DOCUMENT RESUME

ED 081 787 TM 003 148

TITLE	Proceedings of the State Advisory Committee on the Assessment of Educational Quality (1st, Harrisburg, Pennsylvania, February, 1968).
INSTITUTION	Pennsylvania State Dept. of Public Instruction, Harrisburg.
PUB DATE	Feb 68
NOTE	49p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Conference Reports; *Educational Quality; Evaluation Criteria; *Program Evaluation; Program Planning; *State Programs; *Student Testing; *Testing Programs
IDENTIFIERS	*Pennsylvania

ABSTRACT

Proceedings of the first meeting of the State Advisory Committee for the Pennsylvania Bureau of Quality Education Assessment are provided. The proceedings are comprised of the following: Superintendent's Message; Charge to the Committee; The Plan for Assessing Educational Quality in Pennsylvania; What Will Quality Education Assessment Measure?; Progress Report on the National Assessment Project; Summary of Comments and Suggestions; and Conclusion. Concerns of the committee were related to the sampling procedure for the April 1968 testing, the drawback of the plan in the reliance on paper and pencil self-report instruments, the items language in terms of difficulty level and of semantics, the working of certain items which seemed to imply a value judgment of the behavior in question, and the need to establish a list of "guidelines" for teachers concerning the nature of the assessment and its purposes. (DB)

PROCEEDINGS

First Meeting of the State Advisory Committee on the Assessment of Educational Quality

February 1968
Harrisburg

PENNSYLVANIA DEPARTMENT OF PUBLIC INSTRUCTION

MEMBERS OF THE STATE BOARD OF EDUCATION COMMITTEE ON QUALITY EDUCATION

Mrs. Albert M. Greenfield, Chairman
Mr. Mario C. Celli
Mr. Parke H. Lutz
Dr. Katharine E. McBride
Mr. Gail L. Rose
James H. Rowland, Esq.
Mr. Duane Wilder
Dr. Leonard N. Wolf

MEMBERS OF THE STATE ADVISORY COMMITTEE ON QUALITY EDUCATION ASSESSMENT

Dr. Neal V. Musmanno, Chairman
Dr. Paul B. Campbell

Dr. Douglas J. Bowman
Mr. H. Clair Henry
Dr. Kenneth D. Frantz
Dr. John F. Yon
Dr. John C. Dunlap
Dr. Thomas R. Heslep
Mr. John F. McHugh
Mr. Marcus Foster
Dr. Joseph Brennan
Dr. Harold Mitzel
Dr. Donald C. McGarey
Miss Pauline Keiper
Dr. Charles F. Wilson
Mr. Ward L. Myers
Dr. Carl A. Newman
Mrs. Howard J. Huegel
Mrs. Norman B. Shain
Mrs. John Yurchenco
Miss Rosana Eckman
Mrs. Henry Gardner, Jr.
Dr. Harry K. Gerlach
Mr. John W. Scott
Mr. Richard Dario
Mr. John Hertz
Dr. Ralph Swan
Mr. George Walton
Mr. Harry Boyer
Hon. Donald W. Fox
Hon. Preston B. Davis

CONTENTS

SUPERINTENDENT'S MESSAGE

The deliberations of the State Advisory Committee on the assessment of educational quality represent the first steps toward systematically determining the quality of the educational programs in the public schools of the Commonwealth.

It is my feeling that the Quality Education Assessment Program is one of the most outstanding programs in the Department of Public Instruction. Difficult as the task may be the results, I am certain, will be most rewarding. As we move toward an educational system of excellence through the improvement of instruction, all citizens of the Commonwealth will benefit.

A program on paper is one kind of plan. The implementation of the program is that part of Quality Education Assessment which we are emphasizing in these Proceedings. If we are to have objective evidence about the adequacy and efficiency of educational programs, with emphasis on the individual school and its capacity for meeting the specific needs of the specific children it is to serve, we must take the necessary first step toward reaching this ultimate goal. Together I am sure we will reach this goal.

We appreciate the comments and suggestions the Committee has given us, and we are grateful for its continuing advice and support in our joint effort.

David H. Kurtzman

CHARGE TO COMMITTEE

Fellow educators and friends of education, since I assumed the responsibilities of the Office of Commissioner for Basic Education I've been doing my best to discourage unnecessary meetings. We try, whenever possible, to rely on memoranda and telephones, rather than waste precious hours of busy people by scheduling another meeting, except for a very good reason.

I tell you this to underline the significance with which we, in the Department, view this meeting of the State Advisory Committee for the Bureau of Quality Education Assessment which is holding its initial meeting here today. This is a Blue Ribbon Committee, not only because it represents a cross-section of top leadership in Pennsylvania Education, but because its deliberations and suggestions relate to a concern among educators that portends national significance in the matter of evaluating the quality of educational product in this Commonwealth. Pennsylvania can truly become a pioneer in the relatively untraveled frontier of assessment of quality education.

You may remember that when the Department Reorganization was announced, the duties of the newly created post of Commissioner for Basic Education were clearly enumerated in the DPI Bulletin.

Of the many broad responsibilities assigned to this office, the transcendent function, it seemed, was that "It shall be his goal to move Basic Education to the highest quality possible in Pennsylvania." This same objective happens to be one of the major concerns of the present Administration of Pennsylvania.

One section of the far reaching School District Reorganization Act of 1963, Act 299, sets forth the General Assembly's desire for Educational Performance Standards in Pennsylvania, as follows:

> "The State Board of Education . . . . shall develop or cause to be developed an evaluation procedure designed to measure objectively the adequacy and efficiency of the educational programs offered by the public schools of the Commonwealth. The evaluation procedure to be developed shall include tests measuring the achievements and performance of students pursuing all of the various subjects and courses comprising the curricula . . . ."

To carry out this mandate, the State Board appointed from its members a Committee on Quality Education, headed by Mrs. Albert M. Greenfield, who was also the chairman of the Board's Council of Basic Education. Upon this Committee's recommendation, the Board entered into a contract with the Educational Testing Service of Princeton, New Jersey, to assist in the development of a plan to implement this provision of Act 299.

