DOCUMENT RESUME

ED 202 367 HE 013 881

AUTHOR Fincher, Cameron
TITLE Beyond Bakke: The Positive Benefits of Testing.
INSTITUTION Education Commission of the States, Denver, Colo. Inservice Education Program.; State Higher Education Executive Officers Association.
SPONS AGENCY Kellogg Foundation, Battle Creek, Mich.
REPORT NO IEP-502
PUB DATE Oct 78
NOTE 22p.; Paper presented at a Seminar for State Leaders in Postsecondary Education (New Orleans, LA, October 1978).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Admission Criteria; Behavioral Objectives; *Competency Based Education; Court Litigation; *Educational Assessment; Educational Counseling; *Educational Testing; Higher Education; Instructional Improvement; *Outcomes of Education; Program Evaluation; *Standardized Tests; Student Development; *Student Evaluation; Student Placement
IDENTIFIERS Bakke v Regents of University of California; *Seminars for State Leaders Postsec Ed (ECS SHEEO); Supreme Court

ABSTRACT

The uses and applications of testing in education and the implications of the Bakke decision are considered. Pluralism and diversity in education have convinced many observers that a different philosophy or theory of testing is needed. Alternative uses are sought in ways that would shift the aptitudes and general intelligence to skills and competencies. The changing emphasis from measurement to assessment reflects the educational need to appraise those changes in behavior and performance that presumably are the result of learning and development. Differences of ranking are not needed as much as an acceptable way to determine if learners have mastered some standard of performance. Testing concepts and methods are needed to help appraise instructional methods and materials, program structure and content, and overall program effectiveness in bringing about desired outcomes and results. The intent of educational testing needs to concentrate on accomplishments rather than potential. It is suggested that positive uses of tests can result from a shift from selection per se to placement, advisement, counseling, and program assessment. Specific uses served by testing include instructional improvement and learning facilitation and assessment of: basic skills, academic competencies in general education, college academic programs, learning difficulties, and learning outcomes. The Bakke decision indicated that if achievement of diversity for educational purposes is to be an admissions function, colleges must prepare to deal with human qualities that are complementary to those that can be handled within traditional testing contexts. It is suggested that the Supreme Court case provides an important context in which public policy concerning the uses of standardized tests must be debated. (SW)

Inservice Education Program (IEP)

Paper Presented at a Seminar for State Leaders in Postsecondary Education

BEYOND BAKKE:
THE POSITIVE BENEFITS OF TESTING

CAMERON FINCHER

Institute of Higher Education
University of Georgia

New Orleans, Louisiana
October 1978

IEP Paper No. 502

Education Commission of the States

Inservice Education Program (IEP)
Education Commission of the States
1860 Lincoln Street, Suite 300
Denver, Colorado 80295

SHEEO

The IEP Program has been supported primarily by the W. K. Kellogg Foundation with additional funds from the Education Commission of the States, the Frost Foundation and the State Higher Education Executive Officers

BEYOND BAKKE:
THE POSITIVE BENEFITS OF TESTING

Cameron Fincher
Institute of Higher Education
University of Georgia

In Bakke, as in previous cases involving standardized tests, the Supreme Court has shown an obvious degree of deference. There has been explicit deference to Congressional intent and administrative guidelines (Griggs v. Duke Power) and an appreciable reluctance to enter the fray of testing controversies (Albemarle v. Moody; Washington v. Davis). While the extent to which the technical or arcane features of standardized tests can be debated in a court of law is open to serious reservations, Supreme Court rulings provide an important context in which public policy concerning the uses of standardized tests must be debated. Policy is the framework in which testing issues must be resolved, and Bakke, with all its ambiguities, is now part of the landscape.

It should be obvious that the results of testing may be positive as well as negative. With all their imperfections, standardized tests remain a valuable source of information and a useful educational instrument. Much of what we know about racial, ethnic, and sexual differences in educational achievement is the result of standardized testing. Other sources of information -- teacher ratings, interview impressions, direct observation, etc. -- frequently lack the precision, accuracy, and credibility sought in standardized testing. There is

little doubt that testing has been a valuable part of educational efforts to discover and develop talent -- and testing has been the basis of a psychology of individual differences that has sustained public education throughout much of the 20th century.

