitive to this on the ground that California has a state testing program, as a lot of people know, and it is a required testing program that we have to go into. It has been in effect only two or three years. What they will do with those scores, we have no idea. All that we are concerned about was that nobody knew about them, and so we couldn't compare them.

When you say this is not a public instrument, we people who work in public education know that everything that is in our office or files today seems to be public. There is no way you can get away from it. We are being badgered all the time for having facts that we are trying to hide from somebody, as though we are trying to keep something away that the public needs, and we are

skeptical here again on the national testing program, because it will compare districts.

Maybe it is moving a little too fast for us, I don't know. Maybe we don't move fast enough as educators or superintendents. It is hard for us. We do have our ways of working. We have been accustomed to local control and local management. When these things come upon us, we are bound to ask for a little time.

On the matter of objectives, we feel that sometimes the objectives of education are set by the instruments themselves, and if the instruments are limited, the objectives of education become what the instruments measured. If we are in that type of movement, we could be concerned and ask your group to put out a little more information about it.

Getting back to the use of the test, I will quote from Dr. Gardner when he was asked to make a further statement on this subject and he went into the matter of the use of it. He said: "Thus it might contribute to a much more active guide than we currently possess to the allocation of public and private funds."

That bothered me a little because teachers who are involved in testing programs for school systems sometimes test for them.

There are two features unique to the proposed national testing program. One is the coverage of the nation rather than a state or local school system or district. The other is the announced use of the result as a guide in the allocation of school funds.

If the latter were carried out, local boards of education would be confronted with mixed emotions. First, the anxiety that their students would not rank high enough. And second, that they would rank too high to qualify the district for federal funds.

We have to get that cleared up a little. Are we working now? If federal funds will be distributed to the lower-ranking districts, we had better tell the teachers to ease off a little with the children when they are taking these tests.

I am not quite sure of that. Take the California situation. Three weeks ago these test scores were released. Our school system got a beating in the local paper because the headlines came out that we were below average, and what was our average on that test compared with these other 143 different districts in the metropolitan areas? We ranked 199.5 as compared to

200 as the average. We were below average! We were 0.5 off the beat on that. I tried to answer and get into the papers, which we did, and talked from the standpoint of achievement of youngsters and whether you measure school systems on that basis, because you have to know whom you are dealing with, to begin with, how far you raised them from, and where you had them to begin with. So the achievement score doesn't mean much unless you go back to the intelligence quotient programs. I indicated that our school system has a number of culturally handicapped students in the school system, and we may be doing better than others that have a higher score. Then I was accused of being disrespectful of minority children, and somebody suggested that I wasn't competent to be in that position.

TYLER: I think Harold Spears has pointed out perhaps one of the most important reasons for a great deal of emotional concern about any sort of assessment. Yet he has illustrated the fact that without this sort of program going on, there is a great deal of effort now being made by people to assess unwisely and inadequately. I do not believe you are going to stop the concern of the public generally for knowing how our schools are doing, because it has reached the point where for every child to get anywhere in the United States or any industrialized nation, he has to have an education; and because education is so important, people care about it, and as long as people care about it, they want some sort of appraisal.

I would rather have an appraisal that has comprehensive validity, that does not pretend to appraise a particular school system, that looks at our total educational problems. People are then able to see where we are coming along, where we don't need to worry, where we need to get more attention. This is what we are concerned with, because we believe that appraisals—and I repeat what both the Superintendent of Detroit and the Superintendent of San Francisco have said since they are under continuing appraisals—we would like to get a more comprehensive and adequate base.

Now, the question that was raised about how this differs from standardized tests, I can only refer you back to the paper to point out some of the technical differences, but the most important one perhaps is the standard job to be done, not a score that is determined on the basis of the mean of the group, so that the problem of individual testing where you are trying to find

individual differences is precisely the problem that has been pointed out by Harold in San Francisco. If you are at the 49th percentile, that is lower than the 50th percentile. If you really talked about the things that kids can do, the public could much more easily see the difference between, and how small the difference is between, 82 percent who can do so and so and 81 percent, than they can the difference between above and below average.

The reason for concerning ourselves again with regions rather than with any state or school system is that the public needs to know where, as a whole, in our areas various problems lie. It is then the responsibility of the constituted school officials with this as a background to look at their own situations, to call for help, or to comment on the progress that is being made. I hope, therefore, that we can look at this assessment in terms of a way of trying to provide information that does not point a finger at any school, any person, or any teacher, but does help the public to get knowledge of where our problems lie, because we do need more and more resources to deal with them.

BRAIN: There was a consensus represented in the discussion today. That consensus I would express as being the imperative of improving the quality of education.

