DOCUMENT RESUME

ED 108 994 SO 008 368

AUTHOR Fleming, Dan B.
TITLE The "New" Social Studies and Standardized Testing.
PUB DATE Mar 74
NOTE 12p.

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Evaluation Methods; Measurement Techniques; *National Competency Tests; Predictive Ability (Testing); Program Evaluation; *Social Studies; *Standardized Tests; *Student Evaluation; Test Bias; Test Construction; *Testing Problems; Test Interpretation; Test Reliability; Test Selection; Test Validity

ABSTRACT

In this paper, the author reviews a number of criticisms that have been made of the use of standardized tests in the social studies and reviews 11 general, skills, and discipline-oriented standardized tests for social studies. Standardized tests evaluating the results of the new social studies courses are open to criticism on the basis of validity, utilization, and restriction of educational change. They culturally discriminate against some individuals, predict imperfectly, and are often rigidly interpreted. The tests may be assumed to measure innate characteristics and thus may influence teacher expectation regarding student potential. They have a harmful effect on the shaping of the student's cognitive style and often are the basis for school organization and curriculum. Further, standardized tests distort the individual's privacy in later life. Finally the results of standardized tests, regardless of their inherent weaknesses, are used as a basis for allocation of resources. The annotated bibliography lists 11 commercially available standardized tests for skill, discipline, and general social studies testing. (Author/DE)

THE "NEW" SOCIAL STUDIES AND STANDARDIZED TESTING

Dr. Dan B. Fleming

Virginia Polytechnic Institute and State University

March, 1974

The "New" Social Studies and Standardized Testing

A word rapidly looming larger and larger on the Virginia educational horizon is that of accountability. The Virginia Standards of Quality for Public Schools call out for the achievement of specific objectives relating to student performance. This demand for accountability requires ways of evaluating instruction. The first method usually selected is that of the standardized test. As a result, a school system could possibly end up with the tail wagging the dog or a standardized test dictating curriculum. This problem can be a severe one for the social studies teacher. Does the standardized test selected evaluate the social studies objectives of a local school system? If not, why use it at all?

There are numerous criticisms that can be made of standardized tests. Listed below are ten of these identified by Holman and Docter.[1]

"Five Criticisms Bearing on the Validity and Utilization of Tests

1. Tests discriminate against some individuals. It has been strongly argued that some testing programs have consistently failed to take into account differences in cultural background and in unique individual attributes. Such failure unquestionable influences test results and may, therefore, penalize the testees.

A major concern is whether tests developed primarily for use with Caucasian subjects can properly be administered to minority-group members. Many of the latter may have educational and cultural backgrounds markedly different from those of the subjects used in the standardization of any particular test.

Employment-selection tests have especially been denounced by minority-group representatives as too often containing built-in bias which favors the middle-class white person and discriminates against the minority applicant. While respected testing professionals may disagree in the interpretation of specific data purported to prove or disprove this point, they agree that tests lacking in job-related validity have no place in selection-and-placement testing programs.

2. Test predict imperfectly. No standardized tests are perfect predictors of future behavior. Even the most enthusiastic proponent of objective assessment techniques would insist that his ability to foretell behavior is highly dependent on such factors as the individual(s) to be tested, the behavior to be predicted, the time over which prediction is to be attempted, and the criterion measures used to establish predictive effectiveness.

But even with all these qualifications, critics of testing have come to the conclusion that many tests are weak and unsatisfactory devices which mislead naive test users and result in harm to those tested. Many critics have just about given up on tests, for they see them as falling far short of the ideal applications envisioned by their creators and their publishers.

The problem of test validation encompasses many issues that go beyond establishment of certain formal psychometric properties

[1]Docter, Richard F., Holman, Milton G., Today's Education, National Education Association, February 1974, pp. 50-60.

which may be present to some extent in any tests. The proper use of tests must encompass a variety of responsibilities independent of the attributes of any particular test. We must not only ask whether a test has been shown to possess some kind of validity for a known group of subjects, but also must investigate many other questions bearing on the particular circumstances surrounding the application of the test.