But abuse of standardized testing has been widespread. In many respects those who construct and develop tests cannot ensure their professional and beneficial use. Technical competency in test research and development has far outpaced professional and socially responsible uses and applications. Users and consumers have not been "mature and wise" always in their administration, analysis, and interpretation of tests. In brief, the typical classroom teacher now receives little, if any, instruction in the uses and applications of tests and measurements. The typical school or college is without an admirable degree of sophistication concerning test technology and educational uses.

The continuing controversies that surround the uses of standardized tests, an era of litigation, and the emerging dictates of public policy imply that traditional test theory and practices are no longer adequate. A host of critics and adversaries of conventional uses, advocates of new rationales and purposes, and a professional literature with bursting seams suggests that: (1) alternative uses and applications of testing must clearly be defined and established, and (2) complementary methods and approaches must be devised.

The continued use of standardized tests in education is assured by the necessities of the foreseeable future, but the re-structured use of such tests would seem clearly mandated. At the same time there is an obvious demand for other approaches and procedures that serve the same educational and social purposes.

Alternative Uses of Tests

The controversies in standardized testing are their uses for educational and employment purposes -- and the social consequences that follow their use. Test construction and development have been dominated by a search for a high degree of reliability or stability for individual scores and an acceptable degree of precision or accuracy in the prediction of future performance. This concern for reliability theory and predictive validity meant that the uses of such tests were inherently selective -- and for racial or ethnic minority groups the exclusionary implications of selection have been their characteristic feature. Because tests were presumably objective, applicants were too frequently, and too easily, told that they had "failed the test."

New Rationales for Testing

Pluralism and diversity in education have convinced many observers that a different philosophy or theory of testing is sorely needed. Traditional concepts and principles may have served well the purposes of education in an earlier era, but

they provide a disservice in schools and colleges that must adjust to a pluralistic clientele. A "sorting and channeling" function that tests have served in the past is now inappropriate because education is a social benefit to be given the broadest possible distribution and is increasingly perceived as lifelong experience rather than a preparatory activity. Changing conceptualizations of education and schooling dictate changing purposes and uses of testing (Tyler, 1977).

Alternative uses are sought then in ways that would shift the emphasis in testing: (1) from measurement to assessment, (2) from people to programs, and (3) from aptitudes and general intelligence to skills and competencies. Each attempt should break testing out of its narrowly conceived role in a predictive, selective sense and broaden or extend its uses for the explicit purpose of facilitating teaching and learning in an educational setting. Selective admissions will remain a necessity in many professional and graduate programs where educational and social costs prohibit an open-door practice or an extended try-out. It will also remain "a fact of life" in courses or programs where the demand greatly exceeds facilities, resources, and instructional capabilities. But traditional selective admission practices are not needed in the same way they were in a period of limited opportunity. For many institutions the prediction of future performance may well be as effective as it ought to be (Fincher, 1974). Testing is now needed for other purposes such as placement, exemption, credit and evaluation.

From Measurement to Assessment

The changing emphasis from measurement to assessment reflects the educational need to appraise those changes in behavior and performance that presumably are the result of learning and development. With increased recognition that students can and do learn a great deal without the benefits of formal instruction and with increasing insistence that "extracurricular learning" be acknowledged and perhaps certified, there is a need for assessment procedures that do not aspire to the precision and accuracy traditionally desired in measurement. Instead of determining "how much" learning has taken place, there is more often a need to determine <u>if</u> learning has taken place. Differences of degree, ordering, and ranking are not needed as much as an acceptable way of ascertaining if learners have mastered, acquired, or accomplished some explicit standard of performance. Assessment is relevant, therefore, not only to basic literary skills such as reading and writing but also to program objectives and standards.

