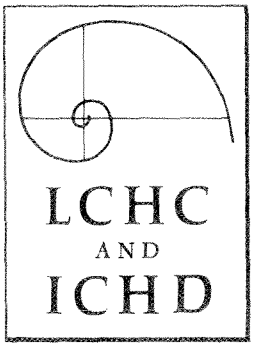
# INTELLIGENCE TESTS:
# A COMPARATIVE PERSPECTIVE

SYLVIA SCRIBNER
The Rockefeller University




The Rockefeller University
Laboratory of Comparative Human Cognition
AND
Institute for Comparative Human Development

WORKING PAPER NO. 1

THE ROCKEFELLER UNIVERSITY

LABORATORY OF COMPARATIVE HUMAN COGNITION

AND

THE INSTITUTE FOR COMPARATIVE HUMAN DEVELOPMENT

Working Paper
No. 1

# Intelligence Tests: A Comparative Perspective

Sylvia Scribner
The Rockefeller University

technical apparatus, and its consequences.

The comparison is directed to the following questions: What is the nature of testing in the two domains? How do test outcomes affect the individual's life? What impact do they have within the institutions? What are their broader social consequences?

Testing in the Medical Domain

An initial observation about medical tests is that they are primarily special-purpose: they are administered on particular occasions for particular diagnostic purposes. There is no general physical ability test analagous to the general mental ability test. In recent years, with increasing emphasis on health maintenance needs, new instruments have been developed to assess an individual's overall state of health. Interestingly, these "health inventory" assessments are designed to secure as detailed a picture as possible of the functioning of different physical systems that can be used both as the basis for individual care and for projecting the health-care needs of the population as a whole. The health inventory does not yield a single summary figure purporting to measure an individual's health capacity--a health quotient, as it were. That there are no H.Q.'s comparable to I.Q.'s is less likely to reflect technical incapabilities than the fact that the concept has no perceived utility.

Of the three standard functions of tests--prediction (prognosis), diagnosis, and selection--the leading function in medical testing is diagnosis; prognosis and selection are

ever-closer relationship between the two, such that available treatments provide guidelines for diagnosis and diagnosis informs treatment.

In addition, prediction is not practiced for prediction's sake, outside of a research context. There is no mass testing of individuals for prognostic purposes only. Suppose a test were developed that permitted almost perfect prediction of glaucoma from eye color. What would the response be to a proposal for mass testing, the only outcome of which would be the labeling of individuals as high or low glaucoma risks? Such mass testing, independent of a program for either treatment or prevention, would not win easy acceptance among the public at large or in the health professions.

In the individual case, a diagnostic work-up providing information about the patient's condition will also yield information about chances of benefiting from available methods of treatment. In most cases, however, the patient's prognosis does not determine her admission to health care; the patient with a poor prognosis is not turned away. We said "in most cases," for the restrictions on this rule are important indicators of the social conditions under which prediction does or does not assume a leading role in testing. "Poor prognosis" may have no-treatment consequences if treatment facilities are limited and must be allocated rationally. Too few hospital beds, insufficient supplies of medicine, a doctor shortage -- these are the conditions in which some social decision-making process comes into play to determine the beneficiaries of the

such as "mentally retarded" or "mentally ill" are commonly affixed to the patient and may result in the adverse side-consequences of stigmatization and loss of opportunities for personal achievement.

We can recall that, in the not too distant past, stigmatization was also attached to certain physical illnesses. There were "dirty" diseases that "dirty" people contracted: gout was a high-status symbol, tuberculosis the unmentionable affliction of the working girl. We like to believe it a sign of progress that evaluative practices of this kind are diminishing, but even if they persist in certain areas with respect to certain conditions, such practices do not rest on alleged scientific support. Because medical testing is special-purpose and oriented to individual diagnosis, the evaluative group-comparison aspect is minimized.

Finally, we would like to make an observation about the relationship among medical tests, treatment, and outcome. Regardless of how the individual has been tested, the medical institution and its practitioners stand accountable for outcome of treatment. Treatments that are ineffective reflect on insufficiencies in our knowledge or inadequacies of our practice. There is no institutionalized policy of blaming the patient for failure to recover. For several years we have been in the midst of a national debate as to the best means of improving the delivery of national health care to many sections of the population whose needs are now inadequately served -- rural communities, urban poor, elderly, and minority

most concern in the schools are the so-called general intelligence tests and tests of the narrower aptitudes that make up "intelligence." On the basis that ability tests have the most widespread use and have been subject to the most "widespread misuse and misinterpretation" (p. 17), the report restricts its discussion to such tests. We will follow a similar course.

