

Genetic correlations, squared or unsquared: Can they be accurately computed?

LEONARD I. JACOBSON

University of Miami, Coral Gables, Florida 33124

Jensen has suggested that the problem of genetic correlations may be understood as analogous to a problem in traditional test theory. In the present paper, the problem of conceptualizing phenotypes, genotypes, and environmental events in this manner was explored. Analysis indicated that when these factors are viewed in this context, many of the assumptions of traditional test theory are found to be grossly violated. Therefore, if the problem is understood in terms of traditional test theory, it may be concluded that the use of heritability coefficients such as Jensen's may be expected to yield results in error to a substantial but unknown degree.

In recent years a number of papers have appeared disputing Jensen's (1971) contention that correlations between phenotypes (observed traits reflecting the heritability component) may be viewed directly as indicative of the proportion of phenotypic variance resulting from heritability (Linn, 1974; Miller & Levine, 1973). The problem is viewed by Jensen (1971) as analogous to a situation in classic test theory in which the reliability of a measure is determined by means of the correlation of equivalent forms. In this light, phenotypes are observed scores, genotypes represent hypothetical true scores, and environmental variables are error variance.

Jensen presents the following equation as an expression of this view: $r_{pp'} = r_{pg}^2$, where $r_{pp'}$ is the correlation between phenotypic measures and r_{pg} is the correlation between a phenotypic measure and its assumed genotype. Jensen further states: "The correlation between identical twins reared in uncorrelated environments is directly analogous to the correlation between equivalent forms of a test. The correlation in each case indicates the percentage of variance in common, or the percentage of genetic (or true score) variance (1971, p. 223)."

Although the conceptualization itself has been a subject of heated controversy, the purpose of the present paper is to demonstrate that even if one accepts such an approach, serious difficulties emerge in viewing the situation as a simple derivative of test theory. These difficulties concern the application of the assumptions of test theory to the conceptualization and measurement of genotypes, phenotypes, and environmental events.

The author acknowledges with gratitude the assistance of Edith E. Lord and Carrol V. Truss. Their perceptive comments and insights greatly strengthened this manuscript. The author also wishes to thank Marc P. Kaufman, whose comments suggested the need for the present paper.

THE ASSUMPTIONS OF TRADITIONAL TEST THEORY

In their discussion, Miller and Levine (1973) specified three statistical assumptions from classic test theory required by Jensen's theoretical formulation. These assumptions are (1) $r_{ge} = 0$, (2) $r_{EE'} = 0$, (3) $\bar{E} = 0$, $\bar{E}' = 0$; where g is the genetic component, e is the environmental component, E and E' are environmental events, and \bar{E} and \bar{E}' are the means of E and E' .

The Jensen (1971) model assumes these assumptions are met; if they are not, a number of problems occur. Violation of the first assumption results in a confounding of purely genetic components and the interaction between genetic and environmental components in $r_{pp'}$. Miller and Levine (1973) indicate that heritability studies, in fact, do violate this assumption. The method of randomly assigning members of the same kinship pair to different groups may be shown to control *within-pairs* sampling bias. It does not control for phenotypic differences *between pairs*. Such errors may result from covariance attributable to similar and systematic environmental influences on phenotypic performance for both members of a pair.

The second assumption assumes an independence of different random environmental events. Prenatal environments of identical twins are not random, nor can it be assumed that twins separated at birth have equal and unrelated prospects for development. Therefore, it seems likely that this assumption is violated also. The correlations among variables such as socioeconomic status, educational level, and intelligence, even for genetically unrelated individuals, provide further evidence that Assumption 2 is not a reasonable one. Violation of this assumption may result in biasing in either direction the correlation between phenotypes on which heritability estimates are based, as well as in biasing estimates of the reliability coefficient.

The third assumption is derived from Jensen's (1969) view that environment is a threshold variable. That is, environment has no positive effect in its own right. However, environmental deficiencies can interfere with the attainment of genetic potential. Thus, environment is a type of gatekeeper; statistically put, environment is a source of error variance that may be conceptualized as a negative factor with a negative mean and a maximum value of zero. The fact that environmental events may operate so as to result lawfully in distortions of the g potential indicates that under certain circumstances r_{ge} cannot be zero (Assumption 1) because r_{pe} is not zero.

Linn (1974) reviews Miller and Levine's (1973) paper and indicates a fourth assumption that Jensen's (1971) model requires. This assumption, $\sigma_{E'}^2 = \sigma_E^2$, follows directly from test theory also, if E and E' are thought of as aggregates of random error variance. Linn indicates that Assumption 4 is violated also in practice for much the same reasons as Assumptions 2 and 3.

THE EQUIVALENT FORMS PROBLEM

Although previous analyses have indicated important problems with the heritability approach, they have not considered perhaps the most critical issue, whether the problem of genetic correlations may indeed be viewed as one involving equivalent forms. That is, can g be equated with true score and p with scores on equivalent forms, in terms of the assumptions of test theory? The primary issue here concerns the adequacy of estimations of p generally, and not merely in twin studies. If the problem is to be conceptualized in terms of equivalent forms, a number of additional assumptions must be confirmed. The following assumption is the traditional definition of equivalent forms:

$$r_{p_1 p_2} = r_{p_1 p_3} = r_{p_1 p_4} = \dots = r_{p_1 p_n}, \quad (5)$$

where $p_1 \dots p_n$ are a series of phenotypic estimates of the same quantity. It follows, then, from Jensen's belief that p is an unbiased estimate of g that:

$$r_{p_1 g} = r_{p_2 g} = \dots = r_{p_n g}. \quad (6)$$

If Assumption 1 that $r_{ge} = 0$ and Assumptions 5 and 6 above are correct, then:

$$r_{pe} = 0. \quad (7)$$

The difficulty for Jensen's (1971) model is that IQ data do not yield results in accord with the assumptions above. For example, Matarazzo (1972) presented the correlations of the Wechsler Adult Intelligence Scale (WAIS) with other IQ tests. The correlations ranged from .50 to .90 with a mean correlation of .75 (the

reliabilities may be assumed to be about .93). If Assumption 5 is correct, all of the r_s should be equal and the correlations should approach that permitted by the reliabilities of the tests involved. It seems doubtful that random errors can be invoked to explain these differential correlations.