RALPH W. TYLER

# Answers to Inquiries

Following Dr. Ralph W. Tyler's address before the AASA convention in Atlantic City, New Jersey, February 1966, many school administrators continued to pose unanswered questions. Inquiries came to the headquarters office of AASA. On March 2, 1966, a letter was mailed to Dr. Tyler asking him to respond to several questions. The questions and answers follow:

QUESTION: What is the approximate timetable for the preparation of instruments, initial field trials, and full implementation?
DR. TYLER: We hope to have the instruments ready for initial tryouts in the fall of 1966. The purpose of the tryouts is to ascertain the feasibility of each exercise and its difficulty. Assuming that the tryouts are successful, we would hope to have the material ready for the first national assessment in the fall of 1967.

QUESTION: Please clarify the relationship between individuals in the sample population and school systems. What, in general, are the procedures which your agency will use if a school district agrees to cooperate, or if a school district does not wish to cooperate?
DR. TYLER: The problem of getting a comprehensive appraisal of a given age group is the problem of the distribution of these young people among public and private schools and within and without schools. After the area sample has been selected (a particular area, a block, a part of a census tract, etc.), the next step is to enumerate the number of persons in the age groupings that live in that sample area and to indicate whether or not they are in school and, if in school, what school. With this information, it becomes possible to request schools, both public and private, to cooperate if they are interested in doing so in giving those exercises that can easily be handled in the classroom. There are, of course, other exercises that will need to be given individually, especially those involving interview and observation. If a school

district does not wish to cooperate, it will be necessary to arrange for these exercises to be administered in the home situation.

QUESTION: How will the results be tabulated and who would analyze the data?

DR. TYLER: The results will be tabulated by exercises for each population. Since the analysis and reporting would be done by official commission—probably one selected by several of the foundations so as to be free from political bias or governmental responsibility—it is likely that the commission would get a competent staff to write the report, and the analysis in general would deal with particular areas or particular kinds of objectives and comment on the progress being made and the problems being encountered about which the American people would need to give support and help to their school districts in attacking.

QUESTION: To whom will a final report be made and what use do you contemplate will be made of it?

DR. TYLER: The final report of each assessment would be made to the interested lay public, and its use, I think, would be very much like the use made of the report of the American Youth Commission in such books as *How Fare American Youth* which alerted the public to the problems of youth in the depression and helped to get support for improvements in curricula and in job opportunities and other needs of youth.



34

# Recipe for Control by the Few

HAROLD C. HAND

I have designed this paper around a principle which states two conditions which any national achievement testing program whose results are used to compare public school systems must satisfy in order to comply with the requirements of the principle of state and local control of the public school curriculum and the principle of government by consent. I shall apply this principle to the national achievement testing aspect of NAEP (National Assessment of Educational Progress). A leading spokesman for NAEP has told us that one "unit of reporting is the area" and, as I shall explain later, one of the stated purposes of NAEP requires for its fulfillment that one such area unit of reporting be the state, and that NAEP achievement test results must be used to afford comparisons among the state school systems in this country.

The two-part principle presupposes that any national achievement testing program whose results are used to compare public school systems would prove to be coercive in respect to the public schools, because it would inevitably force teachers to teach for the tests. NAEP promoters would say that this two-part principle does not pertain to the national achievement testing aspect of NAEP because probability sampling is to be used in NAEP, and they claim that the use of this statistical device will prevent their achievement tests from becoming coercive. In the words of one of these promoters, it is claimed that under conditions of probability sampling " 'teaching for tests' and revising curricula to conform to test content would not be stimulated."

---

Dr. Hand was formerly on the staff of the College of Education at the University of Illinois and has been a distinguished writer in the field of secondary education.

This article is an adapted version of an address given at the AASA Annual Meeting on February 16, 1966. It is reprinted, by permission, from the March 1966 issue of *The Educational Forum*.

achievement test scores are badly needed, and, given its present leadership, would in time take over at least the national achievement testing aspect of NAEP as soon as it could contrive to do so.