A series of studies during 1965 for this purpose resulted in a report by the Educational Testing Service to the Board. I will not take the time to summarize this report. You will find it in the Highlight Booklet you received at registration.

In the recently completed reorganization of the Department of Public Instruction, a Bureau of Quality Education Assessment has been created to spearhead this program, cooperatively with the Bureaus for Research, Curriculum, Guidance, Instructional Services and other related activities, and in concert with administrators throughout the Commonwealth.

Even without a legislative mandate, the case for educational assessment seems to present a pursuasive argument these days. When we are spending 52 billion dollars of federal, state, and local tax money for education this year, more than the gross national product in 1935, it seems reasonable to expect that educators and taxpayers alike should want some demonstrable proof that the product we turn out is improving in quality.

The functions spelled out for our Bureau of Quality Education Assessment are:

> In cooperation with school districts and intermediate units, assess quality education in the public schools.
>
> Coordinate with other Bureaus in developing through research, experimentation and evaluation instruments for measuring quality education.
>
> Develop procedures for applying measures of quality education in cooperation with local school districts.
>
> Analyze and report to the State Board of Education on quality education development.

I have read these charges to the Bureau purposely to emphasize the repetitive use of the words "cooperation" and "coordination."

These are the keys to the success of this new program of Quality Assessment. As a practitioner with 32 years in public education, I, like you, would view with alarm any state test arbitrarily imposed on the schools.

It is not the purpose of the Department to assume an inspectorial or regulatory function. Rather, we envision this service as a cooperative venture to be shared and developed by all organizational levels of educational concern.

We will work on the basis of programs and facilities now available in our schools. We will measure skills and achievements of children and teachers. We also hope to develop instruments to determine attitudes of administrators, teachers and pupils.

Accomplishment of this goal will require the best thinking and the combined judgment of all educational leaders in our schools, as well as of the Department of Public Instruction professionals.

To this end I beseech your understanding and your help.

This, then, is our purpose and our hope. The charge to this Committee is to ask for your help by giving us your thoughtful understanding, your professional judgment and your sympathetic concern.

B. Anton Hess

THE PLAN FOR ASSESSING
EDUCATIONAL QUALITY IN PENNSYLVANIA

Pennsylvania's Plan for educational assessment is of necessity and with full intent based upon the planning and thinking of professionals who have been previously concerned with problems of a like nature. Isaac Newton's comment about standing on the shoulders of others applied wholesale to our efforts. Our Plan utilizes, where possible, the work undertaken for the State Board of Education by the Educational Testing Service (1). It has been modified from the ETS Plan in order to fit more closely with the working realities of the Department of Public Instruction and of the public schools in the Commonwealth. It attempts to take into account the experience of Ralph Tyler and his associates with their plans for the national assessment of education, particularly those facets which caused so much concern among school administrators (2). The experience gained by our colleagues who have been engaged in the New York Quality Measurement Project has also been considered. The Plan reflects the position represented by Alexander Mood concerning the necessity of using, in many instances, approximate, empirical and only logically valid devices. These are items or exercises which look and sound as if they would measure the kind of achievement we want, although they have not been subjected to rigorous trial (3). We agree with Dr. Mood that we cannot wait until a final polished instrument is available for every area which we wish to measure.

We are willing to begin with imperfect instruments because we think that by their very use, refinements in the instruments themselves will be possible. We are attempting to measure a broad range of achievements to avoid directing educational outcome toward some limited, test defined objective.

The Pennsylvania Plan as presently conceived consists of three major tasks. The first of these is the process of selecting or developing measurement devices for student output. Output is defined as those behaviors which are within a student's range of accomplishment at a given point in his school career. Can Johnny accurately multiply 429 x 325? Does he know where to find out about the requirements of a certain college or career? Does he have enough confidence in himself to try a difficult but possible task? Measurement devices must be selected or devised for each of the ten goals which were accepted by the Quality Education Committee of the State Board of Education in 1965. These goals stated in brief form are:

1. Self understanding and self acceptance.
2. Understanding and appreciation of social, cultural, and ethnic groups different from his own.
3. Mastery of the use of words and numbers.
4. Positive attitudes toward school and learning.
5. The habits and values associated with responsible citizenship.
6. Good health habits and knowledge of the conditions necessary for maintaining physical and emotional well-being.
7. Experience with creative processes in a variety of fields.
8. A full grasp of the opportunities for preparing for a productive life.
9. Understanding and appreciation of the natural sciences, the social sciences, the humanities, and the arts.
10. Preparation for effective participation in a changing world.

Depending upon the state of the art, measurement will consist of the degree to which a student is able to perform behaviors which are consistent with, necessary to, or sufficient for, the accomplishment of these goals. Continued discussion of the goals themselves among lay leaders and professional educators is also required.

The second part of the Plan involves classifying student output according to modifying conditions. A review of the literature suggests that factors such as financial resources, community effort, average socioeconomic level, median teacher salary schedule, teacher/pupil ratio, and factors of student potential such as the ubiquitous IQ, individual socioeconomic level, as well as mother's educational level, all are correlates of output in significant ways. In other words, the higher the median teacher salary schedule, the higher the student's socioeconomic level, the higher his mother's educational level, the more likely he is to score well on an output measure.

We intend to report our data in the form of multiple classification expectancy tables. What we mean by this can be illustrated as follows:

Resources You know well the problems of finding adequate financial resources to operate the kind of schools you want for your communities. If we divide the school districts in the state into three categories, on the basis of financial resources, we would expect to have at least three different distributions of student output for each of these conditions. It's asking a bit much to expect the same output from a school with a $10,000. tax base behind each pupil as we might expect from a school with $100,000.

Effort Another factor which may well influence output and may be independent, or at least partially independent, of resources is the amount of money the community is willing to put into education, or, an index of community effort. One wealthy district may put a much smaller proportion of its total resources into education than another, equally wealthy. On the other hand, a very poor district may be putting a very large proportion of its resources into education, and, therefore, be able to compare favorably financially with another district more affluent than itself. If such a variable as this is considered, we have additional groupings possible. For example, we might have a category of districts with high resources and low effort, or medium effort and low resources. Neither of these two variables is directly controllable by the school district itself, although, of course, the school may influence to some degree what the community is willing to do.