The concern for assessment is also a reflection of the returning desire for absolute or positive standards as opposed to the relative standards that are implied in traditional concepts of testing. Standardized tests of educational achievement have usually involved a comparison of the individual's performance with "norms" established on some identifiable group. Criterion-referenced tests are an

attempt to specify performance in educational or learning terms and to decide rather directly if individuals have reached or obtained some specified level of performance. The National Assessment of Educational Progress (Johnson, 1975) is the most extensive application of assessment concepts and procedures to date.

From People to Programs

A broad concern for program evaluation suggests that assessment techniques are needed for programs and projects as well as for student achievement. Accountability would seem to be a particularly strong insistence that the worth of programs be demonstrated by methods other than the status of individuals as the conclusion of formal education. Both concerns imply an obvious need to evaluate input, process, and output variables that characterize educational programs. It is the program rather than its participants who are thus evaluated.

Testing concepts and methods are needed then to help appraise instructional methods and materials, program structure and content, and overall program effectiveness in bringing about desired outcomes and results. This need lies close to the heart of the controversy that has accompanied teacher evaluation and statewide assessment in the public schools. Proponents of program assessment believe it should be the program that is evaluated and neither the teachers nor the

students as individuals. Performance review for teachers and learning achievements for students are objectives that must be realized under other conditions and rules-of-the-game. Both may be a part of program assessment but only in ways that clearly protect the professional integrity of the teacher and the learning needs of the student. Program assessment calls for instruments, procedures, and techniques that are different.

From Aptitudes to Skills

Concurrent with the shift from measurement to assessment and from people to programs, the purpose and intent of most educational testing needs to be changed from a concern with potential to a better judgment of actual accomplishments. Aptitude, readiness, and general intelligence tests have been the subject of most litigation involving standardized tests, and without exception, it has been difficult to satisfy the court's requirement of a logical, rational relation to explicit educational objectives. By and large, the courts have shown little infatuation with predictive validity and usually skirt its merits, arriving at their decisions on other grounds. Federal agency guidelines, on the other hand, have pegged the use of tests in employment quite tightly to predictive validity because of the vulnerability personnel tests have in that respect.

Educational policy would seem to have dictated long ago that achievement should be the object of testing, but there

has been a consistent failure to tie testing concepts and practices firmly to instructional objectives. The most widely used test of academic potential, the Scholastic Aptitude Test (SAT), has been both useful and effective despite the fact that it is a general test of verbal and mathematical ability and not specifically based on courses or programs taught in secondary schools. In short, the test presumably measures abilities that are learned or acquired but not taught. Only recently has there been any genuine acknowledgement that it could be taught.

Concepts of test fairness have increasingly implied that learners should be tested for what is taught. The logical, rational uses of tests in education would seem to be served only when the basic assumption of "equal opportunity to learn" can be met. Although generalized abilities may predict future performance or behavior, the courts have clearly preferred concepts of achievement or accomplishment that could be seen as a logical expectation from the training or instruction previously received. The push for competency-based education is a function of the growing belief that tests ought to measure skills, competencies, or abilities that are the direct outcome of learning and development. If testing is concerned with performance or mastery and related to instructional objectives or learning opportunities, there is less likelihood of a successful challenge in a court of law. By the same token, there is more likelihood that the testing serves a useful, educational purpose.

The Constructive Uses of Tests

The new rationales for testing are based on the premise that there remains positive, constructive benefits to be gained from systematic, educationally sound testing in the nation's schools and colleges. No useful purpose would be served by a moratorium on testing as such -- or by continued controversies that are spurious or ill-adaptive to the solution of educational problems or issues (See Houts, 1977 as one example of the needless literature of controversy).

Positive benefits are to be gained from testing if the focal concern can be shifted once and for all from selection per se to placement, advisement, counseling, and program assessment. Despite incessant advice that test scores should be but one source of information in selection decisions, the dominance of predictive validity has assured that the use of standardized tests in selective admissions will continue to be controversial because test data may be more specific and concrete than other forms of information usually are. Any use of test information in selective admissions should be an explicitly weighted component of the decision process, however, and the test's particular contribution or use should be known. The Georgia Power Case is explicit in its ruling that tests must be validated in terms of their actual use in the decisions affecting employees. It is not sufficient for a test to have predictive validity if such is not known at the time decisions and judgments affecting the applicant are made.

