

Ability Testing: Uses, Consequences, and Controversies

Alexandra K. Wigdor
National Research Council

The Committee on Ability Testing was convened by the National Academy of Sciences early in 1978. Its central task was to examine the role of testing in American life. Because the charge to the Committee called for a study of testing from a social perspective, the members were particularly sensitive to the need to go beyond questions of technical adequacy, and to explore the implications of test use for individuals, minority groups, institutions, and American society as a whole. The project was sponsored by the Carnegie Corporation of New York, the Ittleson Foundation, the National Institute of Education, the National Institute of Mental Health, and the Office of Personnel Management.

This synopsis of the report is designed to convey a sense of the entire document, but will focus particularly on the Committee's findings on educational testing. Before proceeding, I should point

out that the report is not written primarily by and for psychometricians and testing specialists; it represents the consensus of a multidisciplinary group of social scientists and legal experts, a majority of whom have no connection with testing, and it is addressed to policymakers, test users, and test takers—people who know relatively little about the technology but are called on to make decisions about tests or using tests. Naturally, we hope the testing community will find it valuable as well.

The Controversy Over Testing

Advocates of testing have long argued that professionally developed and validated tests are the best available means of impartial assessment of abilities. Quantitative assessment of human performance has seemed well suited to the conditions of modern society because it offers rapid, relatively

inexpensive, streamlined methods of obtaining information about large numbers of people.

Testing also has been attractive to Americans because it seemed a more democratic method of selection than the traditional reliance on family connection, social class, dress, accent, or other social cues. It appealed to 20th-century American notions of fairness because it was more obviously related to performance than the traditional criteria and because it was cloaked with the respectability of science.

Testing has not been without its critics, however. The Committee sympathizes with the misgivings expressed by Walter Lippmann early in the century, particularly his objections to the testers' claim that their instruments were capable of measuring intelligence when no one could define intelligence. In recent years critics have begun to question whether the goals of identifying ability and enhancing productivity are as well served by testing as has been asserted, or whether these goals sufficiently define the social good. Indeed, hundreds of employment tests have been challenged in the courts. The Committee gave antitest arguments a careful hearing and agreed with many of the points made by testing critics.

This led us to include in the report a fairly lengthy discussion of the limitations of tests, limitations that all too often are not sufficiently appreciated by test users and test takers. Among the issues discussed are the limited predictive powers of tests (e.g., first-year grades, not later performance in a profession), the constraints on assessment introduced by the very process of standardization, the traditional concentration of tests on a limited range of cognitive skills which clearly do not



Alexandra (Sandy) Wigdor is Study Director of the Committee on Ability Testing of the Assembly of Behavioral and Social Sciences, National Research Council. Ms. Wigdor was staff historian of the National Study Commission on Records and Documents of Federal Officials, and taught history at Lincoln University (Missouri) and Northeast Missouri State University, and was a reader at the Folger Shakespeare Library for a year. Ms. Wigdor's interests include the legal dimensions of testing, development of government policy, and social issues associated with the rise of testing.

exhaust the complex elements of human cognition, and the inability of tests to assess adequately other important characteristics such as motivation, creativity, or perseverance.

One of the most important and least understood limitations of testing derives from the fundamentally quantitative and empirical nature of psychological testing. Psychometrics, or mental measurement as it used to be called, developed in the latter part of the 19th century as an alternative to the epistemological theorizing of earlier philosophers and theologians. Experimental psychologists rejected philosophy as a route to understanding human cognition in favor of exact measurement of performance from which they hoped to build an empirical basis for understanding mental traits.

From the beginning, psychometrics was based on the idea that people exhibit various mental traits in varying degrees, and this independently of hereditary social status. The new science of statistics could be used to demonstrate those individual differences and to predict any one individual's likely performance in relation to that of a given population of individuals. The theory of probability allowed the scientist to say with a known degree of confidence whether the differences in measured abilities of two individuals reflected real differences or only measurement error, and the notion of correlation was used to relate such measurements to expectations of performance.

One cannot overemphasize the centrality of the statistical underpinnings of psychological testing from that day to this. Testing is essentially a behaviorist enterprise, which seeks to link up test scores with external correlates. While this external approach to the study of abilities has important uses, it also has weaknesses as a means of explaining abilities.