School Characteristics The third factor which may influence output has to do with the characteristics of the school itself. These include such things as the proportion of the school's finances it puts into instruction as compared to overhead items such as maintenance, administration, transportation, etc. Schools cannot completely control these factors, but can probably have more influence in this area than in the amount of resources a community is willing to devote to education. For example, student/instructor ratios would show up in this category as well as staffing patterns, educational level of staff members, etc.

If we divide the range of school characteristics into three parts, we should now have many more possible sub-populations for which different sets of expected output, or norms, might occur.

Student Potential The fourth kind of variable which needs to be taken into account is that which the student himself brings as an individual, his own potential. At this point we propose to develop for each of the ten goals, an expected output for each student, thus giving us a distribution of expectancy in each of the several categorizations we have defined. In other words, for a high effort, medium resources, medium school characteristic school, we might expect twenty percent (20%) of the students to score high for a given goal, thirty percent (30%) moderately high, thirty percent (30%) about the middle, and so on. (These are hypothetical figures.) With this kind of an arrangement we are no longer placed in the position of comparing a wealthy suburban district with new buildings and favorable teacher/student ratios to a school district in a community with old buildings, declining population, reluctant taxpayers, and severe recruiting problems. In order to assure that the expectancy table categories accurately reflect the correlating conditions and in order to determine the relevance of previously identified correlates to the school districts of Pennsylvania, we are administering a measurement package to collect correlational data in April of 1968. It is this package which you will be reviewing tomorrow.

Following the refinement of our descriptive categories through this first study, data on the status of accomplishment for each of the ten goals will be collected to provide norms for varying kinds of learning situations throughout the state. Establishing norms in this manner will help us to avoid comparing unlike student populations. It will also provide a comprehensive picture across the Commonwealth of student response to the items judged consistent with the goals, and will show what conditions are associated with varying responses.

The Bureau of Quality Education Assessment will provide several kinds of services to local schools as the third part of the plan. It will provide a measurement package for assessing students in each goal area. With the normative data, it will assist teachers and administrators in looking at their students' output in terms of what might be expected from students in districts operating with similar controlling conditions. Also, the Bureau staff will be prepared to assist the school staff in considering possible reasons to account for differences which may occur between the output of the students and their expectations, regardless of the direction of the differences. The Bureau will provide a coordinating function to assist school districts in modifying the conditions or processes which lead to unsatisfactory outputs and in determining those which contribute to outputs greater than expectation. Furthermore, the Bureau will assist local school districts in testing the effects of program alterations. It will assist in the development of measurement devices and in subsequent comparisons with appropriate expectations or norms.

Procedures for Implementing the Assessment Plan The process of implementing the Assessment Plan can be illustrated by tracing one of the ten goals from the beginning of measurement development, through administration to utilization of results by teachers and administrators. For this illustration I would like to share with you our struggles with Goal VII, Creativity. After abandoning, as inadequate, two or three promising models, we have secured a tentative beachhead on this nebulous island. We think that there are ways to differentiate the degree of creativity in the expression of pumpkinness in children's drawings of Jack o' Lanterns, and that there is something to be learned from taking a picture through a goldfish bowl. In fact, on our latest model, the Metropolitan Museum's classic Greek Horse, which was recently updated to 1917, survives as an example of somebody's creative output.

Talking about creativity, however, always seems to generate a lot of heat for some reason. Perhaps a slightly less controversial subject, Goal II, might be a less distracting illustration.

Now Goal II states that quality education should help every child acquire understanding and appreciation of persons belonging to social, cultural, and ethnic groups different from his own. Let us see how Goal II fared in the first phase of our plan. How does a child behave if he is fulfilling the requirement of this Goal?

In cooperation with the Bureau of Research, we began to work on this problem by considering existing measurement devices related to attitudes toward people and by expressing behavioral objectives through the construction of single items which were specific to some kind of behavior a student might exhibit. We considered and utilized items such as "How would you feel about having as a best friend a person whose family is much poorer than yours?" We considered and rejected items such as "I would ride with anybody who attends this school." The limitations of this single response type of question, however, led us to consider another type which would not only allow a student to express his preference for a kind of action, but also would permit him to state some reason for such an action. We, therefore, began to recall from our own school experience, and to imagine, social situations with possible kinds of behaviors responsive to these situations, and reasons for such responses. After a few attempts of this sort, we discovered an acceptance/rejection rationale emerging inductively from the behaviors specified through item writing. For example, if a Negro child enters an all Anglo-Saxon classroom, we postulate several possible behaviors. Some children might immediately try to get to know him better, others might wait to see what he is like before being friendly, and still others may stay away from him.

It became apparent that the reasons why a child might immediately seek acquaintance could differ considerably. One reason could be that the child is genuinely open and friendly. Another child could feel obligated because he has been told that he should be nice to everyone. Still another might desperately be seeking interpersonal relationships.

The rationale which developed, accounts for nine classifications of motives, ranging from the quality of "open acceptance" as being most desirable to "need gratifying avoidance" as being least desirable. Falling in the middle range are statements such as "there are good and bad people in every group" and "one should be nice to everyone", exemplifying "stereotyped caution" and "stereotyped acceptance" respectively.

After testing this rationale by the application of different sets of behaviors and through presentation and discussion with small available samples of students, the model evolved to the form in which it is presented in this paper.

	OPEN	STEREOTYPED	NEED GRATIFYING
Acceptance	I like people	One should be nice to everyone	I need friends
Caution	I like people only if they are not <u>too</u> different	There are good and bad people in every group	I'm afraid to meet new people
Avoidance	I like people only like us	Foreigners are <u>too</u> different	He might be a bad influence

Following this rationale a set of items was then constructed and included for tryout in the measurement package for Goal II.

The second phase of our work on Goal II is yet to be accomplished. We are ready to study the influence of school, community and student characteristics on the incidence of the behaviors represented by measurements for Goal II. The output data for Goal II will be correlated with data for each of the characteristic variables. Those factors which seem to show the highest relationships will then be used in defining the categories for subsequent measurement on this Goal. Concurrently with this study, we plan to involve school personnel and laymen in further discussion of the Goal and its measurement implications. Norms for students within each category, so defined, will then be collected. When this step is completed, we will be ready to begin the actual task of assessing the accomplishment of progress toward Goal II on the part of individual school districts. The appropriate measurement package will be administered and classifying data collected. The results will be reported to each district in the form of degree of agreement with expectations. The expectations will be based on data from districts having similar conditions of operations.