The gist of new rationales in testing should be that the value of any testing or assessment procedure should be judged primarily by the extent to which it aids learning or improves instruction. Testing theory and practices should be firmly established within a context of educational policy and practices. Where tests do not serve educational purposes and functions, their continued use in educational institutions should no longer be justified. Specific uses served by testing include:

1. The assessment of basic or fundamental skills of literacy. There is an urgent need in education to establish the minimal competencies that can be expected in elementary and secondary education. "A common and persuasive thread" running throughout a series of seminars conducted by the College Board was the belief that "while some reform and more creativity were needed in the admissions process, the real obstacles to equal educational opportunity were rooted far down in the grades, in the elementary and secondary schools (p. 31)." This consensus is a plaintive echo of a conclusion reached much earlier by the Carnegie Commission on Higher Education (1970) in recommending that "the first priority in the nation's commitment to equal educational opportunity" should be "increased effectiveness of preelementary, elementary, and secondary education programs."

2. The assessment of academic competencies in general education. A significant feature of the decline in liberal or general education is the absence of consensus about the competencies that should be evident in individuals who are liberally educated. The Carnegie Foundation (1977) has recently written of "advanced learning skills" that should be acquired by college students. "The competent college student" has been described by four seasoned scholars for the Tennessee Higher Education Commission (1977). Both publications are explicit in their disappointments concerning the decline of general education and the unflattering intellectual skills evidenced by many college graduates.

3. The assessment of academic programs at the college level. Declining test scores and grade inflation are part of a larger picture suggesting that a college degree no longer has the mettle it should have. This has produced, in return, a demand for program assessment at the baccalaureate level. Departments of academic instruction are challenged to demonstrate the value or worth of the programs they provide students. Greatly needed are tests and assessment techniques that will permit a better understanding of what students have learned in college

and how well academic programs have enabled them to realize their learning needs and interests. (See Academic Degree Program Assessment, 1977).

4. The diagnosis of learning difficulties. Although an avowed purpose of testing from the beginning, most standardized testing in recent years has failed to provide useful information about the actual difficulties students may have in learning what schools and colleges are trying to teach. Compensatory or remedial programs have failed because diagnostic resources and capabilities simply were not adequate. Instructors in developmental or special studies greatly need diagnostic information about student learning difficulties that will permit them to plan and organize remedial efforts. Teachers cannot pinpoint student errors, deficiences, and weaknesses because they so often lack systematic, functional methods and materials. The efforts of national testing agencies to develop more applicable instruments for diagnosis is indicative of a "growing market" for tests more closely related to learning difficulties.
5. The assessment of learning outcomes. In addition to testing that will help diagnose learning difficulties there is a need for testing that will provide useful information about student progress

under conditions of conventional instruction. Systematic testing is needed at frequent stages of education to inform both students and teachers about the accomplishment of educational objectives. Statewide assessment programs now provide better information about educational progress than they have in the past, and systemwide testing in higher education is beginning to establish levels of performance that must be met at entry, mid-program, and exit points. There remains, however, too much confusion about testing or assessment objectives, how they relate to educational objectives, and the specific uses and applications that will be made with testing or assessment results. Greatly needed are assessment procedures that will tell students and their parents how well the students are progressing in school, identify important educational objectives they are not meeting, and suggest ways students can get back into full scholastic stride.

6. The improvement of instruction. A frequent criticism of statewide assessment is that test results do not give the classroom teacher information of a sufficiently detailed nature at a time the teacher can actually use that information to help students. The instructional benefits of assessment are thereby

often lacking, and the actual value of the assessment effort is not what it should be to teachers. Much needed are assessment concepts, principles, and procedures that can be accommodated within a framework of instructional policies and practices. In brief, there is a need for testing that will tell instructors how well their particular teaching efforts are reaching students. While teacher-made tests, exams, and quizzes presumably provide this information, the testing and examining skills of most teachers and college instructors remain highly suspect (Milton & Edgerly, 1977). Over 25 years ago, Ralph Tyler (1951) suggested that testing should help teachers: (1) identify educational objectives, (2) select content, learning experiences, and procedures of instruction, and (3) coordinate instructional efforts. Testing and teaching should, once and for all, be brought within a workable framework that serves educational purposes.

