DOCUMENT RESUME

ED 152 839 TM 007 051

AUTHOR Dowell, David A.
TITLE Methodological Considerations in Intelligence Testing.
PUB DATE [77]
NOTE 14p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Heredity; Intelligence; *Intelligence Differences; Intelligence Tests; *Nature Nurture Controversy; Racial Differences; *Research Problems; *Research Reviews (Publications); *Statistical Bias; Test Interpretation; Test Reliability; Test Validity
IDENTIFIERS Jensen (Arthur R)

ABSTRACT

This paper examines the question of the hereditary nature of intelligence and the validity of some of the statistical procedures which have been used in measuring the degree of hereditability. The author feels that proof of the question lacks sufficient scientific rigor for the support of any conclusion, particulary for a question of such political and emotional importance. (CTM)

METHODOLOGICAL CONSIDERATIONS IN INTELLIGENCE TESTING

I. Introduction

Jensen's (1969) article suggesting the possibility that capacity for intelligence is largely inherited has ignited a strom of controversy. During a period in American history marked by concern over the lot of minority citizens, it is understandable that any suggestion of Black racial inferiority is a touchy subject. This is true whether the heritability of intelligence issue has logical social or political implications because so many people believe that it does have such implications.

The controversy is nearly unique in science because of the juxtaposition of violent emptions and complex technical questions. Rarely does so much emotion hinge upon obtruse mathematical subtleties. There is no doubt that the technical questions are not simple. Honest dispute of the correct application and interpretation of methods is an important aspect of this problem. But, in an important sense, the technical aspects of this issue are separate from the political and social implications purported to follow by such writers as Hernstein and Terman. Without a consensus among the scientific community on so complex an issue, political and social conclusions do not carry the weight of scientific findings. The conclusions drawn by those who choose to do so despite scientific disagreement ought to be taken as seriously as any other political allegations, given the credibility of the source, but do not carry the weight of scientific findings. This separation of the scientific and the social-policical implications is not always appreciated.

The paper which I am presenting today deals with a narrow but important aspect of the intelligence testing controversy: the mathematics of the heritability of intelligence. Jensen's interpretations of data are at the center of the controversy and I will necessarily spend some time discussing these. The following presentation is organized into three parts. First, I will introduce the problem and describe the elements of the argument that intelligence is largely inherited. Next, I will present a critical review of this argument. Interpretational problems to be discussed include restriction of range, within group versus between group heritability, the internal validity of monozygotic twin studies, and the question of regression toward the mean. Finally, the technical issues are summarized and the conclusions which can be drawn are presented.

II. The Problem

The argument that IQ testing data yields a meaningful and high estimate of the heritability of intelligence involves several logical steps. First, IQ tests are presumed to be valid measures of intelligence; that is, they are assumed to be a representative sample of the abilities which comprise intelligence. To the extent that the abilities tapped by IQ tests do not fairly represent intelligence (however defined), then IQ tests are not valid.

IQ tests assume that intelligence has an underlying structure which is monotonic. By this is meant that true intelligence scores can be unambiguously ranked as higher, equal, or lower with respect to each other. It is not permissable that one individual be regarded as more intelligent in one respect and less intelligent in another respect relative to a second person. This requirement is necessary for the statistics applied to IQ tests to be appropriate.

Of course, there is an important difference between the heritability of IQ and the heritability of intelligence. If we are satisfied to study the heritability of IQ, the foregoing assumptions are unnecessary.

If it is assumed that IQ tests are a valid measure of intelligence, we next come to the question: how much of intelligence is inherited? The technique for measuring heritability grows out of correlation and analysis of variance. It is possible to calculate the porportion of variability in a dependent variable which is perfectly correlated with or predicted by the independent variable. If the independent variable, X, is suspected to cause the dependent variable, Y, then the porportion of variance predicted by the independent variable is a measure of how much X determines Y. It is important to remember that porportion of variance predicted is calculated on and applie to the sample data.

The variance in intelligence scores determined by genetics has been studied using twins reared apart and adopted children. If the correlation between the IQs of identical twins reared by different families is high despite their (apparantly) different environments, then, the argument runs, the twins whose genes are the same must have inherited the similarity of their intelligence. It is believed that the "different" (sic) environments of the twins cannot be the source of the similarity in IQs. Studies of foster children have the same purpose. If the IQ correlation between children and their natural parents is higher than that between adopted children in the same family and their (adopting) parents, then this is taken to show that despite similarity (sic) of environments, the difference in genes still produces a difference in IQs.

Finally, statistical geneticists correct the heritability values obtained in these studies on two grounds. First, unreliability of a measure decreases the correlations between that measure and other variables as outlined by McNemar (1969). Correcting heritability values for unreliability of IQ tests, therefore raises heritability estimates as carried out by Jensen (1969). Secondly, it is presumed that the range of genes determining intelligence is less in the studies carried out than in the population as a whole. Because restriction of range also decreases correlations as outlined by McNemar (1969), the heritability estimates are raised by the correction for restriction of range carried out by Jensen (1969).