Staff from the Bureau of Quality Education Assessment will assist the local staff in considering the implications of the comparisons. Where differences too large to be attributed to chance exist, the Bureau's staff will work with local staff in developing hypotheses to account for the differences and, in the instance of positive differences, in suggesting ways to disseminate and further test the processes which seem most likely to have contributed to the successful performance.

In the case of negative differences a similar procedure will be followed. First, a hypothesis will be developed to account for the differences and then, depending upon whether or not the school district is in a position to directly influence the presumed causes, either compensatory programs or modifications will be suggested. Other Bureaus in the Department of Public Instruction, particularly the Bureau of Curriculum Development, and the Bureau of General and Academic Education will contribute these suggestions. Universities and state college instructional centers will also be involved.

After modifications of process or program have been attempted, the Bureau will assist as needed in program evaluation. For example, suppose a school finds its students lacking openness in dealing with people of different ethnic backgrounds. Given this situation, a school might engage in some type of a cultural interchange through interschool visits or exchange student programs. The effect of this compensatory program can be evaluated against the normative data as well as against the previous measurement in the same school. It is expected that as more and more schools experience this kind of assessment, processes which are particularly constructive will emerge and may be disseminated across the state to all districts who recognize their need for modifications.

In summary, the Pennsylvania Plan for assessing educational quality consists of three tasks. The first of these, the selection and development of measurement devices, has occupied most of our time to date and is near completion for an April tryout. The second task of classifying data according to modifying conditions is also scheduled for the coming Spring. The establishment of norms according to these conditions will follow probably in the Spring of 1969. The third task of providing services to local schools will not be accomplished until actual assessment gets underway in the 1969-70 academic year.

If the Plan as conceived is realized in Pennsylvania, we shall then begin to have objective evidence about the adequacy and efficiency of the educational programs offered by the public schools of the Commonwealth, with an emphasis on the individual school and its capacity for meeting the specific needs of the specific children it is to serve. There are many problems yet to be solved. Some of them are these:

1. We are relying heavily on self-report for measurement in the less tangible areas, for example; We need to develop ways of assessing student output in these areas which can be considered both valid and economical. Perhaps self-report will meet these specifications, but at the very least we need to conduct additional validation studies.

2. Some of the commercially published materials are at best a compromise fit to our measurement rationale. We need to locate or devise new materials which will better suit Pennsylvania's Goals.

3. Since continued measurement is required, we must develop or select comparable alternate forms of all measurement devices.

4. Many areas of the curriculum are as yet unmeasured. Example: Technical and vocational skills and high school electives.

We are anticipating your understanding and assistance with these and other problems. We expect that we can solve them together.

Paul B. Campbell

REFERENCES

1. Educational Testing Service. Volume one: the basic program. A plan for evaluating the quality of educational programs in Pennsylvania: A report from Educational Testing Service to the State Board of Education, 1965.

2. Tyler, R. W. Assessing the progress of education. Science Education, 1966, 50, 239-242.

3. Mood, A. L. Measurement of quality in education. Science Education, 1966, 50, 279-283.

WHAT WILL QUALITY
EDUCATION ASSESSMENT MEASURE?

During the last two weeks of April and the first few days of May, approximately fifteen hundred (1500) fifth graders and fifteen hundred (1500) eleventh graders in one hundred (100) schools throughout the Commonwealth will be administered an Assessment Package which I shall be discussing today and you will be reviewing tomorrow. However, as Dr. Campbell stressed in his presentation, adequate assessment of educational quality must take into account the context in which education takes place. In order for us to be able to test for any relationships among what a student brings with him, what a school can offer him, and what a student achieves, we must collect what we refer to as input variables. These input variables are school, community and student characteristics which will be used as correlates of achievement in each of the ten goal areas.

In order to choose from among the hundreds of available school, community and student characteristics, we reviewed five of the most recent major educational research studies.

The first of these, The Equality of Educational Opportunity study by James Coleman and others for the U.S. Office of Education, is currently one of the most controversial, widely discussed, and widely disputed.

The survey concludes that schools bring little influence to bear on a child's achievement that is independent of his background and general social context; and that this very lack of an independent effect means that the inequalities imposed on children by their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life. It is a grim conclusion and we shall be joining the ranks of the many who will try to show that schools do make a difference.

The Fels Institute Study on urban school districts in Pennsylvania sought to measure differences in educational requirements of public schools in urban districts, compared with non-urban districts. Their findings show that substantial numbers of urban pupils present significantly different educational needs related to their socioeconomic backgrounds and such needs require new, expanded or intensified educational programs and services which are costly. Correlations between achievement and socioeconomic characteristics were most significant in urban districts and less significant in non-urban districts.

The extensive testing and research studies that have been taking place in the state of Iowa have indicated that the educational background of parents, the occupational background of parents and the IQ of students correlate most significantly with student achievement.

Findings from the New York Quality Measurement Project indicate that the variable most highly related to student achievement is the educational level of the mother. The New York Quality Measurement Project, begun in 1958, had many of the same aims and purposes as the Pennsylvania Plan. We visited Albany in November, hoping to be able to benefit from their experiences, only to find that the Project is in its last year.

Project TALENT is impressive not only for its size (the sample consists of five percent (5%) of the population of high schools in the Country) but also for its durability. The first data were collected in the Spring of 1960 and, despite the fact that the Project has recently moved from the University of Pittsburgh, they are alive and well in Palo Alto. Project TALENT contains a wealth of information if only one had the time to review its numerous volumes. We did review one of the major volumes to find that factors such as housing quality, teacher salaries, teacher experience and background, number of books in the library, and school size show relationships to student achievement.

