

# Testing Testing

by Kenneth Kaye

### Bias in Mental Testing

by Arthur R. Jensen

Free Press, 1980, \$29.95

"If you want to make it in the white man's world," black parents used to tell their children, "you've got to be twice as smart and work twice as hard as a white child." That could be said today of Arthur Jensen. Anyone else would not have had to produce such a superb book as he has done in *Bias in Mental Testing*: He had to surpass all standards just to win the right to have it judged on its merits, without prejudice.

Jensen is the Berkeley psychologist who upset the Great Society applegart with his 1969 argument that Project Head Start's goals were unreasonable: that the differences in IQ and scholastic achievement between blacks and whites were inevitable because they were largely genetic. He has not changed his views. However, that is not the point of this book. Here he is concerned with defending standardized mental tests against the charge of bias, whether racial or cultural. Consistent group differences in test scores, he argues, are not the fault of the tests. Whether the lower scores of the average disadvantaged child, the average Mexican-American child, and the average black child are due to inferior nature or inferior nurture, he says, they are not due to the nature of test items, the sample on which the test was "normed," the testing situation, or the method of scoring.

It is important that Jensen has dropped the concept of heritability from his explicit argument, even though it may still be alive in the back of his mind. For Jensen's original position contained a fallacy. It treated estimates

---

*Kenneth Kaye is Associate Professor of Education and Human Development at the University of Chicago. His recent studies of communication between infants and adults in feeding, play, imitation and instruction are discussed in his forthcoming book, THE MENTAL AND SOCIAL LIFE OF BABIES, to be published by the University of Chicago Press.*

of IQ heritability (usually somewhere between .60 and .80, which means that most of the variance among individuals is due to differences in their genotypes rather than in their environments) as if these estimates were not mere estimates but inevitabilities. It neglected the fact that heritability itself is different in different environments; so if scholastic achievement is highly heritable in our society, this may simply demonstrate that our schools fail to serve as instruments of social mobility—not that they *must* fail.

An additional problem with Jensen's argument was that the strongest sources of evidence for the heritability of IQ came from the work of Sir Cyril Burt. Sir Cyril, now known to have fabricated some of his data (as well as fabricating a number of his co-authors and even opposing letters to which he replied in the journals), has been wisely dropped from Jensen's pantheon. Although there is still no reason to doubt that IQ scores are highly heritable, certainly this book would not have the credibility it has if either heritability or Burt played a significant role in it.

Most of Jensen's critics over the past decade, instead of attacking the irrelevance of the fact of heritability, tried to deny its existence. One way to do this was to discredit Burt, but the principal way was to discredit the tests. Jensen has no sympathy for those who have pulled out isolated test items and argued that they are patently biased. He has devoted his career to the statistical analysis and validation of tests, and he refuses to argue on any other basis than statistics. Consequently this book is less an argument than it is a thorough handbook on the statistical basis of test construction, validation and use in selecting individuals for schools or jobs. (It contains about 400 thoughtfully designed tables and illustrations, 96 technical notes, a glossary of 46 terms, and more than 800 references. Its main limitation as a handbook is its poor index.)

### *Tests Should Discriminate*

Jensen's essential point is the distinction between discrimination and bias. A test is supposed to discriminate—that is, to distinguish among individuals. If Group A gets higher scores on the average than Group B, the question of *bias* has to do with whether Group A really performs better on whatever criterion (in other words, future performance) the test is supposed to predict. A biased Scholastic Aptitude Test, for example, would be one that failed to predict college performance equally well for all groups. In more rigorous terms, Jensen defines any test as biased for a particular group if that group differs significantly from the majority group in the slopes, intercepts, or standard error of estimates of their regression lines. A test might be biased in any combination of these ways, but the evidence on blacks vs. whites (obviously the most popular comparison in the literature)

is fairly clear: most studies find no difference in the regression functions and those tests that have been found to be biased have been found to favor blacks. This does not mean that the blacks have scored higher on the tests but that their actual performance on the criterion (for example, grades in Officer Candidate School) were somewhat lower than predicted by the selection test (the Navy's Officer Qualification Test). In short, where bias has been found it has usually overpredicted rather than underpredicted the minority group's performance.

Given these statistical tools, the issues surrounding fairness in selection become thorny. It is possible to improve the predictability of individuals' scores first by correcting for the known degree to which the test is not perfectly reliable (by transforming scores to "estimated true scores") and then by taking account of group membership along with the test score, in a multiple-regression equation. Is that fair? Jensen writes, "Until we find out what the relevant psychological predictors are for which

racial classification per se is merely a 'stand-in' variable, we have no other choice, if we wish to improve predictive accuracy, but to include race (or other group membership) as a predictor variable along with the test scores or other predictive measures. On the other hand, if the overprediction of the minority's group criterion performance is not too extreme, it may seem reasonable to many to leave it uncorrected, thereby giving the benefit of the slight predictive bias to the presumably less advantaged group."