Because restriction of range also decreases correlations as outlined bu McNemar (1969), the heritability estimates are raised by the correction for restriction of range carried out by Jensen (1969).

These are the basic arguments necessary in making estimates of the heritability of intelligence. Now, I shall critically examine each of these and also some other related arguments.

III. Heritability of Intelligence and Related Issues

The effect which restriction of range has upon statistical associations is a critical methodological consideration in intelligence testing. Jensen inflates the porportion of IQ variance due to heredity from 45% to around 85% on the basis of corrections for restriction of range and unreliability. If sample variance is less than population variance in the independent or dependent variables or both, then resrriction of range is a consideration which may affect correlation coefficients. Small samples suffer from restriction of range since the chance of picking up very extreme cases in a small sample is low. In general, restriction of range depresses statistical associations because the ratio of explained to total variance is reduced.

Before specifically examining the logic of the restriction of range correction, let us review a few elementary facts about the correlation coefficient. The correlation coefficient may be expressed as the square root of a ratio. That ratio consists of predicted variance in the criterion divided by total variance in the criterion. Total variance in the criterion is comprised of predicted variance plus error variance. In analysis of variance terms, the source of predicted criterion variance is the predictor variable. Variance attributed to error arises from all other sources. Since the correlation coefficient is the ratio of these two variances, then corrections to the correlation coefficient must be based upon the relative size of the numerator and denominator obtained from a sample compared to what would obtain in the ideal experiment or study.

The restriction of range correction is based upon the notion that for variaous reasons, in some samples, the range of the independent variable is less than occurs in the population to which results will be generalized. If the range ofthe independent variable is restricted then the obtained sample correlation coefficient may be attenuated because the numerator of the ratio is smaller than it otherwise would be. The restriction of range correction is appropriate when the independent variable is restricted and when all other sources of variance are unrestricted. If, however, the ranges of other sources of variance which contribute to error are also restricted, then the denominator will also be attenuated. In such a case, the effect fo the several restrictions of range on the size of the obtained sample correlation coefficient depends on the <u>relative</u> restriction of sources contributing to predicted and error variance. For such a case, the correction for restriction or range is inappropriate.

In applying the restriction of range correction to the studies of the heritability of intelligence, the question to be asked is this: is the variance in IQ scores attributed to error less in the study samples than in the population. In the crucial studies of the heritability of intelligence, the error term included all variance not arising from the genetic variable. Of course, when the error term is defined in this way, an important source of variability contributing to error is environment. The crucial question then resolves into this: is the variance due to environmental factors less in the study samples than in the population? If so, then the correction for restriction of range will be inappropriate.

Kamin (1975) describes the techniques used in some separated twin studies to recruit subjects. In one study, by Newman, et. al. (1937) newspaper and radio ads solicited pairs of twins who reported that they were "so strikingly similar that even your friends and relatives have confused you." Some were rejected because they admitted they were not raised apart or because, while they were identical, they had "different dispositions." Those who qualified were rewarded with a free trip to the Chicago Exposition. This took place during the depression of the 1930's. Such selection techniques potentially produce very biased samples although the direction and amount of bias is partially unknown. It seems clear that selection by the methods outlined will produce a sample of twins reared in very similar environments. As Kamin outlines, all four important studies of separated monozygotic twins suffer from such poor sampling designs. Such studies have no external validity because the sample cannot be viewed as representative of any larger population of interest. The correction for restriction of range applied to these studies by Jensen is clearly inappropriate since it clear that environment is restricted at least as much as genentics.

The problem of within group versus between group heritability is closely related to the restriction of range consideration. We can illustrate this using an example cited by Layzer (1975). Consider two populations of wheat. The first is highly inbred and therefore has a completely homogenous genetic makeup. If grown under varying environmental the height of the stalk will be a highly variable trait. The gene-height correlation, that is heritability, will be zero since the variation in genes will account for none of the variation in height. The second population of wheat has a heterogenous genetic make-up and is grown under uniform environmental conditions. Heritability in this population will approach 1.0 since virtually all variation in height depends upon genetic make-up. Heritability estimates based upon one of these populations does not apply to the other. The reason for this is the same that prevents generalizability of biased samples. The first population represents an instance of restriction of genetic range In the second, environment is restricted.

Consider the analogy to black and caucasian environmental and genetic factors. If complaints of minority groups reflect reality, then the environmental conditions under which the two groups live differ systematically. Not only do the modal environments differ but there is little overlap between the extremes of the distributions of the two groups especially at the high end.