We are collecting other student, school, and community variables which do not have support in the literature as being significantly related to student achievement, mainly because they are the kinds of characteristics most researchers do not collect. It is our opinion that if we are to show that schools do make a difference in what pupils know, think, and believe, we should pay more attention to the educational process and less attention to factors such as size, per pupil expenditure, and salaries, important as they have been shown to be. For example, we are asking teachers "What characteristic do you think actually counts in gaining professional recognition in this school system?" and "What characteristic do you think should count the most in gaining professional recognition in this school system?".

We are asking principals to check a listing of innovative practices as they pertain to their schools, practices such as independent study, teacher aides, work experience programs, and instructional television.

We are asking students "What is the best single way for a boy or girl to be important and looked up to by other students in this school?"

1. Being bright, well-informed and interesting
2. Doing well in tests and grades
3. Being an athlete or cheerleader
4. Coming from the right family
5. Being attractive

At this stage we are only dealing with hypotheses. We do not know which school, community, and student variables will show relationships to the achievement of students in Pennsylvania, but we are in the process of collecting a great deal of information in order to find out.

Given the Ten Goals of Quality Education, most of which pertain to attitudes and opinions rather than factual knowledge, how do we discover whether or not students in the Commonwealth are fulfilling the requirements of these Goals?

When we began six short months ago, one of the first things we did was contact each of the major test publishing companies, inviting them to send a representative to visit, in order for us to discuss our Project with them. We hoped that they would have measuring instruments, either in published form or in the developmental stage, that we would be able to use. Most of the companies accepted our invitation. However, as we began meeting with these test representatives, it soon became apparent that the Goals we were attempting to measure were not the subject areas into which the test publishing companies were putting their talents or their money. In the areas of self understanding, understanding others, interest in learning, good citizenship, creativity, vocational development, and preparation for a world of change, the test publishing companies had very little to show us.

Even more disappointing was the fact that they did not expect to have anything to show us in the near future. Their efforts are going into the areas of scholastic achievement with the claim, as we hear with most other products on the market, that, "this is what the public wants." They wished us luck and admired our courage.

If we were to measure the Ten Goals of Quality Education, we were going to have to find measuring devices somewhere else, or develop them ourselves. We do have an Assessment Package ready for April testing. With help from the Bureau of Research Administration and Coordination, most of the items we developed ourselves. For two of the Goals we are using unpublished inventories being validated in university settings.

Goal I states that Quality Education should help every child acquire the greatest possible understanding of himself and an appreciation of his worthiness as a member of society. We expect to find differences in the way a student feels about himself, about his school and social situations. The literature suggests that how a student sees himself, his adequacies, his inadequacies, his values, has a strong relationship to how he achieves in school. The student with the positive self concept will be able to focus attention on classroom activities. He will be able to accept and recognize his weaknesses, he will not feel less worthy as a person because of them and, most important, he will be able to capitalize on his strengths.

A second dimension to be measured in Goal I is based on the theory that the sense of control which a student feels concerning his environment is significantly related to how he performs in school. Does the student view his world as one in which he will be able to fulfill his hopes, his ambitions?

Can he feel personal responsibility for planning his future and proceed with direction and purpose, discounting for the most part, the element of luck as the deciding factor? Can he recognize that everyone faces barriers and limitations and these need not be immovable obstacles?

Goal II states that Quality Education should help every child acquire understanding and appreciation of persons belonging to social, cultural, and ethnic groups different from his own. Correlating the results of our Goal II questions with the student, school and community input variables will allow us to test certain assumptions. For example, where a student body is heterogeneous with regard to educational and occupational background of parents, are these students more likely or less likely to show greater understanding and appreciation of persons belonging to social groups different from theirs?

Does the growth, stability or decline of a community show any relationship to a student's understanding and appreciation of others?

Questions for Goal II center around social class, religious and ethnic differences. Our questions are worded in such a way as to not make references to any particular race, religion or social class. Although our sample schools have been, for the most part, positive and receptive to these kinds of questions, we are aware that it is in this area where apprehension and doubt are most likely to exist, and we look for your comments and advice in these matters.

Goal III states that Quality Education should help every child acquire to the fullest extent possible for him mastery of the basic skills in the use of words and numbers. There is no difficulty in finding a reliable and valid Scholastic Achievement Test. The difficulty arises when one tries to choose from among the many.

The problem is further compounded when one discovers that the typical Scholastic Achievement Test takes from three to five hours to administer. When we considered the fact that our aim was to administer the full battery in no more than five hours, it became obvious that we could not simply choose and administer a typical Scholastic Achievement Test.

A survey conducted by the Bureau of Guidance Services reveals that practically all students in the Commonwealth are administered an achievement battery in the elementary grades as well as the secondary level. Upon reviewing the survey more carefully, we discovered that the number of achievement batteries that were used centers around five or six major ones. We then diverted our attention to the possibility of using existing achievement test scores. The problem then became one of "How do we compare, say, scores from the Stanford Achievement Test, with scores from the Metropolitan Achievement Test, or with scores from the School and College Ability Tests?" The opinions of the test publishers were not very encouraging, but we have not abandoned the idea.

For our April testing, however, Harcourt Brace and World is building us a survey test based on the Stanford Achievement Battery which will yield a total score and take just one hour to administer.

Goal IV states that Quality Education should help every child acquire a positive attitude toward school and toward the learning process. In the positive aspect of this dimension the student will view education as being helpful toward obtaining a job, making decisions as a voter, enjoying leisure time activities, keeping informed of world events, participating effectly in community affairs, and maintaining a home of his own. He will express the interest and desire to graduate from high school. He will consider school and learning interesting, valuable, pleasant, and active.

He will express the opinion that learning does not end where formal education ends, and will express the desire to return to some kind of educational setting from time to time in his adult life.

Correlating the results of the Goal IV questions with the input variables will allow us, for example, to test whether or not students show more positive attitudes toward learning where teacher attitudes toward learning, innovation, and education are more positive. Does the size of a class or a school have any relationship to student attitudes toward learning? Where college attendance rate is higher, are student attitudes toward learning more positive?