7. The facilitation of learning. Although overlapping to some extent other testing needs, the facilitation of learning is important enough to bear repetition. Irrespective of other testing purposes and intents, assessment efforts should never lose sight of learning. Little has happened to alter the commendable viewpoint (Cook, 1951) that testing ought to:

(1) help adapt the curriculum to individual aptitudes and abilities, (2) help adapt instruction to specific accomplishments and deficiencies, and (3) provide motivation for better learning. In short, instruction should serve learning -- and testing should serve both. More recent research into the interactions of learning behavior and teaching efforts suggests that the situation is far more complex than once imagined but not hopeless (Cronbach & Snow, 1977).

Complementary Methods and Approaches

It is unfortunate that Regents v. Bakke has focused too narrowly on the admissions function in education. And it is unfortunate that there is more to read between the lines than within. Be that as it may, there are implications in Bakke that are being carefully considered by those charged with admission responsibilities. If Bakke is a confrontation between individual and group rights, the ruling would seem to indicate that the individual has rights that may not be denied by group considerations. If race is a relevant consideration for the individual but not for groups, it would seem to be an experiential factor that has direct relevance for the individual's aspirations and his or her expectations for success. And if the diversification of admissions criteria is now required by public policy, there must be full recognition that: (1) single standards will no longer suffice, (2) there must be alternative routes to commonly desired destinations, and

(3) much of the decision data will inevitably be subjective, intuitive, or "soft" but must be intelligently applied.

Justice Powell writes of "exceptional personal talents, unique work or service experience, leadership potential, maturity, demonstrated compassion, a history of overcoming disadvantage, ability to communicate with the poor, or other qualifications deemed important." His list would appear to include many qualities that admission procedures have often claimed to consider -- qualities or characteristics that standardized tests are often criticized for not tapping or adequately measuring. But more importantly perhaps, his ruling states that the courts do not compel institutions to use only "the highest objective academic credentials" for admission to their educational programs. If fully granted and if the "achievement of diversity" for educational purposes is to be an admissions function, institutions of higher education must prepare to deal with human qualities that are "complementary" to those that can be handled within traditional testing contexts.

The development of complementary methods and approaches in admissions should be predicated on a "principle of complementarity" yet to be articulated. There should be explicit recognition and acceptance of the possibility that other ways of looking at applicants and their promise for further growth and development are not only desirable but necessary. Much of this is implicit in all that has transpired in the past 15 or 20 years, but it must be gleaned from vague, often ineffectual, efforts to define pluralism and diversity in a democratic

society that would yet claim to be meritocractic. Whatever pluralism and diversity may mean, it must surely acknowledge that academic standards and credentials can no longer be singular or uniform. "Cut-off scores" ought to be as obsolete as the dodo.

An occupational hazard for admissions officers and committees has always been the charge of "a double standard" and there will always be faculty members who won't understand why athletes and music majors "can't meet the same admission standards everyone else does." Efforts to establish pluralistic admission criteria are suspect and further litigation claiming "reverse discrimination" may or may not clarify the issue in the future. The point remains that public policy now influences significantly the admission policies of institutions of higher education but does not completely dominate. Colleges and universities still have appreciable "elbow room" in working through the conflicting demands of social and educational policy.

It would be most desirable if institutions of higher education could search for complementary methods and approaches without litigious harrassment and unrealistic expectations. If they are to succeed, they must be permitted to fail. Colleges and universities must remain free to experiment without paying "court costs and damages."

Equally important is the clear possibility that if complementary methods are sought, some of the decision rules

and bases for human judgment will be contradictory to conventional or traditional criteria. The possibility of contradiction is implicit in a principle of complementarity and would simply mean that some applicants were admitted for one purpose under one set of admission criteria and others were admitted for another. The two sets of admission criteria may or may not overlap and may be completely orthogonal when viewed from a third perspective. It will all smack of double talk, and listeners -- in court and out -- must be patient.