3. Test scores may be rigidly interpreted. Test scores provide one opportunity to establish a data base for the arbitrary classification of individuals. Anyone interested in labeling people can have a field day with test results. This fact notwithstanding, the properly trained user of tests is supposed to know that test scores are not fixed measures, that they are estimates of human attributes at best, and that they necessarily encompass various kinds of sampling errors.

But test scores are often applied in rigid and arbitrary ways. In schools, this can result in assignment of children to ability groupings based on measures which may be indefensible. The quality of professional practice associated with tests usage leaves much to be desired.

4. Tests may be assumed to measure innate characteristics. Some critics of ability testing have argued that tests provide scores that may be naively interpreted as measures of innate characteristics, such as "intelligence"; many harmful consequences are said to flow from this misconception. It has occasionally been assumed that, if tests were not available, people would not make arbitrary classifications of individuals. Tests are therefore condemned as anti-humanistic and as fostering a view of mankind that sees human abilities as fixed or rigidly limited.

Even worse, some critics have reasoned that tests influence individuals to conceive of man in categorical terms, such as "mentally retarded" or "gifted." They conclude that thinking of this kind is undesirable.

At first glance, this seems to be nothing more than a variation on the practice of making rigid use of test scores. The essential difference, however, as expressed by some critics, is that not only do tests foster the belief that man has fixed "intelligence" based on innate characteristics, but also that the use made of test scores depends heavily on such a belief.

The kind of school program offered and the energy invested in preparing a youngster for the future may be directly influenced by an educator's belief that tests measure innate intelligence. The egalitarian ethic in America frowns upon labeling based on some arbitrary measurement supposed to reflect innate characteristics.

5. Test scores may influence teacher expectation regarding student potential. In a classical study, Robert Rosenthal and Lenore Jacobson showed that, when teachers' expectations regarding student potentials were based on fictitious information about the students' abilities, the actual achievement of students reflected these expectations. Those who were expected to achieve less did achieve less, and vice versa.

Critics of ability testing have argued with considerable force that tests of "intelligence" have highly undesirable consequences for student performance because at least in part, teachers tend to relate to students differentially, according to their supposed intelligence. Students who are singled out as "gifted" or "low ability" are given different assignments, rewards, and teachers, and they are systematically taught what is expected of them.

There seems little argument that teachers' expectations contribute to student performance. It is less clear what factors shape teacher expectations. Test scores may be important in determining differences among students for some teachers; however, we need to know far more about the entire matter of teacher expectancy, for many other variables may help to determine their attitudes.

Five Criticisms That Are Independent of Test Validity

6. Tests have a harmful effect on the shaping of cognitive styles. The widespread use of multiple-choice test items, matching items, and other test components with a single correct answer is said by some critics of testing to contribute to undesirable styles of thinking. Some claim that the young student is carefully taught that all problems must have a right or wrong answer, and thus the student is led to think in this manner about all questions.

7. Tests shape school curriculums and restrict educational change. When teachers know that the evaluation of their students will be based on a particular kind of test of some more or less predictable content, they make extensive efforts to assist their students to perform well of these tests. The proponents of statewide testing programs would probably argue that this is exactly what they have in mind -- that teachers ought to be encouraged to cover material which their colleagues consider essential. "What's wrong with this?" they ask.

Critics of testing say that experimentation with new ways of teaching, the introduction of new subject matter, and the whole process of individualizing instruction in terms of the needs and interests of individual students are hamstrung by a slavish adherence to standardized achievement testing. The question seems to come down to finding an acceptable balance between the need to know what has been learned during a given period of time and the encouragement of innovation, change, and experimentation in the classroom.

8. Tests distort the individual's self-concept and level of aspiration. Of all the criticisms of tests, one of the most penetrating and difficult to dismiss is that young persons may generalize from test results and make conclusions about themselves which are not warranted or intended. For example, consider the teenager who may be struggling to establish a more positive and more realistic self-concept. How helpful is it for him to be shown his low test scores which may make him conclude that he is far less capable than his classmates?

How many high school students have received brief and inappropriate counseling recommendations, usually based in part that they are not "college material"? One large school district, for example, regularly presents junior high school students with test result summaries printed on cards that the students take home to their parents. These cards offer a lucid and easily understandable summary of what the various achievement and aptitude scores mean. Although the intent is to make information available to parents, there are obviously risks in terms of shaping the attitudes of students toward themselves.