Goal V states that Quality Education should help every child acquire the habits and attitudes associated with responsible citizenship. It seems as difficult to find an inventory to measure responsible citizenship as it is to find responsible citizens these days. We not only have an inventory, but a responsible citizen as well, for in this short time, Dr. Ernest Peters from the Bureau of Research has not only developed the Pennsylvania Citizenship Assessment but has also been able to administer it in two school districts to collect reliability data. Students will respond to questions such as:

> When my parents give me a job to do, I do it.
>
> If a store clerk gave me too much change, I would keep the extra money.
>
> A man has to cheat a little if he wants to get ahead in this world.

For Goal VI, understanding good health habits and the conditions necessary for the maintaining of physical and emotional well-being, we reviewed the relatively few standardized instruments and selected the California Health Behavior Inventory.

In order to expand our measurement in this area, we are considering the use of a teacher check list, whereby the teachers can check personal observations of grooming and health habits. We, perhaps, can also utilize health records which are available in every school.

Creativity, Goal VII, was defined by the Quality Education Committee as including worthwhile activities that the student initiates and pursues on his own, activities producing an outcome which is perceived by the child himself and by others as a contribution to some part of this world. Our concept of measurement in this area rests first on the theoretical basis that there is a student potential for creative output. Second, that there are conditions under which creative output is more likely to occur and, finally, that there are characteristics which are common to creative production. This theoretical construct thus includes the idea of the student's potential, the conditions under which he works and the quality of his output.

For the elementary student, Thomas Rookey, a research intern, had developed the Pennsylvania Appraisal of Creative Tendencies. Students will respond to a five point scale ranging from strongly agree to strongly disagree to questions such as:

> It's great fun to make up stories.
>
> It would be a waste of time to take a photograph through a fishbowl.
>
> You should not change the rules of the game.

The secondary student will check "yes" or "no" to a list of seventy-nine (79) activities, each of which has been assigned a quality rating. He will also be given a series of questions as a measure of creative potential.

Our efforts in this area, as well as everyone else's efforts in this area, are strictly experimental. There are many who maintain that creativity cannot be measured. In fact, there are those who maintain that many of our goals cannot be measured. At this time we are taking a more optimistic view and are operating on the opinion that using less than perfect instruments is better than limiting our evaluation to those areas where more perfect measuring devices do exist. The Committee on Quality Education was very emphatic in their defining of Quality Education to include the many attitudes, habits and accomplishments which are not easy to measure. To limit our measurement to those one or two goals which are relatively easy to measure would not be, in our opinion, fulfilling the wishes of the Committee nor the purposes of the Act.

Goal VIII states that Quality Education should help every child understand the opportunities open to him to prepare himself for a productive life and should enable him to take full advantage of these opportunities. Guidance counselors have long been interested in the area of vocational development and there have been many theories, many inventories, and many pages written concerning this process. However, the majority of this work has been with secondary students. Research in the area of vocational maturity with elementary students has been practically non-existent. We considered the question, "How can fifth graders reasonably be expected to behave in matters concerning choosing a life's work, choosing a life's style, making decisions and preparing for independence?" I had the pleasure of raising some of these questions with small groups of fifth graders in the surrounding area, and based on our experience with these students, together with the position taken by the Bureau of Guidance Services, we are hypothesizing that fifth graders are aware of different kinds of work and workers and have a growing understanding of the relatedness of educational and vocational opportunities.

By the eleventh grade the more mature student will show involvement in the vocational choice process by actively seeking information about the world of work. He will take the responsibility for making a career decision. He will not depend upon others to make these choices for him. Finally, the student will base his career choices upon his interest, ability, and aptitude using a realistic appraisal of his potential as a basis for his decision making.

Goal IX states that Quality Education should help every child to understand and appreciate as much as he can of human achievement in the natural sciences, the social sciences, the humanities, and the arts. Our items for Goal IX attempt to measure the understanding of human achievement, but we have neither found nor developed a measure of appreciation of human achievement.

By the time we reached Goal X, I sometimes think we ran out of steam, or ideas, for the majority of items we are using to measure in this area are being used to measure in other areas as well. Part of our difficulty may also be due to the fact that our understanding of Goal X seems to overlap with Goals IV, VII, and VIII. However, Dr. Martin Yanis from the Bureau of Research has developed several items which are unique. Goal X states that Quality Education should help every child to prepare for a world of rapid change and unforeseeable demands in which continuing education throughout his adult life should be a normal expectation.

Although being able to assemble this Assessment Package for your review tends to give us a feeling of having something completed, we must keep in mind that this Assessment Package is just a beginning. Test development for the Quality Education Project must be one of ongoing revision, and constant refinement.

Joan S. Beers

REFERENCES

1. Coleman, J. S., et. al. Equality of educational opportunity. Washington: Superintendent of documents, U.S. Government Printing Office, 1966.

2. Fels Institute of Local and State Government. Special education and fiscal requirements of urban school districts in Pennsylvania. Philadelphia: University of Pennsylvania, 1964.

3. Iowa testing programs research report. University of Iowa, 1966.

4. Project TALENT. A survey and follow-up of educational plans and decisions in relation to aptitude patterns: studies of the American high school (cooperative research project No. 266.) Palo Alto: Project TALENT Office, 1962.

5. Wohlferd, G. H., Armstrong, C. M., and Curtis, L. P. Scholastic achievement and relevant factors. The University of the State of New York: The State Education Department, 1967.

PROGRESS REPORT ON THE
NATIONAL ASSESSMENT PROJECT

Educational evaluation has been undergoing profound changes in both theory and practice in the past two decades. These fundamental developments are to be found in the several major uses of educational measurement and evaluation, such as in guidance, admission and placement, and in the awarding of scholarships; in the diagnosis of student learning; in developing programs, courses, instructional procedures, materials, and equipment; in monitoring and guiding educational programs in the school and in assessing the progress of education to provide bases of public understanding and public policy. These changes have arisen both from the changing educational situation and from the development of new knowledge and technology in education.

The applications of science and technology in agriculture, industry, defense, commerce and the health services have shifted the nature of human occupations from those largely based on physical strength and manual dexterity to those involving large components of intellectual activity and social sensitivity and skills. When I was born, thirty-eight percent (38%) of the U.S. Labor Force was engaged in agriculture. Now only seven percent (7%) is so employed. Then, another twenty-three percent (23%) were utilized as unskilled labor in non-farm occupations. Now only five percent (5%) is so employed. Very few young people can find jobs who have not attained the functional literacy represented by the average achievement at the end of the fifth grade.