Part of the evidence for believing this to be true is to be found in a section of Mr. Keppel's justification to the Committee on Appropriations of his 1965-66 budget request. Under the project heading of "Assessments of Educational Achievement" here in full is what he said:

> One of our most pressing needs in the field of education is to know how well schools are accomplishing their purpose and how effectively students are learning subjects which are generally taught. This will enable *comparisons between geographic areas* and groups within the educational framework which are not now possible. Educators and scholars generally are giving increasing attention to the problems of assessing achievement in the schools and desirability of evolving measures of *comparative educational effectiveness*.
>
> Preliminary discussions with the Carnegie Corporation of New York point to the possibility of the development of tests which will help accomplish the purposes of this project. The Office is interested in encouraging the Corporation and allied agencies in the development of tests and procedures. We believe that the Office can be very instrumental in the implementation of tests and analyses of various kinds that may be conducted and can serve as a catalyst for bringing together academic and educational authorities for accomplishing the objectives of this project. Thus, our principal contribution to this activity during the fiscal years 1965 and 1966 will be conduct of conferences which will be developed jointly with the Carnegie Corporation and other interested agencies. [Italics added]

Harold Howe, who has been named to succeed Mr. Keppel as Commissioner, left no doubt as to where he stands on the question of a national achievement testing program in his introductory news conference. According to an AP dispatch as reported in the December 23, 1965, issue of the *Tampa Tribune*, Mr. Howe said: "The Congress has a right to get some picture of what's happening in public education as a whole, 'because it is putting money into it.'" Since he certainly knows that achievement testing is a major component of NAEP, his statement is

tantamount to saying that Congress has a right to know what the results of national achievement testing are, "because it is putting money into it."

5. The next logical step would be to make participation in NAEP a requirement for receiving federal aid. It is now legal for the Commissioner of Education to require such participation on the part of all state school systems which receive federal aid under either the 1963 Vocational Education Act or the 1965 Elementary and Secondary Education Act. Both Acts make it obligatory upon the Commissioner to require that evaluations be submitted to him by all state school systems which receive federal aid under either Act, but neither Act stipulates what the required evaluation shall or shall not consist of. Instead, this is left to the judgment of the Commissioner.

6. From here on out, virtually no state and but relatively few communities can support a quality public school program without federal aid. To say that states and public school districts any longer have the freedom to refuse federal aid is a cruel fiction. So virtually all of them would then be forced to go along with the national achievement testing program.

But even if the national achievement testing aspect of NAEP continued to remain in the hands of a private group, with the test results used by the USOE for purposes of making a wiser—hence a differential—use of federal aid, the NAEP achievement tests would become coercive, and control of the public school curriculum would be in the hands of the private group who has the final say-so as to what goes into the NAEP achievement tests.

We shall see why this is true as we take note of the seventh statement.

7. Probability sampling cannot discourage, much less prevent, a great many teachers from teaching for the tests—quite the contrary. It is the *use* made of the test *results* that will be determinative. The NAEP promoters tell us that the NAEP assessment "seeks to describe . . . by different geographic areas." This euphemism means that the public school systems are to be compared state by state. What is true of the sample is true of the whole; this is what probability sampling means. If the performance of the sample public schools in State X is reported to be low in comparison with the sample performances in other



39

states, every teacher in every public school in State X would be under a cloud. This would be true because no school in State X could prove to its patrons that it rates any higher than the general low level of the sample public schools in its state. Since no one knows which schools would be caught in the sample the next time around, and for other reasons noted below, we can be confident that more than enough public pressure would be generated to force the teachers (a) to do everything possible to find out from the sample schools what is in the tests (from the pupils tested, if they have to), and (b) to teach for the tests.

Under probability sampling the work of all teachers in all schools in any state would be judged by the results of the achievement testing done in their state regardless of whether or not any of their pupils had been tested. This means that the images of the public school systems in the 25 below average states would, of course, be clouded.

It is not difficult to see what the consequences of this comparison would be in an era in which nearly all communities must either attract new industry or retrogress into stagnation and decay. Nobody knows better than the business and professional men who make up the power structure of the community that new industry goes where it thinks the good public school systems are, and shuns as it would the plague, those places where the image of the public schools is mediocre, to say nothing of those places where this image is poor.

*We can be confident that the power structures in the 25 or so states with the damaged public school images would turn on the heat.* Although they would know which schools had been caught in the sample in their state, the heat these influential people would turn on would be of about equal intensity all over the state, this because they would know that every public school system in the state would have an equal chance of being caught in the probability sample the next time around. The pressures on teachers to teach for the NAEP achievement tests would be too enormous to resist. Further, the frustrations which local public school boards would experience would be acute. For the vast majority of school boards would not be caught in the probability sample, yet they would be held accountable for the poor test results in a few school systems elsewhere over which they have no control.