In our view, the proper handling of test results calls for neither a strategy of silence and secrecy nor for open distribution of data without discussion, clarification, and interpretation of meanings.

9. Tests select homogeneous educational groups. A common procedure in organizing a school is to assign students to classes on the basis of estimates of learning ability. Very often these estimates are based on ability testing. It is a short step to conclude that tests have determined the organizational style of schools, and it may surely be argued that tests do indeed contribute to the way in which students are assigned.

Critics of the ability-track system, as this arrangement is often called, frequently see educational testing as the bad guy. But, were no test data available, an educational administrator dedicated to ability-track grouping could find numerous critera, such as grades, teachers' ratings of ability, and so forth, for making these assignments.

Concerns about homogeneous grouping in schools have acquired strength with recent research which suggests that this allocation procedure tends to do more harm to the low groups than can be justified. The proponents of heterogeneous assignment to classes argue that children with lower ability need the stimulation and the role models provided by higher-ability students if they are to achieve as much as they possibly can.

Contemporary approaches to school organization stress the importance of providing a program of individual instruction for each child, regardless of the range of competences within a class. Educators are now stressing the positive influences of heterogeneous grouping, with the result that the track system is generally thought to be on the way out. But for the parents of children who are

assigned to low groups, the track system is an unpleasant reality based primarily on test results. Hence, since tests are often painted as the villain in the situation, it is assumed that banning tests will eliminate the track systems.

However, with regard to a school district set on the perpetuation of homogeneous ability grouping, the problem is not so much one of testing or not testing, but rather one of adherence to a questionable concept of educational organization.

10. Test invade privacy. School attendance is mandatory for young children. Once in school, the children are generally required to participate in activities, including testing, which some parents consider to be invasions of privacy.

Certainly few would argue against allowing schools to give tests to determine what a student has learned in some course of study, but should schools be allowed to require students to take intelligence tests? What good is such information to a school? Can data from some tests be used to the disadvantage of a student without his knowledge that such information even exists? How can the line be more clearly established between information that a school requires to help reach a legitimate decision and information that it has no business acquiring in the first place? The right to privacy is precious to the citizens of a free society; only when there is compelling justification should tests invading privacy be used."

The most significant criticism of standardized tests according to Holman and Docter is that tests too often serve as "gatekeepers" in the allocation of resources. Of major concern to all teachers is the importance given a standardized test that fails to evaluate the program objectives of the curriculum.

"In many testing situations the large majority of the available tests can be discounted due to inappropriateness of content or grade level, or inadequacy or lack of norms or technical data. If only the tests of the major publishers offering social studies tests are considered, the choice is further narrowed; in fact, in some content areas there is no choice at all at some grade levels. (These statements pertain only to separately published social studies tests. If use can be made of a test battery yielding a social studies subscore in addition to other scores, a few other tests may be considered.) If a test is wanted for use in a formal evaluation program, selection will in many cases be limited to the tests of the larger publishers. These tests, though they too may have limitations, are likely to show greater care in development and standardization than do most of the lesser-known tests. For less formal evaluation, choice among lesser-known tests may be feasible and in some cases necessary. Certainly, some of these test could be considered for informal practice or review.

In examining available tests for possible use, the teacher should give consideration to (1) such practical matters as publication date, cost, and the time needed for administration, and (2) such technical matters as reliability, validity, norms, and test difficulty. Test manuals should provide information on both points, although statements concerning reliability, validity, and norms may be less than adequate. As important, however, is for the teacher to study carefully each of the tests receiving consideration; the quality of items and what the test seems to "cover" in terms of content, skills, and understandings. Such subjective analysis is of utmost importance to insure that the test selected meets the purpose for which it is to be used."[1]

For a detailed explanation of each of the points to be considered, see the thirty-fifth yearbook from the National Council for the Social Studies, Evaluation in the Social Studies.

