I.Q. HERITABILITY, RACE DIFFERENCES, AND EDUCATIONAL RESEARCH

BY N. L. GAGE

The most important way to disprove Jensen's hypothesis is to reduce race differences through education and other kinds of environmental influence. Unless educators and others who influence the environment are effective in making Negroes as successful educationally and occupationally as whites, Jensen will for practical purposes have won the argument. The data are equivocal, as endless clashes of opinion among well-intentioned scientists indicate. But the opinions of scientists are not, in any event, our primary concern, whether we are black Americans or white. What we really want for our society is the elimination of racial differences in school and job success.

Jensen hypothesized that "genetic factors are strongly implicated in the average Negro-white intelligence difference...."[1] This hypothesis provokes intellectual and political fights because of its implications, if accepted, for legislation aimed at improving the education of Negroes and other low-income students. It can be construed as implying that such legislation is futile. If that belief is accepted by political leaders, our governments will not give educators,

and their research and development arms, the money and other resources they need for work toward reducing the educational and employment disadvantages suffered by Negroes and other minorities. It is because of what Professors Jensen and Shockley say to the President and the Congress that educators, who want a fair chance to try their approach, should be concerned.

In what follows, I intend, first, to examine some underemphasized aspects of the data on whites that are cited in the controversy. Then I shall consider the relevance of these data, however valid they may be for North European and American whites, to the question of genetic determination of Negro-white differences in mean I.Q. Third, I shall call attention to the need for the educational research and development that can produce necessary improvement in the achievement and attitudes of Negro youth and, ultimately, provide the only definitive test of Jensen's hypothesis.
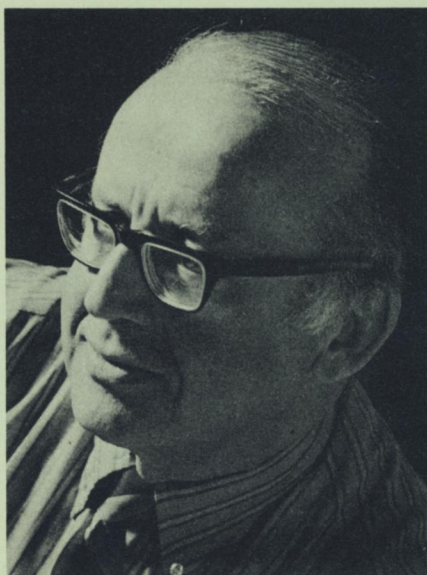
Identical Twins Reared Apart

The logic of the problem of estimating genetic and environmental variance

in I.Q. leads us to identical twins reared apart as the source of the most relevant data. Such twins provide natural experiments in which the genetic source of variance is eliminated. Hence any remaining differences in I.Q. between such twins can be attributed (apart from errors of measurement) to variance in environmental factors — ranging from the intrauterine position and nutrition of the fetus to the quality of the home, neighborhood, and school in which the child receives his education. It is little wonder that — although they also refer to data on identical twins reared together, and on fraternal twins and ordinary siblings reared together and apart, and on real and foster parents and children, and persons in other blood and environmental relationships — students of the problem find their most convincing data to be the I.Q.'s of identical twins reared apart. Thus, Jensen, subsequent to the storm aroused by his 1969 paper, published a new analysis of the four largest studies of such twins.[2]

But it is not enough merely to know that identical twins were reared apart. How "apart," or different, were their environments? If two identical twins were separated during their first six months of life and raised in different families, their environments may nevertheless have been very similar. If both were raised in the families of, say, college professors, our traditions would lead us to expect that both twins had "good" homes and received much intellectual stimulation. If both twins were raised in lower middle-class homes, where both sets of foster parents had a high school education, we should again infer that these twins had fairly similar environments. If both twins were raised in isolated rural or urban slums by poor and uneducated people, without the books and conversation that we expect to foster intellectual development, then also they were not reared very "apart," even though they lived in houses separated by three thousand miles.

These ideas about the meaning of

N. L. Gage (2448, Stanford University Chapter) is professor of education and psychology at Stanford University and chairman of the Executive Board, Stanford Center for Research and Development in Teaching. He edited the classic Handbook of Research on Teaching *for the American Educational Research Association and was AERA president in 1963-64. His* Teacher Effectiveness and Teacher Education: The Search for a Scientific Basis *has just been published by Pacific Books, Palo Alto, Calif. Mr. Gage says, "For quick help in sending literature to me at Stanford in Germany, where I wrote this paper, I am very grateful to Arthur Jensen, Henry Kaiser, and Douglas Pidgeon. To Mr. Pidgeon I am also grateful for suggestions of ways to clarify some sentences and for reminding me of Bloom's work, to which I refer."*

"apart" are not new, but they have been underemphasized. Jensen devoted only 11 lines to the matter,[3] in describing the "most interesting" of the available studies (Cyril Burt's):