No less heat would be turned on the schools if, as one NAEP promoter has recently said would be the case, the NAEP

test results were to be reported by larger regions rather than by states. If this is done, and if the stated purpose of NAEP of affording "a much more accurate guide to the allocation of . . . public funds" is still operative, the USOE would still have to rank the states on the basis of NAEP test results. But instead of there being only one state with Rank No. 1 there would then be two or more states with Rank No. 1, two or more with Rank 2 and so on down the line. About half the rankings would be below the average, of course. So the image of all the public schools in all the states in all the regions with below-average rankings would be damaged. New industry would be frightened away from groups of states rather than frightened away on a state-by-state basis. The heat which the power structures in the threatened regions would turn on the public schools to "get those test scores up or else" would not be one whit less intense. Frustrated public school boards would then be joined in their misery by frustrated state boards of education. For each state board of education would be held accountable for the poor test results in all the states in its region. Under regional reporting, no state board of education could prove that its state deserved any higher rating than the rating for the region as a whole.

All that I have said regarding the coercive consequences of NAEP achievement testing would still hold even if the USOE never has any official knowledge of the results or never makes any use of them. *All that is required is that these results be known to the public.* For once they are known, the ranking by states and/or regions would follow. We can be sure that this ranking would be taken seriously, so great is the prestige of the sponsoring Carnegie Foundation. The upshot would still be that the private parties who determine what goes into the tests would control the public school curriculum.

Nor would the question of coercion depend upon the kind or kinds of questions which might be embodied in the NAEP achievement test items. For it is the resultant image of the public school systems of a state or region, and not the kind or kinds of test items used, which would count with the power structures of the states or regions.

## PROTECTIVE PRINCIPLE PROPOSED

We come now to the two-part principle I wish to propose. But before stating this principle, let us take note of three specifica-



41

tions which it was designed to satisfy: (1) it must say something about each of the major aspects of public schooling to which the principles of state and local control and of consent apply; (2) it must stipulate that the test items used must be restricted to matters *already decided through the exercise of these principles;* (3) it must be restricted to the "what is now going on" aspect of what has already been so decided, for to base test items on intended future modifications in the school program would make all such test items both technically invalid and unfair to teachers.

The two-part principle I wish to propose states that in order to be consistent with the principles of state and local control of the public school curriculum and of government by consent, any national achievement testing program whose results are used to compare public school systems must satisfy *both* of the following two conditions:

1. No test item can be used which presupposes the desirability of any educational objective, or any subject matter content, or any teaching method, or any treatment of pupils which is not now being sought or employed or practiced in any public school system.
2. No test item can be employed which presupposes the desirability of emphasizing any educational objective, or any subject matter content, or any teaching method, or any treatment of pupils to any substantially greater or lesser extent than it is now being emphasized by any public school system.

Now let us apply this two-part principle and its premises to what has thus far been done in connection with NAEP. The promoters of NAEP have ignored both the apparatus of the principle of state and local control and the apparatus of consent. The acid test of government by consent is whether those who make public policy are legally responsible to and legally removable by the public, and the question of whether or not the public schools are to be assessed by a private body free to do whatever it wishes with the results of the assessment is a question of public policy.

The NAEP promoters began by holding a series of seven invitational conferences attended only by people acting in their capacities as private citizens. They held no conferences with sample state boards of education, sitting as such with a quorum present, or with sample local public school boards, sitting as such with a quorum present, to find out what the likely reaction of

such bodies the country over would be to the new public school policy which participation in NAEP would represent. Instead, and despite the fact that these bodies and they alone are legally empowered to establish new public school policies, the NAEP promoters consulted only with persons acting in their capacities as private citizens. True, some state and local school board members may have attended the invitational conferences, but if they did, they could speak there only as private citizens in respect to any new policy for the public schools. In sober fact, then, what the NAEP promoters did was to ignore completely the only bodies legally empowered to establish the new public school policy which these promoters are attempting to bring into being.

Furthermore, the NAEP promoters did not make a systematic factual study of and restrict themselves to ongoing public school practices set up under policies made by legally appointed or elected officials. To operate within the principle of consent the NAEP promoters would either (a) have to restrict their achievement test items to what is 100% common among the vast range of ongoing public school practices in this country or (b) ask each public school caught in their probability sample to delete from the NAEP tests all items not appropriate for that particular public school. Yet the NAEP promoters didn't even draw a probability sample of public school systems to get official reports indicative of what is included in even a representative sample of the universe of ongoing public school practices, to say nothing of making a study to see which if any of the ongoing practices within this sample of the said universe were everywhere manifest.

Instead, the promoters of NAEP proceeded at once to award contracts to private testing agencies for the development of achievement tests to be used in public schools.

Every achievement test item presupposes the desirability of the educational objective to which the item relates. So the private testing agencies had to start by setting down what *they* think the educational objectives of the public schools should be, and to go on from there. The private commission in charge of NAEP is to have the final say-so in respect to the tests. But they are in the same boat and so must pass judgment on the tests in terms of what *they* believe the educational objectives of the public schools should be. A private body, the Carnegie Foundation, is footing the bill.