The NCSS yearbook referred to above was published in 1965. Included as an Appendix to that book was a "Bibliography of Social Studies Tests." Presented below is an updating of that bibliography. The views offered concerning these tests is that of the reviewer. As a general statement about most of the tests reviewed, they seem to stay remarkedly the same ten years after the last bibliography was written. For a school system supporting an inquiry approach for social studies, there is considerable doubt that the bulk of the tests reviewed below are suitable. In fact, the use of many of these tests may do definite harm to a "new" social studies program. Many commercial tests are not treated in this bibliography, but those most widely used are included.

BIBLIOGRAPHY OF SOCIAL STUDIES TESTS

This bibliography is not intended to be all inclusive for the many commercial social studies test available. It does include many of the most widely used commercial tests. Only tests available in separate booklet form are included. The format used here parallels that used in the 1965 Evaluation in Social Studies Yearbook. No attempt is made to examine the tests for cultural bias.

GENERAL SOCIAL STUDIES

1. Achievement Series: Blue, Green, and Red Levels. Consists of three separate, but overlapping levels of graduated difficulty intended for use in grades 4 through 9; 1971; each of the three test has 40 items; Robert A. Naslund and others; Science Research Associates. Each test is very readable. Half of the items in each are devoted

---

[1] Berg, Harry D., Editor, Evaluation in Social Studies, National Council for the Social Studies, 1965, p. 220.

to elementary skills in cartoon, picture, map, and graph interpretation. The map questions are very simple and quite limited. The remaining half of the tests are general social studies information questions requiring the recall of specific facts. The tests are not challenging and are of no particular value to a school system featuring an inquiry approach to social studies. Accompanying the series is an excellent teacher's guide, Using Test Results.

2. Iowa Tests of Educational Development: Social Studies Background. Grades 9-12; 1972; 2 forms; E.F. Lindquist and others; Science Research Associates. A general social studies developmental test. It deals primarily with contemporary political, economic and social affairs. Since only thirty items are used, this provides a very sketchy cross section of great amounts of content areas. Emphasis is given fundamental ideas and terms such as the definitions of culture and the meaning of capitalism. The test attempts to determine what social studies knowledge is of most worth and tests for same. Very little attention is given world cultures, history or geography. No questions deal with skills. This is a test of limited value for an inquiry oriented program.

3. Sequential Tests of Educational Progress: Social Studies. Grades 4-6, 7-9, 10-12, 13-14; 1969; 2 forms; Cooperative Tests and Services, Educational Testing Service. Claims to test "skill in organizing, interpreting, and evaluating information as well as knowledge of many concepts of history and the social sciences." The tests for grades 4-6 and 7-9 consist of fifty questions each. Questions deal with picture interpretation, map skills, cartoon interpretation, graph reading, determining main theme, and reading charts. Both tests are very process skills oriented and should be useful for a inquiry oriented social studies program. Both tests are readable and do not depend on recall of information. The 10-12 and 13-14 tests consist of two parts, 35 questions in each part. As with the lower level tes ; major emphasis is given map, graph, chart, photography, and reading interpretation skills, only going into more depth.

4. Minnesota High School Achievement Examinations. Grades 7-12; Social Studies Grade 7, 1970; Social Studies Grade 8, 1973; Social Studies Grade 9, 1970; Social Studies Grade 10 (American History) 1973; Social Studies Grade 11(World History) 1973; Social Studies Grade 12 (Introduction to Social Science) 1973, American Guidance Service, Inc.
   Grade 7 - Consists of 102 items following a chronological survey of American history. Questions ask for low level recall answers. No skills questions asked.
   Grade 8 - Consists of 100 items drawing from world geography. Ten items deal with map reading skills. The remainder are generally low-level recall questions dealing with different world regions with two sections on conservation and world affairs.

Grade 9 - Consists of 82 items designed for a civics course. Topics included are national and local government, occupational and educational planning, and current events. Basically low-level recall questions asked with no skills questions.

Grade 10 - Consists of 114 items drawn from American history. Generally low-level recall questions with no skills questions.

Grade 11 - Consists of 115 items drawn from world history. Organized into 12 units dealing with different areas of the world with the last unit on "Pressure of Population Upon World Resources." Generally low-level recall questions with no skills questions.