Theories of cognition do not play a central part in test development and one could argue that we are not much closer to a definition of "ability" or "intelligence" than we were at the beginning of the century. One of the Committee's important recommendations to the testing community, therefore, is that it is time to devote more effort

to the question of what tests measure and to work more closely with cognitive psychologists.

Despite these acknowledged limitations, the Committee does not agree with those critics who would do away with testing. It is the Committee's position, based on a thorough review of the available evidence, that ability tests *can* be useful predictors of subsequent performance. (The qualifier is important because that usefulness depends on the test being carefully developed, on its being validated or otherwise shown to be appropriate to the circumstances of actual use, and on the user knowing enough about testing and scaling to interpret the scores.)

Fairness Questions

The Committee further concluded that ability tests predict equally well for *all* groups of test takers. Research evidence does not support the notion that tests systematically underpredict the performance of minority group members. There are certain important exceptions to this generalization. One cannot give a Spanish-speaking child an English-language test and expect the scores to mean the same as those of a native English speaker. But the research evidence does not support the frequent contention that tests are unfair because they underpredict the performance of certain subpopulations. In this sense, tests are not biased and can be called "fair."

I would like to emphasize that the report makes a strong distinction between the *characteristics* of tests and the *social impact* of tests. The widespread use of tests, no matter how scientifically valid, may have effects that this society finds unacceptable. When the focus is on social impact, different fairness questions arise. The rather large differences in average performance on most ability tests between blacks and whites, for example, will tend, when test scores dominate a selection process, to screen blacks out of jobs and into special education programs. Because of the exclusionary effects of tests, they are no longer simply the concern of the employer or a matter of school policy. Since the passage of the Civil Rights Act

of 1964, tests have become an important focus of civil rights enforcement efforts.

This politicization of selection issues has had both good and bad effects on testing. At its hearings in November 1978, the Committee was presented with a good deal of evidence that employers have tightened up their testing practices and have devoted more resources to personnel selection in response to the promulgation of federal guidelines on employee selection procedures. In addition, the enforcement activities of the Equal Employment Opportunity Commission and the courts have done away with many employment testing programs that were making no verifiable contribution to productivity.

Nevertheless, ability testing can advance the genuine societal interest in promoting an efficient and productive work force, a goal that few would argue with in these bleak times. The Committee found that the interpretation of legally acceptable testing practices promoted by EEOC, and largely accepted by the courts, has made it exceedingly difficult for employers to defend even state-of-the-art selection procedures. We have reached the point where relatively good tests are being thrown out along with the bad.

There is, in the Committee's view, no psychometric solution to the problem of adverse impact. The solution must come from policy, not science. But it is important that the two complement one another. The central recommendation of the report regarding employment testing is that the validity of the testing process should not be compromised in the effort to shape the distribution of the work force. Those who administer the laws should not require tests to do things they cannot do, such as guarantee that score distributions will not differ for different racial or ethnic groups. Instead, we urge government authorities to provide employers with a range of legally defensible selection rules that balance the competing goals of productivity and equal opportunity. Where tests are of demonstrated utility their use should be encouraged, but only as one element in a selection formula that increases minority participation in the work force.

Educational Testing

Some of the most vociferous debate over testing has involved admission to college and professional schools. Ironically, one of our central findings is that most undergraduate institutions are not so selective that test results are crucial to the selection decision. Most students will be admitted to the college or university of their choice. Test scores are likely to be a barrier only to the few applicants who are marginal and, at the other extreme, the few applicants who want to attend the most prestigious institutions. Consequently, the Committee recommended that undergraduate institutions that now require applicants to take a test for admission (virtually all 4-year schools) reexamine the wisdom of that requirement.

By contrast, there is tremendous pressure for places in the professional schools, particularly law and medical schools. It is not unusual for these institutions to have several thousand applicants for the 100 or so available places. Because of this crush of applicants, a majority of whom would be capable of fulfilling the requirements of the program of study, there has been a tendency for admissions officials to exaggerate the importance of small score differences. The Law School Admission Council has decided to replace the 200-800 scale on the LSAT with a 10-50 scale for its new test to discourage decisionmakers from attributing a false precision to the numbers.