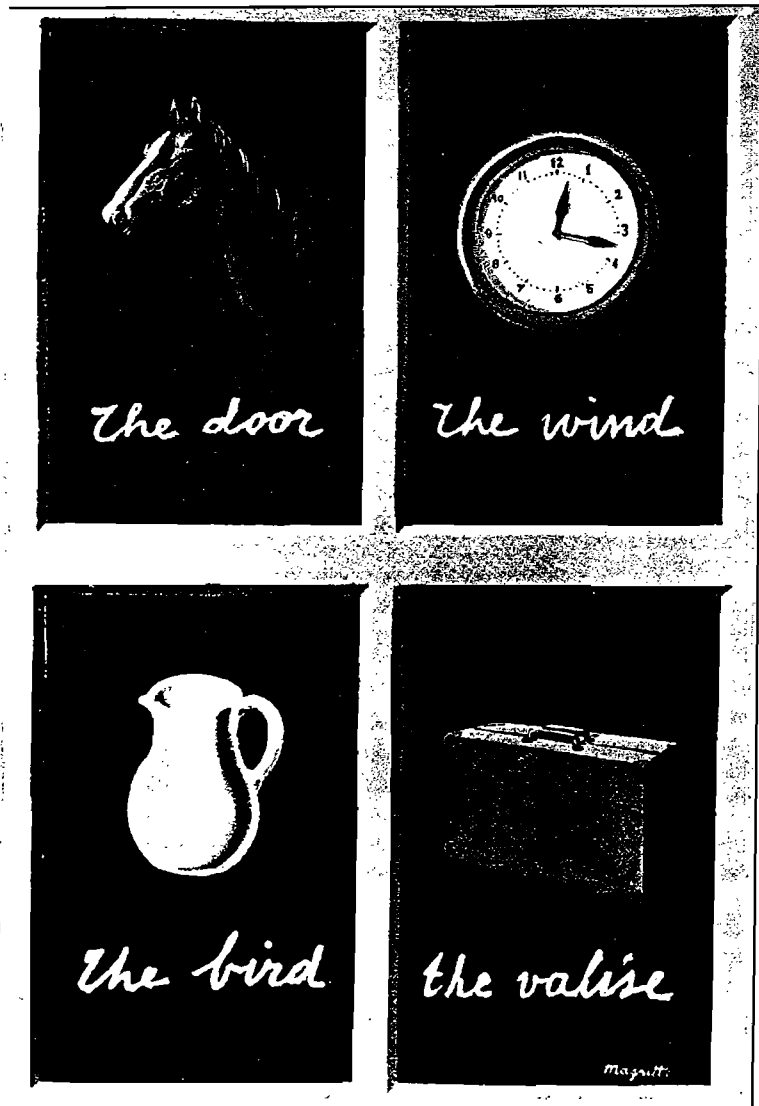
### *The G Factor*

Jensen has some ideas about what those relevant psychological variables might be. He is a big believer in *g*, which is the psychometrician's symbol for the general factor accounting for the high correlation among many diverse tests of particular abilities. Race and social class differences usually amount to about one standard deviation (15 IQ points) on those tests that are highly loaded on the *g* factor, like the Stanford-Binet. Specific abilities high in *g*, such as verbal analogies and arithmetic reasoning problems, differ substantially between the races while low-*g* abilities such as speed of addition, counting, and rote memory do not. Late in the book (around page 700) Jensen proposes that biological differences in reaction time of the neurons may explain the individual differences that underlie *g*, differences in reasoning and in seeing abstract relations.

Though such ideas are not central to the thesis of this book, Jensen, like Charles Spearman, Clark Hull, and others, is obligated at least to speculate about the neural explanation of *g*. For there is a competing view that attributes the *g* factor to psychometricians' inability to differentiate the tests further into specific abilities that would tell us more about the nature of intelligence. We understand little about what the human brain is doing when the human mind is thinking, but what is known suggests processes of such great complexity that it is hard to believe individual differences in intellectual power can be largely reduced to a single dimension of neurophysiological functioning. That is what the believers in *g* claim, and the reason the hypotheses at stake have much greater psychological importance than merely the verification and arbitrary labeling of a principal component in a correlation matrix. The frontier of research and controversy in the field of mental testing is moving away from educational policy back to the questions that motivated the great pioneers in the field: questions about the nature of human intelligence.

### *Black and White Issues*

One severe limitation of this entire book, however, is the constant dichotomizing of people into comparison groups, as though race were a sufficient classification scheme either for selection purposes or for psychological



The Key of Dreams (1936) by René Magritte. Collection of Jasper Johns. G.D. Hackett.

research. It is not. Adam is white, but he is also an unmarried male Lutheran mechanic of Swedish descent born and raised in New York City. Eve is black, but she is also a married woman, a Catholic, an Indiana farmer, only daughter of a veterinarian, a mother, and a crossword puzzle freak. They both take a Civil Service exam. As Jensen frames the issue of fairness, it would be this: Should Adam's and Eve's "group membership"—i.e., race—be taken into account in interpreting their test scores? But Adam and Eve belong to many groups which may be more important to them than their racial ancestry. Why is Eve more "black" than she is "farmer" or "puzzle doer"? Furthermore, most American so-called blacks and many so-called whites are really gray; in this sense the concept of race has little real biological meaning.