Grade 12 - Consists of 90 questions centered around the following units; "Achieving Personal and Social Competence," "Man and Society", "Being An Effective Citizen In Our Democracy," "Living In An Industrialized Society,""America: A Land of Many People". Questions are generally low-level recall with no skills questions.

5. Stanford Achievement Tests: Intermediate and Advanced Battery. Two Forms. Intermediate II test, grades 5.5-6.9. Advanced Battery, is grades 7.0-9.9. Truman L. Kelley and others, 1964. Harcourt, Brace Jovanovich, Inc.
The Intermediate II test is divided into two parts. Part A, Content, consists of 45 items, mainly low-level recall questions. Several deal with value choices that are debatable as to suitability. Part B, Study Skills consists of 28 items dealing with reading tables, graphs, maps and interpreting a political poster. Part A is rather dry and seems a carry over from textbooks of the fifties.

The Advanced Test, grades 7.0-9.9 is organized the same. Part A, content consisted of 52 items and Part B, Skills, has 39 items. The content questions are similar to those on the Intermediate Content section. Skills included are map and graph skills; the use of a textbook bibliography, a library index card, and the use of the index.

The tests do not seem suited for evaluation of a process oriented social studies program. Although the skills section is generally useful, the quality of the content section tends to negate the value of the skills section.

SKILLS TESTS

6. ANPA Foundation Newspaper Tests. Consists two levels of tests, two parallel forms each. Level one is for the junior high school. Level two is for the senior high school. 1972, American Newspaper Publishers Association Foundation, Education Testing Service. Each test has 30 items based on a four page simulated newspaper. Students read the newspaper as they take the test.

7. Watson-Glaser Critical-Thinking Appraisal. Grades 9-12, College, Adult. 1964, Goodwin Watson and Edward Glaser. Two Forms, Harcourt, Brace, Jovanovich, Inc. The Watson-Glaser test consists of 100 items centered around five areas: drawing inferences from a summary of facts; recognizing assumptions implied by a statement; reasoning logically by interpretation; and discriminating between strong and weak arguments. The test is unique in its effort to measure critical thinking. The test is generally too difficult for the poorer reader and the instructions are somewhat difficult to understand. The test seems to have value in determining logical thinking for a limited advanced clientele.

DISCIPLINE TESTS

8. Test of Economic Understanding. High School and College. Two Forms, 1963, the Joint Council of Economic Education, Science Research Associates. The test seems much the same as that reviewed in the 1965 Evaluation Yearbook. It consists of fifty items including three questions on chart interpretation. Emphasis is on the U.S. economy with a few questions on international economics. Many questions test for understandings beyond the recall level. Test could be used for pre-post test of a course in high school economics.

9. Crary American History Test, Revised Edition. Grades 10-13. Two Forms. 1965. Ryland Crary, Harcourt Brace Jovanovich, Inc. The test consists of 76 items dealing with the major periods of American history ignoring the past 25 years. Four items interpret a quotation; three items deal with best sources; and six items deal with locating events on a U.S. map. The remainder are primarily recall questions. Generally not very useful for an inquiry oriented program.

10. Cooperative Social Studies Tests. Grades 7-12. The series includes: American History (Jr. High School), Civics (8-9), American History (Sr. High School), Problems of Democracy (10-12), American Government (10-12), Modern European History (10-12), World History (10-12). Two forms, 1964, Educational Testing Service. The tests generally try to combine content and skills items. Each test has 65 to 70 items with working time for each estimated at 40 minutes.

Skills items generally include map, cartoon, graph and data interpretation. The history tests also include time lines. Much of the tests depends on low-level recall and there is no inquiry emphasis evident in the tests. The tests have some value but should be updated.

11. Topical Tests In American History. Grades 10-12, 1963, Educational Testing Service. The series consists of eight tests. The topics are presented in chronological order going from Test 1: Exploration, Colonization, and Independence to Test 8: The Second World War and After. The tests run consist of 60 items and stress ability to interpret data such as quotations, maps, and cartoons. Unit 8 needs to be updated. A useful handbook accompanies the series. The tests could be useful although some updating of content and approach would be helpful.