This is a step in the right direction. The report suggests other steps for improving the use of admissions tests in selective institutions. The first is a call for local validation efforts. Few colleges and universities perform such studies, so officials cannot make fully informed interpretations of the meaning of test scores at their institutions. The Committee recommended that where tests play an important part in selection, admissions officials should consider participating in a cooperative validity study of the kind offered by the Graduate Record Examinations Board.

The Committee further recommended that the selection process encourage diversity in the student populations. Overemphasis on test

scores will tend to exclude members of racial and ethnic minorities. But for the present, the basic selection problem for medical and law schools is an overabundance of qualified candidates. When there are many applicants capable of succeeding, admissions decisions should be based on social and educational values that are broader than a comparison of predicted grade averages.

Truth in Testing

On the issue of test disclosure, the Committee viewed the truth-in-testing movement as an important assertion of the interest of students, and of society in general, in the allocation of educational opportunity. In addition to our belief that it is generally desirable that the entire college selection process be open and known, the Committee feels that there are specific reasons for encouraging test developers to make more information available about their testing programs than has traditionally been the case. One of the biggest sources of problems with testing in America stems from the misunderstandings and misinterpretations of test results by test users, test takers, and the public in general. Insofar as disclosure leads to more informed use of test scores, the truth-in-testing movement will be of benefit. This might well be the case if decisionmakers and students are given sufficient information about the nature of a particular test and what it purports to measure, the scaling methods used to give scores meaning, and the limitations of the test.

Despite its support for openness in admissions testing, the report stops short of endorsing disclosure legislation such as that proposed by Representative Ted Weiss of New York on the federal level or the Act passed in New York a couple of seasons ago. We were not convinced that full disclosure, that is, disclosure of the actual test questions and answers, would have the effect of significantly improving test quality, as supporters have claimed. Nor did we think their expectation realistic that full disclosure would make the admissions process fairer by improving the competitive posi-

tion of minority and disadvantaged applicants. It is perhaps more likely that privileged students will reap the greater advantage.

At the same time, we were equally unconvinced that full disclosure would cause serious problems with regard to test validity and reliability in large-scale testing programs. The experiments of the last few years, including the passage of the New York and California laws and the decision of several important admissions councils to disclose their tests nationwide, have not brought the immediate and total collapse of standardized admissions testing programs as some predicted, and, indeed, the industry seems to have been resourceful in developing new equating techniques. Of course, the long-term effects for good or for ill remain obscure at this point.

Finally, the Committee was not convinced that federal regulation of the testing industry is the best way to promote openness. We did not see evidence of systematic abuse by the testing companies of a kind that would warrant government intervention. Nor did we find it self-evident that the state laws currently in operation will significantly improve tests or the way they are used.

The lack of clear-cut answers about the effects of disclosure led the Committee to propose as its major recommendation that instead of speculating about the technical costs and consumer benefits of disclosure, the states, Congress, admissions officials, and the research community monitor the developments of the next few years so that all parties can make an informed judgment about creating a workable balance between openness in testing and objective assessment of academic ability. To this end, the report recommends that the Department of Education support empirical investigations of the effects of various disclosure plans.

Testing in the Schools

The selection situations described thus far are of general social importance partly because they affect the life chances of the individuals involved. When the Committee turned to the subject of testing in the primary and secondary grades,