At the other extreme, the great employment opportunities are in science, engineering, education, the health services, recreation services, social services, management, and accounting. In our time, the role of the school has shifted from that of selecting out a small percent of the pupils for more advanced education, while the others dropped out and went to work, to that of reaching every child effectively to enable him to go on learning far beyond the expected level of twenty-five (25) years ago. The task of the college is not to find a favored few, but to identify a wide range of potential talents and help each student to realize this potential both for his own self-realization and to meet the ever-increasing demands of a complex technological society. This changing situation requires new instruments and new procedures of educational evaluation and measurement.

New knowledge and technology in education are also influencing evaluation. For example, the recent findings of many studies regarding the powerful effects of the student's home, culture, and community environment upon his learning have clarified the need for evaluating these factors in order to guide and improve education. As another example, a series of investigations, like those of Newcomb and Coleman, have shown the strong influence of peer group attitudes, practices and interests upon the learning of its members. This indicates the need for evaluating the nature, direction and amount of peer group influences in developing effective school programs.

These highly significant changes have brought about new conceptions and new practices in several areas of education evaluation. I shall report on the developments sometimes called the "National Assessment."

The purpose of the project on assessing the progress of education is to provide that intelligent lay public with census-like data on the educational levels of important sectors of our population in order to furnish a dependable background of information about our educational attainments, the progress we are making and the problems that we still face in achieving our educational aspirations. Since education is now being recognized as a necessity for everyone in order to participate in our complex social, civic and industrial life, the public is asked to support programs for extending both the quantity and quality of education. Without perspective regarding the progress we have made and the difficult tasks we face, our citizens have an inadequate basis for making judgments. As a result, decisions are frequently made on hearsay or widely published assertions rather than on a reasonably clear picture of the educational situation.

Recognizing the lack of comprehensive and dependable data about the educational attainments of the various sectors of our population, Carnegie Corporation, a private foundation, in the Summer of 1964 appointed an Exploratory Committee on assessing the Progress of Education. I was asked to serve as Chairman. The Committee's Staff Director, for an initial six (6) month's period was Stephen Withey of the University of Michigan. Then we were able to obtain Jack Merwin on leave for two years from the University of Minnesota and since the Summer of 1967, Frank Womer of the University of Michigan has been Staff Director for the project.

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 more than three years. In 1966, the Fund for the Advancement of Education joined in supporting the project.

Since several reports have been made on the purpose and plan of this assessment project, I shall only outline them briefly. The conferences with teachers, supervisors and administrators resulted in strong recommendations that the initial assessment should include more than the 3R's and that ultimately all important educational areas should be included. Hence the four test construction agencies, American Institutes of Research, Educational Testing Service, Psychological Corporation and Science Research Associates have contracts to develop assessment exercises in reading and the language, arts, mathematics, science, social studies, literature, music, the visual arts, citizenship and vocational education. New instruments had to be developed because currently available tests are constructed to provide dependable means or median scores for groups but do not furnish information about the achievements of the lower and upper ends of school populations. Yet an assessment of educational progress should provide information about the progress being made by the less advanced and the more advanced children as well as those who are in the middle of the distribution.

Furthermore available tests are constructed to be given to every pupil normally in a classroom setting. To assess the progress of various segments of our population it is not necessary for every child to take every exercise. Samples of the population can furnish a dependable picture of the progress on the whole.

Some exercises can be performance tasks, some can involve observation or individual interviews as well as paper and pencil tests. This permits a more adequate assessment of the progress of our children, youth and adults toward the several major objectives of education.

The assessment will report on the educational achievement of two hundred and fifty-six (256) parts of the total U.S. population. These two hundred and fifty-six (256) parts are obtained by sampling separately each of four age groups: 9 year olds, 13 year olds, 17 year olds, and adults between 26 and 35 years of age; each of four regions: northeast, southeast, midwest and farwest; each of four sizes of communities: big cities, small cities, suburban areas, and rural-small town areas; two socioeconomic levels and the two sexes.

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

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 ten 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 two hundred and fifty-six (256) populations. In subsequent years, the progress made by each of these populations since the preceding assessment would also be reported.

In conducting this project, a number of technical problems have been encountered that are relatively new to educational measurement. In dealing with them, the Committee and its staff have not only had the benefit of the experience, intelligence and skills of the contractors but we have had very competent technical advice. The technical advisory committee consists of John W. Tukey, Chairman, who heads the Department of Statistics at Princeton University, Robert Abelson, Professor of Psychology at Yale, Lee Cronbach, Professor of Educational Psychology at Stanford and Lyle Jones, Professor of Psychology at the University of North Carolina.

Among these technical problems are the following: How do we communicate with children who come from home backgrounds of limited education and who are unable to read? What kinds of errors are introduced in administering the same exercises to some children individually and to others in groups? What biases, if any, are introduced by administering assessment exercises out of school in comparison to in-school administrations? How can a probability sample of children, youth and adults be chosen that represents the various sectors of the population and at the same time, minimizes costs of administering the assessment exercises?

These are a few illustrations of the many technical questions that have arisen.

The National Opinion Research Center has conducted a study for us, testing the feasibility of administering the exercises to adults in the home and also in small group settings. Surprisingly enough, adults are interested in taking the exercises, and sessions of one and a half (1½) hours with each adult are quite feasible.

The Research Triangle Institute of Duke University and the University of North Carolina are not only designing and testing the sampling procedure for identifying those who are to be assessed but the Institute also is conducting comparative investigations of assessing youth in school and out of school. The American Institutes of Research are conducting studies involving individual administration of exercises followed by interviews with children tested to check the comprehensibility of directions and wording of exercises. They will also obtain findings regarding pupil motivation in taking various kinds of exercises.

Educational Testing Service is conducting extensive tryouts to check the feasibility of various types of exercises and also to obtain indications of the range of achievement levels for which given exercises are appropriate. This special study is a large and comprehensive one.

These illustrations should suggest the kinds of published reports that will be available in connection with the study of these technical questions. They will be of value to students of educational measurement concerned with a wide variety of uses of evaluation.