Private conferees, private test makers, private test approvers, and a private financial angel—there is not a person among

them who is in any way either legally responsible to or legally removable by the public in respect to what the consequences of the NAEP achievement tests would be in the public schools.

Virtually all teachers want their pupils to do well on whatever achievement testing is done with their pupils, so every achievement test that is given by any agency external to the teachers themselves is coercive to some degree in respect to the teachers in question. If this testing is done pursuant to purposes which have been set or approved by the duly constituted authorities provided for by the law, none of the consequences of the resultant coercion can be charged to any violation of the principle of state and local control of the public school curriculum. But when, as in the case of NAEP, the tests used are not restricted to purposes which are known to be thus set or approved, violations of the principle of state and local control would be certain to occur.

I anticipate it will be said by NAEP promoters that I am guilty of making much ado about nothing in my fulminations regarding the violation of the principle of consent. It will be said, I anticipate, that the business of consent will be cared for properly when the state boards of education and local public school boards whose public school systems are caught in the probability sample either accept or reject the invitation of the private commission to permit the use of the NAEP achievement tests in these public school systems. This disclaimer will stand up only if every NAEP achievement test intended to be used in its state accompanies the invitation to any given state board of education, and only if every NAEP achievement test intended to be used in its district accompanies the invitation to any given local public school board, and, further, only if the *stated* expectation is that whatever items or sections of the tests which may be deemed inappropriate by the state or local school officials are to be inked out or otherwise deleted under the supervision of these officials.

I also anticipate that the two-part principle may be objected to on the grounds that it is impossible for the private commission to secure from its sample public school districts the vast amount of information about them which compliance with this principle would require. The task confronting the private commission in this respect would indeed be titanic. For it would require this commission to find out what is now going on in a probability sample of public school districts in regard to the educational



44

objectives, the subject matter content used, the teaching methods employed, and the way pupils are treated—all this to whatever extent it is intended that this, that, or the other aspect of the total program of any sample school system is to be assessed by part or all of the NAEP assessment battery.

There are three answers to this objection. One is that if any such external agency as the private commission in charge of NAEP cannot secure this information, then this agency could not possibly know what would constitute technically valid measures of achievement for the schools in question.

The second answer is that if the private commission cannot secure the information, then it is virtually certain that it could not possibly conduct the appraisal either within the apparatus of state and local control of the public school curriculum or within the apparatus of government by consent—unless, of course, the appropriate NAEP tests are to accompany ever, invitation to participate, and unless the *stated* expectation is that all the NAEP tests are to be modified by state and local officials to make them conform to legally sanctioned state and local conditions.

The third answer is that this information is known, but for the most part only to the school people within any given public school system itself. Some of the most important parts of this information relate to intangibles or quasi-intangibles which are sensed but seldom a matter of written record, e.g., the concepts of self-worth and self-confidence on the part of pupils the teachers are trying to help develop, the human relationships among the youngsters the school is trying to engender, the kinds of viciously miseducative neighborhood environments the staff is trying to compensate for, and the appreciations and aspirations the teachers are hopeful of instilling in the children. But whether a matter of written record or merely sensed, a significant amount of this information can be communicated only partially to any outside group. What is required, then, is that all tests and other instruments by which public school systems are to be assessed must be shown to and their validity judged by the teachers, counselors, supervisors, and administrators of the various school systems if a partial, warped, and unfair judgment of the public schools is to be avoided.

I am aware that something of this sort presumably is or will be going on in the few schools in which the NAEP instruments are now being or will be tried out, but we are unbelievably naive

if we think this will insure more than partial validity if and when these instruments are put to use in other school systems. In his *Phi Delta Kappan* article, Dr. Tyler tells us that as the schools in which these tryouts are being or will be conducted "find the assessment meaningful and helpful, it will then be appropriate to recommend its use on a nationwide scale." To find any assessment "meaningful and helpful" a school must necessarily know what the findings for that particular school are. Yet only a few paragraphs earlier in the same article we are told that the NAEP is being so planned that when it gets under way "no summaries of achievement . . . by school can be made." It would appear, then, that the fact that the tryout schools may find the assessment meaningful and helpful will be no warrant whatsoever that any of the sample schools will so regard the proposed assessment program once the computers have begun to hum. If any claim to the contrary is made it cannot be taken seriously.

The question before us, then, is basically a question of public policy. It is whether the national interest would best be served (a) by embarking on a national achievement testing program in the public schools at the certain cost of relinquishing the principles of state and local control and of consent as these now apply to the public schools, or (b) by holding to these principles at the certain cost of losing whatever the potential values unique to such a national achievement testing program may be.