Mental ability tests are "general purpose" instruments that lack the specificity of medical tests. Such nonspecificity is consonant with their main functions, which historically have been those of prediction and selection, rather than of diagnosis. Cronbach's (1960) comparison between the objectives Binet outlined for testing and those of Terman and other early testers indicates that, from the outset, American instruments were intended to serve the function of mass evaluation rather than individual diagnosis geared to treatment. Binet was concerned with the description and explanation of individual differences, and he saw this task as involving two types of diagnostic study:

> 1. The study of how psychic processes vary from individual to individual, what the variable properties of these processes are, and to what extent they vary.
>
> 2. The study of relations among the different psychic processes in a single individual...." (1895, quoted in Herrnstein and Boring, 1965, p. 429)

Binet believed that these studies could be carried out by mental tests, but he explicitly rejected the notion that a single mental test could be devised that would shed light on

and de-emphasize the total summary score in favor of interpreting scores on the test subscales (Rapaport, 1968). Moreover, it stands to reason that tests designed to be diagnostic could more usefully serve this purpose. Diagnosis is a peripheral practice, however, within the school system, where standardized group testing holds sway.

The very success of IQ tests in predicting childrens' careers through school has served to freeze the tests in essentially their original form. Cronbach (1960) and others point out that, although the psychometric properties of IQ tests have improved and models of prediction have become increasingly sophisticated, test instruments of today are not very different from those of the '20's and their original prototype, the World War I Army Alpha test. A half-century of continuous psychological investigation of learning processes, memory, and reasoning skills have not eventuated in the development of tests yielding information of the kind Binet sought. The scientific technical enterprise of testing has become increasingly divorced from the diagnostic and treatment enterprise.

Many factors are undoubtedly implicated in the failure to develop diagnostic tests with prescriptive value in education. For one thing, the task of diagnosis has turned out to be infinitely more complicated than Binet envisioned. But the difficulty of the task strikes us as a lesser factor than the strong social functions which the tests fulfill in their present form.

valuable to transmit was converted into the practice of preparing children for their "station in life." IQ tests predicted what level the children could reach, and the end-point was built into the differentiated school curriculum (see Tyack, 1974, for a detailed historical account of this development).

Terman, for example, argued that occupations could be ranked by the level of intelligence they required and that children could be matched to occupations by IQ tests:

> Intelligence tests can tell us whether a child's native brightness corresponds more nearly to the median of (1) the professional classes, (2) those in the semi-professional pursuits, (3) ordinary skilled workers, (4) semi-skilled workers, or (5) unskilled laborers. This information will be of great value in planning the education of a particular child and also in planning the differentiated curriculum here recommended. (Terman, 1920, p. 23)

Terman advocated a three-track system, supplemented by a special track for the gifted and a special track for the intellectually disadvantaged, all of which corresponded to the five classes of occupations into which he saw the child population channeled. For some, that long-range selection function of the public school system was of greater utility than its traditional function of education. This position was made explicit by Professor Pillsbury (1920), who argued against the accepted wisdom that "education is essentially a process of creating intelligence" (p. 62). Schools, he thought, mainly winnow out the unintelligent and select the capable. Pillsbury foresaw the possibility that improved

is any demonstration of significant relations among diagnosis, treatment, and outcome. Indeed, individualized treatment, the function that both legitimizes and makes necessary the development and use of medical tests, remains an external consideration in the educational testing movement. In historical perspective, the interaction between testing and treatment in the educational domain has taken a form quite different from that in medicine. To a considerable extent, test results have been used as the basis for modifying treatment goals (varying the content and end point of education for different groups of children), rather than modifying treatment practice (e.g., adapting teaching methods of a single curriculum to specific skills or disabilities of individual children). As McClelland (1973) has persuasively argued, such present-day practices as the system of tracking within elementary schools and selective access to higher levels of schooling on the basis of IQ and aptitude tests have had the result that such tests both predict treatment outcome and help to produce it. The prevailing paradigm of IQ tests as "independent predictors" of an educational outcome criterion can be maintained only if one ignores the instrumental role of test scores in determining educational input.

Moreover, the effects of test scores on educational input may be more than a matter of track or group placement. Test scores can also affect educational input by helping to define teachers' expectations of individual children, and thus subtly influencing the quality of interaction between teacher-and-student which itself constitutes an important aspect of the

the testing movement as one attempting to justify existing societal inequities by giving them the force of natural law: "The alleged 'mental levels' representing natural ability, it will be seen, correspond in a most startling way to the social levels of the groups named. It is as though the relative social positions of each group are determined by an irresistable natural law" (Chicago Federation of Labor, in Counts, 1928, p. 107). It was clear to the Federation that IQ tests merely reflected different life conditions determined by the unequal distribution of wealth and social status in industrial America, and it vociferously protested the use of the schools to perpetuate this order.

The argument that occupational social stratification is the consequence of innate differences in intelligence has been vigorously revived in the last decade (Herrnstein, 1971). According to some analysts (see, for example, Bowles and Gintis, 1973), the role of the IQ test in supporting this ideological position provides the principal social justification for its continued use. To the extent that mental testing serves as rationale and support for unequal opportunities based on color, class and ethnicity, in a profound sense, it has serious implications for society as a whole.

Rapaport, D., Gill, M., and Schafer, R. 1968. Diagnostic Psychological Testing. New York: International Universities Press.

Terman, L.M. 1920. The use of intelligence tests in the grading of school children. Journal of Educational Research, 1: 20-32.

Tyack, David. 1974. The One Best System: A History of American Urban Education. Cambridge, Massachusetts: Harvard University Press.