In addition to furnishing greatly needed information to the general public, there are some special values of this project which were not anticipated when it was begun.

One of these is the development of assessment exercises appropriate for those children and youth in each age group who are among the lowest third in achievement. Because currently used achievement tests are focused on average performance, their exercises are concentrated at a middle band of difficulty. Only about five percent (5%) of the exercises in commonly-used tests are relevant to what is learned by students who are in the lowest third of their group. As a result, we have no way of evaluating the progress being made by this lowest third. A somewhat similar situation exists in the case of the highest third.

It was difficult for our contractors to construct valid exercises for both tails of the distribution because it has never been done before. The first efforts proved on tryout to be unsuccessful, but by obtaining the services of teachers and supervisors who are working effectively with children from low socioeconomic backgrounds to help in developing exercises and by careful revisions in the light of tryout results, it has been possible to produce assessment exercises that can appraise the educational progress of children in all three segments of the distribution. This will be of great value in the future as we seek to evaluate the effectiveness of programs that have been developed to serve special school populations.

A second kind of value of the assessment lies in the demonstration it will furnish of a wider range of useful evaluation procedures and studies than those that have commonly been employed. School testing programs now are concentrated on obtaining mean or median values for classes and schools and with the reporting of relative standings of individual pupils in standard scores, quartiles, percentiles, and the like. These have very limited usefulness in evaluating programs, curriculum materials, teaching procedures, learning and instructional devices and so on.

As a profession we give lip service to appraising progress toward all important educational objectives using a variety of evaluation techniques, but in fact, we do not commonly cover all our important aims, nor do we usually depend on techniques other than paper and pencil tests. The assessment project reminds us that we can make appraisals in terms of our several objectives, and it is demonstrating one way of doing it. It is showing us that we do not need to have every pupil take short, undefined or undesignated tests. It is demonstrating the feasibility of using samples of students so that the range and depth of evaluation can be greatly extended. It is showing us the possibility of systematically sampling the student behavior we are seeking to develop, at the particular stage of each child's progress, through the use of a range of devices including not only paper and pencil exercises but also observations, interviews, questionnaires, performance tests, and samples of pupil products. In my opinion, this contribution to the development of more adequate evaluation within our own schools may be the most important contribution of the current assessment project.

In summary, assessing the progress of American education provides a means of helping the public understand the instructional purposes, achievements and progress of our schools and furnishes the professional staff with means for evaluating much more comprehensively than in the past, the effectiveness of educational programs designed to serve particular purposes. Both the results and the procedures of the assessment project can make a constructive contribution to education.

Ralph W. Tyler

SUMMARY OF COMMENTS AND SUGGESTIONS

Made by the Advisory Committee for the
Bureau of Quality Education Assessment
at its First Meeting on February 1 and 2, 1968

Following the presentation by members of the Bureau on Thursday afternoon, the Committee divided into two groups to react to the Pennsylvania Assessment Plan.

Concern centered about the sampling procedure for the April testing. It was explained that the sample of one hundred (100) schools was randomly selected from within a total list of school districts stratified by market value of property and size of school population. One of the purposes for the April testing is to identify other significant dimensions by which school districts can be characterized. The data from the April testing will form the bases by which schools will be classified for the normative study, scheduled for the Spring of 1969.

For the most part, the Committee was favorable and supportive of the efforts of the Bureau, while at the same time cognizant of the inherent difficulties underlying any large scale assessment plan. Committee members lauded the work of the original Committee on Quality Education for emphasizing that factual knowledge is only one aspect of student achievement and that the feelings, attitudes and opinions a student learns are deemed equally important. Most of the members voiced their approval of the efforts of the Bureau to measure achievement in the attitudinal goal areas.

It was brought out, however, that one of the most significant drawbacks of the Plan is the reliance on paper and pencil, self report instruments. Several members of the Committee expressed the fact that the process of assessment can take other forms, such as observational techniques. Some Committee members expressed the opinion that measurement in Goal II, for example, might be more valid if the other techniques were used.

The Friday morning sessions were devoted to a review of the assessment package. With the questions in front of them, the Committee members were able to be more specific in their reactions to the measuring instruments. Their major concern centered about the language of the items-language in terms of difficulty level and in terms of semantics.

Although the Bureau applied the Flesch Reading Ease Chart to the majority of items for fifth graders, some of the Committee members felt that some fifth graders might not understand the terminology of certain items. Reading ease is a particularly important consideration. As Dr. Tyler noted, where a student does not understand the language of the questions, we cannot reliably assess the dimension.

The concern with semantics involves differences in meanings among different social classes. For example, the word "scolded" may have a different connotation for the middle class youngster than it has for the underprivileged youngster.

Some Committee members suggested that our items be reviewed by, for example, students, and, perhaps, teachers of the underprivileged, to detect such nuances in the terminology.

One further concern related to the wording of certain items in a manner which seemed to imply a value judgement of the behavior in question.

Other Committee members were concerned with "establishing a climate" for administering the assessment package. The consensus of opinion was that the Bureau should establish a list of "guidelines" for teachers concerning the nature of the assessment and its purposes. They maintained that although the actual testing will be done by monitors and not by teachers, it is the teachers, the day or so before the test date, who will influence the attitudes with which the students will approach the testing situation.

The more specific item revision suggestions involving grammar, punctuation and syntax, although not specified in this report, are also being utilized in the refinement of the instruments.

CONCLUSION

With the administration of the assessment package in the Spring of 1968, the first phase of the Pennsylvania Plan for the assessment of educational quality will be realized. The Bureau will then begin the task of analyzing the data in order to determine which school, community, and student characteristics are significantly related to the achievement of Pennsylvania students in terms of the ten goals of quality education.

At another conference tentatively scheduled for the Fall of 1968, we shall be examining in detail the results of the April testing. In light of our findings, we shall be considering the proposed goals of education in the Commonwealth and again shall be reviewing the assessment package.

We are confident that the Committee, together with educators and other citizens of the Commonwealth, can make a significant contribution to the vital task of assessing the quality of education in the public schools of Pennsylvania.

Neal V. Musmanno

"To be important, an educational outcome must make a difference. If it makes a difference, the basis for measurement exists." R.L. Ebel