I venture to suggest that it is the second of these alternatives which state boards of education, local public school boards, and the teaching profession in the public schools must embrace unless and until the American people repeal the pertinent laws which are now in the statute books.



46

# The Assessment Debate at the White House Conference

A lively clash of ideas among leading educators arose at the White House Conference on Education last July 20 [1965] when Panel 2-B ("Assessment of Educational Performance") chose to discuss the question of national assessment now being studied by the Carnegie Corporation.

Panel consultant John Goodlad opened the discussion with a proposal that a nationally televised colloquy accompany the beginnings of assessment to insure a general understanding of the twin problems of clarifying educational goals and of assessing their attainment.

Should there be periodic assessments of American education? A discussion summary follows.

CARR: Yes, and this is being done every day by those who should continue to do it, the teachers in the classroom. We should not substitute a national assessment for this. Pluralism in testing should be retained. Assessment should be used to help teachers do a better job.

DUNNAN: Ingenuity, flexibility, and initiative can be lost in a national testing program. [Note the shift from the question of assessment to "a national testing program."] We have two monsters now, College Entrance Examinations and Merit Scholarship Exams, keeping youngsters from developing intellectual creativity.

---

Reprinted from the September 1965 issue of the *Phi Delta Kappan*.

Participants in the panel discussion were: *Chairman:* Harold B. Gores, president, Educational Facilities Laboratories, Inc., New York; *Consultant:* John I. Goodlad, director, University Elementary School, University of California at Los Angeles; *Questioner:* Stephen K. Bailey, dean, Maxwell Graduate School of Citizenship and Public Affairs, Syracuse University; *Panelists:* Hedley Donovan, editor-in-chief, Time, Incorporated, New York; Jack Arbolino, director, Advanced Placement Program, College Entrance Examination Board, New York; Thomas W. Braden, president, California State Board of Education; William Carr, executive secretary, National Education Association; and Donald W. Dunnan, superintendent of schools, St. Paul, Minnesota.



47

BRADEN: Overcoming prejudice, establishing a value system, and developing factual habits of mind are the real goals we must learn how to assess.

DONOVAN: Assessment is going on all the time, not only by teachers but by parents, voters, employers, legislators, and others. I question our ability to construct any valid system of central assessment. The divisive effect of any effort to quantify our strengths and weaknesses may not be worth the gain.

BAILEY: And if our goals are based simply on non-prejudice, a set value system, and pure rationality, how do we create the warmth, the uniqueness, and the unpredictability so important in our society? Our discourses must produce warmth as well as light.

DUNNAN: We must be most concerned with understanding the *minds* of our students. Man has lived on this earth three-quarters of a million years and has only begun to learn the rudiments of the way the human mind works. The future is limitless.

GOODLAD: And we have only begun to utilize the potentials of assessment. Today we are happy if a fifth grader with an IQ score of 100 reads at the sixth-grade level, unconcerned with where he should be reading, based on *his* ability, not the group average of attainments. All fifth graders can be reading at the fifth-grade level and *all* be deficient in some aspect of reading going undiagnosed. We have done almost no assessment based on criterion measures rather than norm-based standards. The difference between assessing individual *students* and assessing our total educational effort with its resulting educational performance is enormous.

We have begun to be able to evaluate such competencies as inquiry and decision making, but we use these assessments hardly at all in determining the adequacy of our educational effort. Especially in teacher education we still assess chiefly knowledge of facts about schools, about child psychology, etc. We must assess various kinds of talents suitable for different kinds of teaching tasks and report the kinds of teaching abilities being developed. We have three distinct jobs: identifying desirable pupil traits; learning to assess them; and deciding which ones to value.

ARBOLINO: Can we assess student achievement without encouraging conformity?

LINDLEY STILES (dean, School of Education, University of Wisconsin): What are we afraid of? We compare people on physical fitness, color blindness, mental retardation, and other traits. Why not compare them on academic achievement? The NEA assesses states on their treatment of teachers and applies sanctions to those which fail to measure up.

JOHN H. FISCHER (president, Teachers College, Columbia): What objection is there to an assessment that seeks to collect data on the condition and progress of education in the several states? The only objections we have heard are hints that if we get more significant information someone might be tempted to misuse it. Isn't that somewhat irrelevant?

JERROLD ZACHARIAS (Massachusetts Institute of Technology): Fact-finding we need; new tests we need; testing the tests we need. But we do *not* need to hang a sword of Damocles over the head of the child by instituting ruinous pressure of competition.