In summary, testing remains a potent source of aid and assistance for the improvement of education -- if properly conceived and rightly used. But not only should there be other uses and applications of testing in education, there should be a serious effort to develop complementary approaches and procedures serving some of the same purposes. Selective admissions will continue to be a fact of academic life and group conflict shows no signs of lessening in contemporary society. A part of the challenge is to keep the issues open long enough to work out an eventual resolution.

REFERENCES

Academic Degree Program Assessment in the University System of Georgia: A Tentative Statement of Rationale, Purposes, and Procedures. Atlanta: University System of Georgia, 1977.

Branscomb, Harvie, Milton, Ohmer, Richardson, John, & Spivey, Herman. The Competent College Student: An Essay Concerning the Objectives and Quality of Higher Education. Nashville: Tennessee Higher Education Commission, 1976.

Carnegie Commission on Higher Education. A Chance to Learn: An Action Agenda for Equal Opportunity in Higher Education. New York: McGraw-Hill, 1970.

Carnegie Foundation for the Advancement of Teaching. Missions of the College Curriculum: A Contemporary Review with Suggestions. San Francisco: Jossey-Bass, 1977.

Cook, Walter W. "The Functions of Measurement in the Facilitation of Learning." In E.F. Lindquist (Ed.) Educational Measurement. Washington, D.C.: American Council on Education, 1951.

Cronbach, Lee & Snow, Richard E. Aptitudes and Instructional Methods: A Handbook for Research on Interactions. New York: Irvington Publishers, 1977.

Fincher, Cameron. "Is the SAT Worth its Salt? An Evaluation of the Scholastic Aptitude Test in the University System of Georgia Over a Thirteen Year Period." Review of Educational Research, 1974, 44, 293-305.

---------- The Access-Placement-Retention-Graduation of Minority Students in Higher Education. A background paper prepared for a conference on "Equality of Access in Post-secondary Education" sponsored by the Southern Education Foundation and the Ford Foundation. Atlanta: SEF, 1975.

---------- "Standardized Tests, Group Differences, and Public Policy." The College Board Review (Spring 1977) No. 103, pp. 19-31.

Hills, John R. Proficiency Testing: Implications for Higher Education. Atlanta: Southern Regional Education Board, 1977.

Holloway, Charles M. The Bakke Decision: Retrospect and Prospect. Summary report on six seminars held by the College Board in July and August 1978. New York: College Board, 1978.

Houts, Paul L. (Ed.) The Myth of Measurability. New York: Hart, 1977.

Johnson, S.S. Update on Education. Denver: National Assessment of Education Progress, 1975.

McCormack, Wayne (ed.) The Bakke Decision: Implications for Higher Education Admissions. A report of the ACE-AALS Committee on Bakke. Washington, D.C.: American Council on Education and Association of American Law Schools, 1978.

Milton, Ohmer & Edgerly, John A. The Testing and Grading of Students. New Rochelle: Change Magazine Press, 1977.

Roueche, John & Roueche, Susan D. Developmental Education: A Primer for Program Development and Evaluation. Atlanta: Southern Regional Education Board, 1977.

Tyler, Ralph. "The Functions of Measurement in Improving Instruction.: In E.F. Lindquist (Ed.) Educational Measurement. Washington, D.C.: American Council on Education, 1951.

---------- "Fifty Years of Theory and Practice." Social Policy, 1977, 8(2), 11-17.

Legal References

Albemarle v. Moody. United States Law Week, 1975, 43, 4880-4895.

Griggs v. Duke Power. Supreme Court of the United States, 1971, No. 124.

Regents v. Bakke. United State Law Week, 1978, 46, 4896-4936.

U.S. v. Georgia Power. Federal Reporter, 2nd Series, 1973, 474, 906-927.

Washington v. Davis. United States Reports, 1975, 426, 229-270.