Considerable difficulties emerge also if Jensen's belief that p is an unbiased estimate of g is correct. In his presentation, Matarazzo (1972) found a lawful pattern in the correlations cited above. Contrary to expectations based on Jensen's assumptions, the correlations varied with the specific test employed, characteristics of the population sampled, and the range of intelligence levels evaluated. Some other factors probably related to these variations in reported r_s are the following: conditions of administration, temporal duration between measurements, instructional set, personality, and sociomotivational factors.

Matarazzo's (1972) data provide a good example of the effects of the specific test employed and the temporal duration between measurements on p . The lowest correlations obtained were for longitudinal studies in which the Stanford-Binet and the WAIS were given over 20 years apart. The mean correlations here averaged about .5. The highest correlations, above .80, were obtained when the temporal interval between the tests was smallest. Now, if p is truly an unbiased estimate of g , and assuming g remains constant, temporal interval and the specific test employed should be of no consequence, apart from the effects of random error variance on test scores. Since these non- g variables are clearly of importance, Assumption 6 cannot hold completely in this case. Similar instances of violations of this assumption could be given for all of the aforementioned factors as well.

Considerable evidence against the belief that environmental events represent only error variance unrelated to both phenotypes and genotypes is provided by studies of the effects of environmental manipulations on IQ scores (Bernal, Jacobson, & Lopez, 1975; Jacobson, Berger, Bergman, Millham, & Greeson, 1971; Jacobson & Greeson, 1972). These studies demonstrate that IQ scores may be increased predictably as a function of the individual's experience with systematic learning programs. The changes obtained are quite durable, as measured over at least 2 years, and can in no way be construed as resulting from training that spuriously biased performance on an IQ test. The finding that IQ scores can be permanently altered as a result of behavioral training conflicts with Assumption 7.

Further support for this interpretation may be found in Hunt's (1973) review of empirical studies concerning heritability and environmental contributions to performance. Reevaluation of the data in the Coleman (1966) report, for example, indicates that the 25% of the variance in scholastic achievement attributed to racial and ethnic group membership resulted primarily

from social conditions not originally investigated adequately. When all of the relevant factors were considered, the variance remaining that could be attributed to hereditary factors dropped to 1.2%.

CONCLUSIONS

It is apparent that genotypes, phenotypes, and environmental events cannot be conceptualized adequately in terms of the assumptions of the traditional test theory model. Therefore, if the problem is understood in terms of traditional test theory, it may be concluded that the use of heritability coefficients such as Jensen's should be avoided, since they result in errors to a substantial but unknown degree.

It is reasonable to ask what type of model *would* be an appropriate one. The answer to this question is currently unknown and the reader is referred to Layzer's (1974) discussion of some of the problems in conceptualizing variables in the area of intelligence. Perhaps it should be noted that if some of the variables evaluated are found eventually to be nonlinear and/or nonadditive and/or nonmonotonic, with covariances obtained where orthogonality was assumed and multiplicative combinational relations present for good measure, then we may still be very far from even understanding the problem in a meaningful way.

REFERENCES

- BERNAL, G., JACOBSON, L. I., & LOPEZ, G. N. Do the effects of behavior modification programs endure? *Behaviour Research and Therapy*, 1975, **13**, 61-64.
- COLEMAN, J. S. *Equality of educational opportunity*. Washington, D.C: U.S. Government Printing Office, 1966.
- HUNT, J. M. *Assessment in a pluralistic society—proceedings of the 1972 invitational conference on testing problems*. Educational Testing Service, 1973.
- JACOBSON, L. I., BERGER, S. E., BERGMAN, R. L., MILLHAM, J., & GREESON, L. E. Effects of age, sex, systematic conceptual learning, acquisition of learning sets, and programmed social interaction on the intellectual and conceptual development of preschool children from poverty backgrounds. *Child Development*, 1971, **42**, 1399-1315.
- JACOBSON, L. I., & GREESON, L. E. Effects of systematic conceptual learning on the intellectual development of preschool children from poverty backgrounds: A follow-up study. *Child Development*, 1972, **43**, 1111-1115.
- JENSEN, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, **39**, 1-123.
- JENSEN, A. R. Note on why genetic correlations are not squared. *Psychological Bulletin*, 1971, **75**, 223-224.
- LAYZER, D. Heritability analysis of IQ scores: Science or numerology? *Science*, 1974, **183**, 1259-1266.
- LINN, R. L. Unsquared genetic correlations. *Psychological Bulletin*, 1974, **81**, 203-206.
- MATARAZZO, J. D. *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins, 1972.
- MILLER, J. K., & LEVINE, D. Correlation between genetically matched groups versus reliability theory: A reply to Jensen. *Psychological Bulletin*, 1973, **79**, 142-144.

(Received for publication November 22, 1977.)