HAROLD TAYLOR (former president, Sarah Lawrence College): We need not more tests but fewer interruptions of learning. At Sarah Lawrence or anywhere else, when you abolish tests you get more learning. Europe, the Middle East, and the Far East illustrate the paralysis of education produced by standard tests linked with government assessment. Let's abolish all tests and concentrate on teaching and learning.

SISTER JACQUELINE (president, Webster College): Any unit —town, state, or nation—that employs a monolithic set of standards is stifling growth. It doesn't have to be a national test.

GOODLAD: The comments about needing more knowledge and needing to keep the pressure off the students emphasize the difference between fact-finding and testing. A national assessment can provide information without setting standards or creating competitive pressures. We do need to know what students know and how they think. And we can discover this without creating pressures.

For one thing, we can select various samples of children from different localities, administer to each section of the sample 10 percent of the total test. Assembling the results, we can achieve the desired knowledge of student attainment in many aspects of life without ever pressuring or punishing or grading children, but simply finding facts.

GORES: To reassure those who fear pressures, conformity, and teaching to the exam, how do you make a national assessment 'antiseptic?

GOODLAD: The national discourse, the televised colloquy we mentioned at the outset, will help to dissipate fears. We can insist that all researchers be more responsible for careful evaluation of their own research efforts. We might have interstate commissions or other cooperative nongovernmental bodies to advise us. And we can schedule top-level colloquies on the purposes of assessment to educate us in the differences between punitive testing of the individual and information-getting assessment of our educational effort. We must raise the level of educational discourse.

RESOLUTION OF COUNCIL OF CHIEF STATE SCHOOL OFFICERS NOVEMBER 1965

# Measurement and Reporting of Educational Results

We recognize that expansion in quality, scope and cost of education brings new responsibilities to the educational agencies of local, state and federal governments to evaluate public education and to report the results of these evaluations appropriately to pupils, parents, teachers and the general public. The people have a right to know about the results of education and its cost as the largest peacetime function of government. The action of the Congress in making evaluations of educational programs at the state and local levels part of the requirements for the use of supporting federal funds reflects the general public desire for educational accounting. Despite the difficulties involved, we believe public school administrators have an obligation to make such an accounting. In the process, it should be possible to assist both the profession and the public to understand more thoroughly that the complexities of teaching and learning limit the extent to which the results of education can be measured with precision.

We believe local and state educational agencies can and should prove themselves capable of assuming increased responsibility in this field. In our times, most important areas of human endeavor are organized and conducted with reference to local, state, national and even international environments. Agriculture, transportation, communication, commerce, finance, labor, science, research and social trends of all kinds operate in increasingly broader environments, and so does education. We believe that both the basic objectives of education and the more limited objectives and methods affecting teaching and pupil learning in schools should reflect these realities. The significance of the widening environment within which education operates is clear. Both the scope of the curriculum and the evaluation of learning achievements in schools should take it into account.



51

We recognize the problems in gaining professional and public support for necessary delimitations in the field of evaluating and testing. Judgments concerning the personal attitudes and values of pupils should be a primary responsibility of their teachers and others who know and understand the pupils best. Personality tests for these purposes are of doubtful value for widespread use in their present stage of development. Measurements of academic achievements of pupils also are primarily a responsibility of teachers, and here standardized achievement tests can be useful to supplement classroom evaluations. One beneficial effect of these tests may be to stimulate learning to make generalizations broader than those that may be learned in many local courses of instruction.

Instruments for assessing progress in education over wide areas should be developed in ways to insure maximum objectivity, reliability and validity of their results. Since the best of such instruments fall far short of perfection and since the general public and many individuals tend to place excessive reliance on their mathematically expressed results, we call for careful restriction on the use of such instruments to purposes which they clearly serve. These hazards lead us to favor restriction of the use of personality tests to limited experimentation and clinical work by experts. On the basis of changing views about IQ and other tests of innate intelligence in recent years, bringing general recognition that environment plays a greater part in the development of measurable intelligence than had been previously supposed, we disapprove of excessive reliance upon ability test scores in the placement of pupils and other abuses of these scores.

We believe that testing instruments now available for measuring academic achievement are among the most objective, reliable and valid of all tests, and that they may be made generally beneficial to pupils, teachers and educational institutions. The hazards connected with their development and use must be recognized and can be controlled. We call on all who develop such testing instruments and use their results to cooperate with local, state and federal authorities in eliminating interference with local classroom instruction. in discouraging factual drill primarily to raise test scores, and in avoiding misuses of test scores. Parents, school staffs, administrators, school board members, leaders in the mass media and intelligent private citizens should work together to achieve general understanding



52

of desirable uses of appropriate instruments for the measurement of academic achievement.