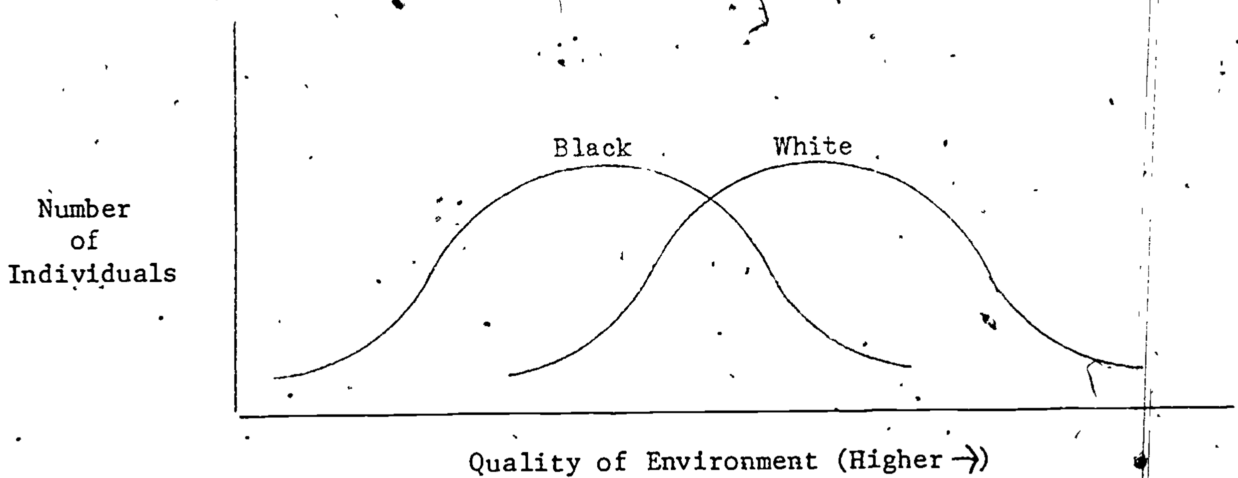

Fig. 1. Putative "Quality" of Environments of U.S. Black and White Populations. "Quality" is an Abstraction Reflecting the Contribution of Environment to IQ Score.

Since we do not have evidence that the ranges and distributions of Blacks and Whites are similar with respect to whatever environmental and genetic variables affect IQ scores, then a heritability estimate based upon White subjects does not apply to Blacks. More simply stated, we have reason to suspect that the range of environments in which Black children develop is more restricted than that of Caucasian children. This may result in less variation in IQ among Blacks than Caucasian. Moreover, since it is the "better" environments that Blacks are generally denied, then the lack of variation would be expected to largely appear as lowered IQs.

The fact that evaluations of compensatory education have failed to show sustained gains on achievement test scores is sometimes cited as evidence that changes in environment have little effect upon mental ability (Jensen, 1969). However, the compensatory education research was plagued with methodological difficulties making results uninterpretable (Campbell and Erlebacher, 1970). Bronfenbrenner (1975) reviews several studies of intensive early intervention which have produced gains of as much as 20 points on IQ tests. The evidence is clear that even short term alterations of environments can affect IQ scores.

Comparability of Environments

Ignoring the lack of external validity in the separated twin studies, we now turn to the question of internal validity. If monozygotic twins are separated at an early age and placed in different environments, then, the geneticist argument runs, any similarity in their IQ scores must be produced by their identical genes. This holds only if the environments in which they are placed are random with respect to each other. In these studies of separated monozygotic twins and also in the adopted children studies does the assumption of random environments hold?

No. First of all, families who adopt tend to be more wealthy than average (Goldberg, 1976). This results in a restricted range of environments. We have seen that this depresses the environment-IQ correlations. Secondly, adopted children tend to be placed in homes more similar to those of their natural parents than random. This is evident in Kamin's (1974, 1976) analysis of Shield's (1962) data. Adopted children are often raised in the homes of friends or relatives. Simply being adopted into the same community is a significant homogenizing factor. Thirdly, genetic-environment interaction is an important factor. For example, two twins with genetic endowment for athletic ability are very likely to both choose to participate in sports. The environments to which they are subsequently exposed will be significantly determined by such a choice and correlated. Finally, it is important to note that each of these factors sends to spuriously inflate estimates of heritability based on studies of monozygotic twins. Because of the unknown magnitude of these effects, heritability extimates as they have been computed are of little more use than the trivial statemant that the upper bound of heritability is 1.0 (Layzer, 1975) and may be considerably more misleading. Kamin (1974, 1976) reviews all four crucial twin studies carried out on monozygotic twins reared apart. Obtained uncorrected correlations were 0.86, 0.77, 0.67, and 0.67. In light of the above considerations alone it is hard to agree with Jensen that heritability of intelligence probably lies between 0.7 and 0.8 with the best estimate being the higher. The more reasonable position would seem to be that the least value obtained in the studies, that is 0.67^2 or 45%, is an estimate of the upper bound of the heritability of intelligence.