Continued on page 26

MESSICK

Continued from page 20

- philosophy meet. *Journal of Educational Measurement*, 1976, 13, 31-41.
- Cronbach, L.J., & Gleser, G.C. *Psychological tests and personnel decisions* (2nd ed.). Urbana: University of Illinois Press, 1965.
- Darlington, R.B. Another look at "culture fairness." *Journal of Educational Measurement*, 1971, 8, 71-82.
- Deutsch, M. Equity, equality and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, 1975, 31(3), 137-149.
- Dunnette, M.D., & Borman, W.C. Personnel selection and classification systems. In M.R. Rosenzweig & L.W. Porter (Eds.), *Annual Review of Psychology* (Vol. 30). Palo Alto, Calif.: Annual Reviews, 1979.
- Einhorn, H.J., & Bass, A.R. Methodological considerations relevant to discrimination in employment testing. *Psychological Bulletin*, 1971, 75, 261-269.
- Ghiselli, E.E. *The validity of occupational aptitude tests*. New York: Wiley, 1966.
- Ghiselli, E.E. The validity of aptitude tests in personnel selection. *Personnel Psychology*, 1973, 26, 461-477.
- Gross, A.L., & Su, W. Defining a "fair" or "unbiased" selection model: A question of utilities. *Journal of Applied Psychology*, 1975, 60, 345-351.
- Kelly, R.M. Ideology, effectiveness, and public sector productivity: With illustrations from the field of higher education. *Journal of Social Issues*, 1980, 36(4), 76-95.
- Linn, R.L. Fair test use in selection. *Review of Educational Research*, 1973, 43, 139-161.
- Linn, R.L. In search of fair selection procedures. *Journal of Educational Measurement*, 1976, 13, 53-58.
- Messick, S. Test validity and the ethics of assessment. *American Psychologist*, 1980, 35, 1012-1027.
- Messick, S. Evidence and ethics in the evaluation of tests. *Educational Researcher*, 1981, 10(9), 9-20.
- Nozick, R. *Anarchy, state, and utopia*. New York: Basic Books, 1974.
- Okun, A. *Equality and efficiency: The big tradeoff*. Washington, D.C.: The Brookings Institution, 1975.
- Peterson, N.S., & Novick, M.R. An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, 1976, 13, 3-29.
- Rawls, J. *A theory of justice*. Cambridge, Mass.: Harvard University Press, 1971.
- Sawyer, R.L., Cole, N.S., & Cole, J.W.L. Utilities and the issue of fairness in a decision theoretic model for selection. *Journal of Educational Measurement*, 1976, 13, 59-76.
- Thorndike, R.L. Concepts of culture-fairness. *Journal of Educational Measurement*, 1971, 8, 63-70.
- Wigdor, A.K., & Garner, W.R. (Eds.). *Ability testing: Uses, consequences, and controversies*. Washington, D.C.: National Academy Press, 1982. ■

WIGDOR

Continued from page 8

this consideration became the controlling one. All the recommendations involving testing in the schools are based on the premise that tests should not be used if they have destructive effects on the life chances of children.

For example, we found that test scores play a central, often a determinative, role in special education placement. When used appropriately and in combination with the many other sources of information about a child's performance that are available to school teachers, tests can help to identify pupils who are unable to prosper in the regular program, perhaps because of learning disabilities or mild retardation. The Committee concluded, however, that there is little justification for removing such children from regular classes if there is no reasonable expectation that special instruction will provide them with more effective education. Evidence presented in several court cases raises the strong presumption that special education classes are sometimes holding pens where no real education occurs.

The minimum competency testing movement provides another

example. Our central recommendation is that minimum competency programs involve instruction as well as assessment. We see little point in devoting educational resources to assessing students' competencies if the information gained is not used to improve substandard performance. Moreover, we feel that the schools should carry the burden of demonstrating that the instruction offered has a positive effect on test performance. Diagnosis without treatment does no good, and quite literally, adds insult to injury.

A second theme that I would like to highlight is this society's tendency to place unwarranted faith in numbers. Quantification encourages the dangerous illusion that what cannot be reduced to a number can be left on the periphery of the decision process. Expressing test performance as a numerical score can often short-circuit the information-gathering process that testing is supposed to aid. The Committee is especially concerned to warn against the practice that apparently exists in several jurisdictions of using rigid numerical cutoff scores (e.g., IQ 80) as a definition of mental retardation. This sort of overreliance on test

scores is a widespread problem which in the past has led to destructive classification decisions such as placing children with a limited command of the English language in classes for the retarded on the basis of low test scores.

In concluding this summary of the report on ability testing, I would like to emphasize a theme that runs throughout the report, and that is a call for balance. Tests can be useful sources of information in educational settings and in the workplace; but they are limited instruments, and do not tell everything of importance about any individual. We hope that the report will help to counteract the widespread tendency to look to ability tests as a panacea for deep-seated social ills, and we hope that it will counter the equally prevalent tendency to use tests as a scapegoat for society's ills.

Copies of the report may be purchased from the National Academy Press, 2101 Constitution Ave., N.W., Washington, D.C. 20418. *Ability Testing: Uses, Consequences, and Controversies*. Alexandra K. Wigdor and Wendell R. Garner, eds. Two Volumes. Part I: Report of the Committee (\$13.95). Part II: Documentation Section (\$24.95). ■