Skin color is a characteristic we all use, with much prejudice but also with some predictive validity, to categorize strangers. When we get to know them as individuals their race turns out to be less important than their occupation, sex, talent, kindness, hobbies or political views. Contemporary legislation and jurisprudence have managed to go beyond the polarizing issue of race to establish the principle that selection must be based on assessments as close as possible to the criterion itself. The technical refinement of probabilistic prediction must give way to freedom of opportunity to prove predictions wrong, which means to be a person instead of a black or a white. While Jensen espouses this view explicitly in several places, it is undercut by the dominant theme in this book, white vs. black test scores, and by his warnings about the dire consequences of affirmative action. For example, he argues that affirmative action must always lead to lower productivity: "It may be regarded as socially and economically important that a minority group be adequately represented in a school or an occupation, even if bringing this about means lowering the selection cutoff for minority applicants and incurring a higher failure rate among the selectees or readjusting the failure rate by tolerating a somewhat lower level of performance."

#### *Uses and Abuses of Tests*

Surprisingly, Jensen expresses in the last chapter the radical view that routine mental testing of schoolchildren serves no useful purpose and is often destructive.

This last chapter should be read first so that the reader does not assume the rest of the book is striving toward some other conclusion. In fact, it is a nagging weakness in the first fourteen chapters that we never understand why scholastic achievement needs to be predicted in the first place. The fundamental difference between aptitude tests and achievement tests is spelled out too late and with too little emphasis. Aptitude tests (IQ and other tests of "ability" both specific and general) consist of items selected for their psychometric properties in a given population and for their statistically proven predictive validity for some single criterion. If it were shown that memory for series of digits predicted success in dentistry, then a digit memory test would be a good aptitude test for prospective dentists. But even if its predictive validity were perfect (1.0), no state licensing board would adopt that test in place of an achievement test consisting of information actually taught in dental school.

This is an important point because it shows precisely where the limits of science's responsibility are and will remain. The technology of mental testing could be carried to perfection, prediction coefficients could be computed for every known demographic group, and it would still be wrong to deny an applicant the opportunity to prove himself just because, for example, unmarried male Lutheran mechanics usually fail in that particular job. Nothing in the first 714 pages prepares us for the fact that Jensen himself reaches just those conclusions. As staunchly as he defends the tests and argues for the existence of real and biologically determined differences between the races, his considered views on the uses and abuses of tests pose a serious threat to the multimillion-dollar testing industry. He opposes ability grouping or "tracking" in schools, advocating what others call "mastery learning": the use of achievement tests to diagnose precisely when pupils have learned the prerequisites for more advanced courses and precisely which units they need to repeat.

Has he changed his views? "We have known for years that much of the variation in level of scholastic achievement among schools is predictable from a number of demographic characteristics of the population that they serve. But the schools themselves can have no control over such community characteristics as socioeconomic level, racial composition, average educational and occupational

level of the adult population, and the like." This is substantially what Jensen said in 1969. It does not imply, to him, justification of white supremacy. "Righting the past wrongs of racial discrimination cannot be furthered by blaming the mental tests (which we admittedly should continue to improve and to use more wisely), but by prohibiting racial discrimination in any form, by legal sanctions when necessary, and by seeking equal educational opportunities for members of those minority groups that have been denied them in the past, so they can compete fairly, as individuals, in selection for employment, technical training, or higher education, without condescending dispensations."

#### *Believers*

Political and social values should not be allowed to control what is accepted as the current scientific truth. The controversy about racism within science in the decade just past led many to propose (sometimes successfully) that research on individual differences should not be funded and that believers in the heritability of IQ should not be allowed to speak. I have heard colleagues say that the *Harvard Educational Review* should not have published Jensen's 1969 article, that he himself was asking for trouble and deserved the abuse he received.

In fact, though, many of Jensen's opponents allowed their scientific conclusions to be far more biased by their political views than he did. For some this was a conscious act of "social responsibility." There was (and still is) confusion between promoting science as a tool for social progress and dictating to science what can and cannot be its conclusions. David Hume said it two centuries ago: "There is no Method of Reasoning more common, and yet none more blameable, than in philosophical Debates, to endeavour the Refutation of any Hypothesis, by a Pretext of its dangerous Consequences to Religion and Morality. When any Opinion leads into Absurdities, 'tis certainly false; but 'tis not certain an Opinion is false, because 'tis of dangerous Consequence."

This era was a bleak chapter in the history of American social science, and it is remarkable that Arthur Jensen has emerged from it with his talent for careful compiling of evidence, his lucid writing style and his courage intact. He is neither absolutely right, nor absolutely wrong. His book deserves to be taken on its merits. □