We recognize the need for improvement in the assessment of American education. Such assessment should be developed, administered and interpreted, however, so as to take into full account the wide diversity in the social, economic and ethnic character of our nation and in the aspirations of our people. Agreements among states providing for interstate assessments of educational results can supplement local and state evaluations and further local, state, and national cooperation in education in terms of local, state and national needs. It is essential, therefore, that any assessment should be prepared in cooperation with and administered by the states and localities.



53

ASCD EXECUTIVE COMMITTEE STATEMENT JANUARY 1966

# Guidelines for National Assessment of Educational Outcomes

The Executive Committee of the Association for Supervision and Curriculum Development has made an effort to keep the membership informed of developments in the area of national assessment of educational outcomes. In addition to reporting developments and indicating problems and issues, the Committee is likewise concerned with the dimensions that should be evident in any project that is proposed or implemented.

Accurate assessment of educational outcomes is essential for sound planning and effective stimulation of growth in our educational structure. Assessment has always been an integral aspect of curriculum development and is a major responsibility of curriculum workers. This responsibility is especially critical in a time of awakened public concern, massive Federal commitment and widespread professional reappraisal of our educational endeavors. It is, therefore, necessary that curriculum workers everywhere develop new procedures for assessment far beyond present levels to meet properly the changing needs of our times.

Whatever goals or procedures for assessment are applied to education have inevitable effects upon the nature and function of the curriculum. This inescapable relationship must be clearly understood by curriculum and assessment workers alike for the consequences of blindness are intolerable. We cannot afford to destroy with one hand what is built at great cost with the other. Any assessment program must therefore meet at least the following criteria:

1. *Assessment must value and maintain the diversity of our people.* The United States is a nation of many cultures and its proudest traditions rest upon the dignity and worth of the individual and the protection of the rights of minorities. To this end assessment must value and encourage a diversity of schools

Reprinted from the January 1966 issue of the *ASCD News Exchange.*



54

properly expressive of the ethnic, religious and geographic backgrounds of our people and fulfilling the multiplicity of objectives demanded by local communities of their schools. Sensitivity of schools to local needs is an essential characteristic of a democratic society.

2. *Assessment must protect and encourage uniqueness in students and citizens.* Even as plurality of cultures characterizes our society so opportunities for individual development and growth must be prized. This includes an awareness of the cultural heritage of the individual and a chance to go beyond that heritage in terms of the individual's own growth and potential. A dynamic society cannot afford to stifle creativity and uniqueness in its young.

3. *Accurate assessment of educational outcomes requires exploration in breadth across the full range of educational objectives.* This includes social, emotional, vocational, health and artistic goals as well as basic skills and intellectual growth. Where valid procedures for assessment of these broad objectives do not exist, they must be developed. Data obtained from a limited sample of these objectives or from limited instruments can provide but a distorted picture of educational accomplishment.

4. *Adequate assessment also requires exploration of learning in depth.* Learning may vary from superficial "knowing" to effective, efficient "behaving." It is not enough that schools produce students who "know" better. The only valid criterion for effective learning is whether the student behaves differently as a consequence of having participated in the process. Proper assessment must be directed to the deeper questions of effective behavior.

5. *Assessment must explore the changes in students over periods of time.* The static assessment of a sample group of learners at a particular point in time fails to distinguish the effects of education from the effects of parental or other environmental influences. Adequate assessment programs must include measures of growth and change to assess properly outcomes.

6. *The effect of the assessment procedures themselves on the nature and function of the educational process must be ap-*

*preciated and continuously subject to scrutiny.* As large-scale assessment of the effectiveness of education begins to take place there are dangers that curricula will be tailored to fit evaluation instruments and that sterile uniformity will replace healthy diversity. The proper outcome of large-scale assessment is the use of data as the basis for further dialog and clarification of the objectives and means of education. It must further rather than inhibit the above stated objectives.

# National Testing and National Curriculum

AASA RESOLUTION FEBRUARY 1966

The American Association of School Administrators opposes any act which would, in effect, establish a national testing system or a national curriculum. Not only does such a high degree of centralization infringe upon the legal responsibilities of the state and the school system and the professional responsibilities of the individual teacher, but we believe it inevitably defeats the declared aim of American education—the individual development of each child. Attempts to evaluate or compare all systems through prescribed national tests will result in a curricular structure which will vitiate attempts of local schools to serve individual pupils.

The AASA believes that the voluntary cooperation of local school districts, education associations, foundations, state departments of education, and regional agencies is capable of producing sufficient data for research, evaluation, or assessment purposes and as much standardization of curriculum as is consistent with the maximum quality desired.