

Education Assessment: Results a Step toward Accountability

Denver—For years leaders in politics and education have been searching for a way to make America's education system accountable to the public. Signs that the search may be producing results appeared with the release on 7 July in Denver of the first results of the National Assessment of Education at the annual convention of the Education Commission of the States.

The development of assessment has been a long and laborious process. Advocates of the project have had to tread the rough water between local school administrators who feared that national assessment would intrude on local autonomy and education reformers who wanted a system which would compare individual schools, reveal weaknesses, and spur improvement.

Assessment's first results brought no respite from the controversy. The *New York Times* lambasted the assessment as "one more example of extensive research that proves what everybody knows." It criticized the assessment for failing to draw concrete conclusions about specific school systems from its data. Yet the real indication of the success of these first partial national assessment results lies not in the conclusions they may have pointed to, but rather in testing the methods which produced them and which can now be used for assessments at the local level.

Testing Systems, Not People

The technique of assessment is a new and unique development in testing. It is designed to measure not how much an individual knows but, rather, what types of knowledge are possessed by various groups of Americans. Assessment is designed to test education systems, not the individuals in these systems.

Assessment was first proposed in 1962 by Francis Keppel, then U.S. Commissioner of Education. In 1964 an exploratory committee, funded initially by the Carnegie Corporation and headed by Ralph W. Tyler, who is director of the Center for Advanced Study in the Behavioral Sciences at Stanford, began an investigation into the feasibility of conducting a national education assessment. The program they

developed provided a major departure in educational testing.

The task which the exploratory committee faced in developing the assessment was formidable. From its first conception, the assessment was opposed by several education groups. The American Association of School Administrators criticized it as the first step toward a national curriculum and a limitation of local control over local school districts. Teachers feared that it might lead to a system of merit pay in which teachers were paid according to how well their students did on the national assessment tests. They foresaw an era in which education was aimed solely at achievement ratings on the assessment tests (*Science*, 5 May 1967).

In the face of this opposition the Tyler committee began to develop the assessment procedures. The committee designed tests in ten areas—science, citizenship, writing, mathematics, music, art, social studies, reading, and career and occupational development. Before preparing the questions for each test, the committee first developed a series of objectives which they thought should be the goals of education in the test's field. For the science test, for example, they stated four objectives for each age. These are: (i) "know the fundamental facts and principles of science," (ii) "possess the abilities and skills needed to engage in the processes of science," (iii) "understand the investigative nature of science," and (iv) "have attitudes about and appreciation of scientists, science, and the consequences of science that stem from adequate understanding."

In formulating the questions themselves and the procedures for asking them, the committee used several innovative testing methods. They developed a system of taped and printed exams and personal interviews to reduce the effects of reading ability on all but the reading tests. In conventional tests, where questions are presented in written form, a student is often penalized by a reading deficiency even when the test is designed to examine a subject other than reading.

Assessment exercises included both very easy and very hard questions to

make it possible to ascertain what knowledge most students have as well as what knowledge most do not have. Conventional tests contain mostly questions of average difficulty.

The most important innovation in the national assessment was the method of reporting the results. Standardized test results are reported in terms of how many questions an individual answered correctly. Assessment results are reported in terms of how many people answered a question correctly. The groups of people answering each question are broken down into censuslike data according to race, sex, family background, region, type of community, and age. Thus it is possible to tell from the tests which groups of people are weak in specific areas of their education.

Administration by the ECS

Once the testing procedures were developed, the administration of the assessment was turned over to the Education Commission of the States, which is an alliance of governors, state legislators, and education commissioners from 43 states and territories. The ECS was founded in 1966 and serves as a forum for state educators and politicians to discuss their common education problems. It also provides a common front for the states in their defense against federal intrusion in education. The ECS is dominated by politicians who are jealous of state prerogatives in education. Many states agreed to join only after being assured that the ECS would not infringe on their educational domains (*Science*, 3 December 1965).

At the ECS convention partial results of the first two assessment tests—science and citizenship—were released. Results were broken down only according to age. Breakdowns by other groupings have been promised for the future. Since these tests were the first in what is designed to be a periodic testing program, no measure of progress could be ascertained from them. Because of these limitations and because the techniques of assessment are still relatively new, the administrators urged caution in drawing any conclusions from the data which the reports provided. Administrators themselves drew only the most general kinds of conclusions, such as "17-year-olds know more than 13-year-olds."

Nevertheless, those at the convention could not resist making some more concrete observations about American education based on the assessment's

initial reports. Science teachers at the convention pointed out that students taking the assessment seemed to be weaker in physics and chemistry than biology. They called for new emphasis in the teaching of physical sciences.

Richard Webster, state senator from Missouri, said that the results of the citizenship test indicate a weakness in Americans' knowledge of state and local governments. He asked educators to reexamine their civics courses in light of this weakness.