> ... most important, the separated twins were spread over the entire range of socioeconomic levels (based on classification in terms of the six socioeconomic categories of the English census), and there was a slight, though nonsignificant, negative correlation between the environmental ratings of the separated twin pairs. When the twin pairs were rated for differences in the cultural conditions of their rearing, these differences correlated .26 with the differences in their I.Q.'s. Differences in the material conditions of their homes correlated .16 with I.Q. differences. (The corresponding correlations for a measure of scholastic attainments were .74 and .37, respectively. The correlation between the twins in scholastic attainments was only .62, indicating a much lower heritability than for I.Q.)[4]

If we turn to Jensen's source,[5] we find that Jensen has chosen the lower of two correlations, i.e., the one for the individual intelligence test. The r for the group test is .43. But only a little more information becomes available in this source, despite the fact that Burt considered such correlations to be "the only satisfactory method" (p.149) to demonstrate the importance of environmental opportunities. In his similarly slight treatment (12 lines) of these data,

Burt wrote:

> For this purpose we have assessed the economic and cultural conditions of the homes in terms of a conventional scale similar to that employed for assessing intelligence and educational attainments, namely, one in which the mean is 100 and the standard deviation is 15. The correlations thus obtained are shown in Table 3 [our Table 1].
> It will be seen that differences in educational attainments are highly correlated with differences in cultural background. There is also a significant correlation between cultural differences and differences in the scores for the group test taken as they stand. But the correlations for the individual test and the final assessment are so low as to be nonsignificant with a sample of this size. The differences in educational attainments show a small but significant correlation with differences in material conditions, chiefly no doubt because the latter are responsible for differences in the children's physical health and school attendance.[6]

Several comments are warranted by this treatment of these data by Jensen and Burt, who was also a strong hereditarian. First, the treatments seem brief, almost grudging, in view of the admitted importance of the data and the pages devoted by these authors to other matters. No standard deviations or means of the environmental-difference variable were provided; thus it is impossible to compute a regression equation for pre-

dicting differences in I.Q. from differences in environment for identical twins reared apart. Second, no distributions or lists of environmental-difference data were given, although Jensen for his *Behavioral Genetics* article apparently obtained Burt's original I.Q. data. He provided lists of the I.Q.'s and various statistics based on them (mean, standard deviation, mean absolute difference between twins, standard deviation of the absolute difference, the intraclass correlation between twins, and the difference correlation between twins, which "indicates the degree of similarity between twins relative to the similarity between persons paired at random from the general population.")[7]

Third, Burt qualified the correlation of .43 between differences in group-test I.Q.'s and differences in cultural conditions by referring to lower correlations (.26 and .15) for individual-test I.Q.'s and his so-called "final assessment" I.Q., respectively. His suggestion seemed to be that the latter measures and the r's involving them are more valid than those based on group tests. But he gave no evidence to support such higher validity for the individual test and, despite much lore to the contrary, individual tests cannot be assumed without explicit evidence to be more valid. The "final assessment" was obtained after the group and individual tests had been given and their results had been "submitted to the teachers [of the children] for comment or criticism; and wherever any question arose, the child was reexamined."[8] Since Burt defined intelligence as "innate general cognitive ability," it is not inconceivable that teachers were unintentionally influenced by him to criticize and question I.Q.'s out of line with children's hereditary backgrounds rather than those discrepant with the cultural conditions in their homes. Thus the "final assessment" could readily, even if unintentionally, have been biased in such a way as to reduce its tendency to reflect a child's environment and increase its conformity to the child's hereditary background. Without better information as to the nature of the "final assessment," we cannot appraise its validity. Hence it seems questionable to intrude the I.Q. differences based on "final assessments" into the consideration of relationships between such differences and cultural environment differences.

Fourth, we should take care not to regard the school attainment data in Table 1 as irrelevant to the social policy issues to which I.Q. data apply. For it is

Table 1. Correlations Between Differences in I.Q. and Differences in Home Conditions for Identical Twins Reared Apart*

N=53 pairs

Test Results	Home Conditions	
Differences in Intelligence	Differences in Cultural Conditions	Differences in Material Conditions
Group test	.43	.21
Individual test	.26	.16
Final assessment	.15	.18
School attainments	.74	.37

Correlations over .29 are significantly different from zero (P<.05).