Now, I would like to turn to the question associated with regression toward the mean. It has been argued by Jensen (1969) that if the IQ score of blacks regresses toward the mean on retest more than a matched sample of whites above the mean, then this indicates that there exists a lower true population mean for blacks than for whites. This statement is based upon the belief that the more extreme is a score, the more it will regress on posttest. There are some difficulties with this argument.

The amount of regression toward the mean which will occur in a subgroup depends upon the mean error term associated with the subgroup. The larger the absolute value of the error term, the more regression will occur. It is impossible to unfailingly separate the error and true score components an the observed scores of any subgroup, of course. However, in theory the size of the error term depends upon the reliability of the criterion and the location in the distribution of the subgroup. My own recent research on regression (Dowell, 1977) shows that whenever matching or selection is carried out on the basis of any variable correlated with a criterion that differential regression will occur. This is precisely the situation to be expected with race differences in IQ scores. Since race is correlated with IQ, differential regression will occur. This merely indicates subgroup differences in reliability and observed score means and indicates nothing about true scores. It is wholly unwarrented to state that true population means may be estimated by such evidence as differential regression arising from unreliability and matching on the basis of a correlated variable.

IV. Conclusions

The mathematics of heredity analysis are complex enough that some honest disagreement among professionals occurs. Social and political assertions based upon conclusions about scientific aspects of the matter do not carry the weight of scientific findings.

An analysis of the methods used in studies of heredity and environmental effects on intelligence revelas that all suffer from very weak external validity. The unknown effects of the bizarre sampling techniques probably resulted in biased samples. Moreover, because of the bias, the use of the restriction of range correction is wholly inappropriate for heritability estimates. Heritability factors may be specific to population subgroups. Hence, estimates of the heritability of IQ and intelligence are very suspect.

Ignoring external validity for a moment, it appears that IQ studies are weak in other ways. The correlated environments of adopted twins as well as gene-environment interactions make internal validity of these studies suspect. Particularly important is the fact that the correlated environment and gene-environment considerations each tend to <u>inflate</u> heritability estimates. Finally, assestions that regression toward the mean can be used to identify population true scores is very misleading and incorrect. Different subgroup error components and different observed score means makes differential subgroup regression likely without implication true score means.

IQ testing is a complex area for further scientific research. The glare of publicity and the heat of controversy are inappropriate in connection with a topic on which no scientific concensus has emerged.

References

Bronfenbrenner, Urie. Is Early Intervention Effective? Some Studies of Early Education in Familial and Extra-Familial Settings. In Montague, Ashley (ed.) Race and IQ. New York: Oxford University Press, 1975.

Campbell, D. T. and Erlebacher, A. How Regression Artifacts in Quasi-experimental Evaluation Can Mistakenly Make Compensatory Education Look Harmful. In Hellmuth, J. (ed.) Disadvantaged Child, Vol. 3. New York: Bruner/Mazel, 1970.

Campbell, D. T. and Stanley, J. Experimental and Quasi-experimental Designs for Research. Chicago: Rand-McNalley, 1967.

Dowell, D. Regression toward the mean: A conceptual model and empirical test. Manuscript, 1977.

Goldberger, A. S. Mysteries of the Meritocracy. In Block, N. J. and Dworkin, G. (eds.) The IQ Controversy. New York: Pantheon Books, 1976.

Hernstein, R. J, IQ in the Meritocracy. Boston: Little Brown and Co., 1973.

Jensen, A. R. How Much Can We Boost IQ and Scholastic Achievement? Harvard Educational Review, v. 39, 1969.

Kamin, L. The Science and Politics of IQ. New York: John Wiley and Sons, 1974.

Kamin, L. Heredity, Intelligence, Politics and Psychology: II. In Block, N. J. and Dworkin, G. (eds.) The IQ Controversy. New York: Pantheon Books, 1976.

Layzer, D. Heritability Analyses of IQ Scores: Science or Numerology? In Montague, Ashley Race and IQ. New York: Oxford University Press, 1975.

McClelland, D. Testing for Competence Rather than Intelligence. American Psychologist, v. 29, 1973.

McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, Inc., 1969.

Newman, H. H., Freeman, F. N. and Holzinger, K. J. Twins: A Study of Heredity and Environment. Chicago: University of Chicago Press, 1937.

Scarr-Salapatek, S. Unknowns in the IQ Equation. In Block, N. J. and Dworkin, G. The IQ Controversy. New York: Pantheon Books, 1976.

Shields, J. Monozygotic Twins Brought Up Apart and Brought Up Together. London: Oxford University Press, 1962.

Terman, L. M. A New Approach to the Study of Genius. Psychological Review, 1922.