Assessment can achieve its full potential for enforcing accountability only if it is applied locally. Then weaknesses in specific school systems can be corrected at the local and state level where power over education policy is lodged. National assessment, however, will not be used for evaluating local school systems. State politicians who control the national project are opposed to such an "intrusion into local jurisdiction." These politicians are, on the other hand, generally eager to use the methods of national assessment in conducting their own local assessments. The state or locality would fund, conduct, control, and take credit for the assessment. National assessment would provide the methodology. How widespread local assessment becomes and how meaningful it will be depend largely on whether assessment gains acceptance as a fair and accurate measure of educational achievement. If it gains such acceptance, those school administrators who have in the past avoided evaluation because of a lack of judging methods will be forced to bow before the public demand for accountability.

The assessment techniques still have a way to go before they gain that widespread acceptance. Included with the reports on science and citizenship were the comments and criticisms of two panels of assessment reviewers. The panels were made up almost exclusively of teachers. Among their criticisms were the following.

1) The assessment did not make any distinction between real and professed beliefs of the people it tested. In the citizenship test, for example, an overwhelming majority of those quizzed indicated that they would not mind living next door to a person of another race. Yet there was no way of telling whether those who answered in this manner really believed this or whether they were simply giving what they assumed would be considered the "correct" answer.

2) In setting its objectives for edu-

cation in citizenship, the assessment stuck mainly to conventional, idealized goals. In developing and testing the objective "know the main structure and function of our governments," they tended, for example, to overlook the views of blacks in Mississippi who may see the function of government as promoting the separation of races.

3) In some cases the assessors made mistakes in summarizing the information which the tests yielded. On one question 9-year-old children were told that big leaves often yield more water than little leaves. They were then shown pictures of five leaves of different sizes and asked to pick the one which gives off the most water. Eighty-nine percent performed the obvious task of picking the biggest leaf. The assessors then reported that this test proved that 89 percent of 9-year-olds *knew* that big leaves give off more water than little ones, despite the fact that that information was given as part of the question.

4) In the science assessment too many questions were devoted to assessing how much knowledge Americans have and too few to determining their attitudes toward science and their abilities to use scientific methods.

5) Some of the new testing techniques may affect the test results. In the questions which are asked verbally by an examiner, for example, it should be determined whether it makes any difference if the examiner is of a different race from the person being examined.

6) The assessors drew only the most generalized conclusions from their data. For the most part they left it up to the public to form their own opinions about what national assessment means for education. In effect this enabled every person with an ax to grind to draw his own particular preformed conclusion from the results. "The assessment proves what I have known for a long time," said one politician in a typical preface to an explanation of what assessment means.

Yet, despite these criticisms the method of assessment was received in general with approval even from some traditional opponents of assessment. There are strong indications that local and state officials will adopt the techniques for their own use. "I fully expect that assessment will go down to the state and local levels," said James E. Allen, former U.S. Commissioner of Education. "The public is increasingly demanding performance. I feel that teachers and other groups will be willing to accept evaluation of educational

performance once they are assured that there is a fair method of doing it," he added.

Ewald B. Nyquist, New York Commissioner of Education, indicated that he favors assessment, and called for a public system of comparing state and local school systems with respect to their educational achievements and progress.

How soon assessment will reach the state and local levels remains to be seen. But the national assessment administrators proclaim that their program is an ongoing one which will grow and improve. If it does, and if it gains widespread acceptance as a fair method for evaluating local school systems, it could be the long-sought means of providing accountability in education.

—THOMAS P. SOUTHWICK

RECENT DEATHS

George H. Conant, 73; former teacher of plant pathology, Ripon College and University of Pennsylvania; 15 May.

Raymond T. Ellickson, 60; professor of physics, University of Oregon; 31 May.

Ronald M. Ferry, 78; retired associate professor of biochemistry, Harvard University; 26 May.

Mary J. Fraps, 71; retired research professor, poultry science department, University of Maryland; 3 May.

Richard M. Fraps, 67; retired senior physiologist, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland; 9 April.

John C. Godbey, 87; retired professor of chemistry, Southwestern University; 14 June.

Alonzo F. Myers, 75; former chairman of higher education, New York University; 24 May.

Robert W. Ramsey, 63; professor of physiology, Medical College of Virginia, Health Sciences Division, Virginia Commonwealth University; 7 April.

Burrell O. Raulston, 84; emeritus dean, University of Southern California School of Medicine; 27 May.

Harold C. Taylor, 64; director, W. E. Upjohn Institute for Employment Research; 6 May.

Erratum: In "Charge-mosaic membranes: dialytic separation of electrolytes from nonelectrolytes and amino acids" by J. N. Weinstein and S. R. Caplan (17 July, page 296), Eq. 1 on page 296 should read

$$\omega_s^* = \left(\frac{J_s}{\Delta\pi_s} \right) \Delta p = 0$$