*Source: Cyril Burt, "The Genetic Determination of Differences in Intelligence: A Study of Homozygotic Twins Reared Together and Apart," *British Journal of Psychology*, 1966, p. 149.

school (and occupational) attainment with which society, parents, and educators are most directly concerned. I.Q.'s are important only insofar as they predict school and occupational attainment. We are concerned that blacks have average I.Q.'s substantially below the average I.Q.'s of whites not because I.Q.'s are important in their own right but only because they throw light on ability to succeed in schools and jobs. If environmental conditions of disadvantaged children can be arranged so as to improve school attainment, then such conditions are important regardless of their effect on I.Q. They provide a basis for hoping that the educational and occupational discrepancies between blacks and whites can be reduced and then eliminated. Hence we stress the great importance of the correlation of .74 between differences in school attainments and differences in cultural conditions in the homes of the separately reared identical twins. This r suggests that, even when heredity is held constant, substantial differences in school attainment can be produced by environmental differences. Along with the correlation of .43 for the group-test I.Q. differences, Burt's results indicate that, given environments different or "apart" enough, we can produce major differences in I.Q. or, even more important, school attainment, even among persons with the same genetic composition.

Unfortunately, we have no units for measuring environments that tell exactly what exists or happens in them. So we cannot tell from Burt's data (or those of a similar study discussed below) just how much environmental difference of what kind was associated with a given difference in the I.Q.'s of the identical twins. This neglect of technical detail in reporting on the environment measures can perhaps be remedied by further examination of Burt's legacy of data. If so, we might have a basis for determining whether the environmental differences that produced 10-point or greater I.Q. differences in 14 of Burt's 53 identical twins[9] resemble those that have produced such differences between American blacks and whites.

Let us turn now to the only other study, by Newman, Freeman, and Holzinger in 1937,[10] in which I.Q. differences and environmental differences were studied to reveal correlations for identical twins reared apart. This study yielded the data shown in Table 2. The scatterplot for the relationship between I.Q. differences and rated differences in educational advantage is shown in Figure 1. As the note for Table 2 indicates, the environmental-difference ratings were obtained from five judges who read the case material on each pair of twins.

It is clear that the I.Q. differences were larger for those twins whose estimated educational disadvantages differed more; *the I.Q. difference and the educational-advantage difference correlate .79.* The correlations between I.Q. difference and estimated social-advantage and physical-advantage differences were, respectively, .51 and .30. The fact that both these r's were lower than .79 suggests the special relevance of the educational-advantage ratings. When considered along with the results of Burt's study, shown in Table 1, where the correlations between I.Q. difference and "material-condition" difference were also smaller, the results indicate convincingly the association between differences in educational environment and differences in I.Q., even when genetic differences are held constant at zero.

Bloom carried the analysis of the data in Table 2 a step further:

We have divided the separated twins into two groups. For one group of 11 pairs, each pair of separated twins had very similar educational environments. The rank correlation for their I.Q. scores was +.91, whereas for the eight pairs that had the least similar educational environments, the rank correlation for their I.Q. scores was only +.24.[11]

Bloom's analysis explicates what is implicit in the data presented in Tables 1 and 2: The very high correlations, averaging .85,[12] between the I.Q.'s of identical twins reared apart result from

Table 2. Some Data from Identical Twins Reared Apart
(Newman, Freeman, Holzinger, 1937; Muller, 1925; Gardner & Newman, 1940; Saudek, 1934) *

Case Number	Sex	Age at Separation	Age at Testing	Environmental Differences			I.Q. Difference
				1. In Years of Schooling	2. In Estimated Education Advantages	3. In Estimated Social Advantages	
11	f	18 mo.	35	14	37	25	24
2	f	18 mo.	27	15	32	14	12
18	m	1 yr.	27	4	28	31	19
4	f	5 mo.	29	4	22	15	17
12	f	18 mo.	29	5	19	13	7
1	f	18 mo.	19	1	15	27	12
17	m	2 yr.	14	0	15	15	10
8	f	3 mo.	15	1	14	32	15
3	m	2 mo.	23	1	12	15	-2
14	f	6 mo.	39	0	12	15	-1
5	f	14 mo.	38	1	11	26	4
13	m	1 mo.	19	0	11	16	1
10	f	1 yr.	12	1	10	15	5
15	m	1 yr.	26	2	9	7	1
7	m	1 mo.	13	0	9	27	-1
19	f	6 yr.	41	0	9	14	-9
16	f	2 yr.	11	0	8	12	2
6	f	3 yr.	59	0	7	10	8
9	m	1 mo.	19	0	7	14	6
Muller	f	1 mo.	30	9	?	?	-1
Gardner & Newman	f	1 mo.	19	0	2	?	-3
Saudek	m	1 mo.	20	0	?	?	+4

Note: The estimated differences in educational and social advantages are in "points," with a maximum possible of 50. From the case material each of five judges rated the environmental differences between every pair of twins on a scale of 10 points, and the figure given in the table is the sum of these five ratings. A minus sign before an I.Q. difference means that the twin who received the higher rating for educational advantages obtained the lower I.Q.

*From R.S. Woodworth, *Heredity and Environment: A Critical Survey.* New York: Social Science Research Council, 1941.
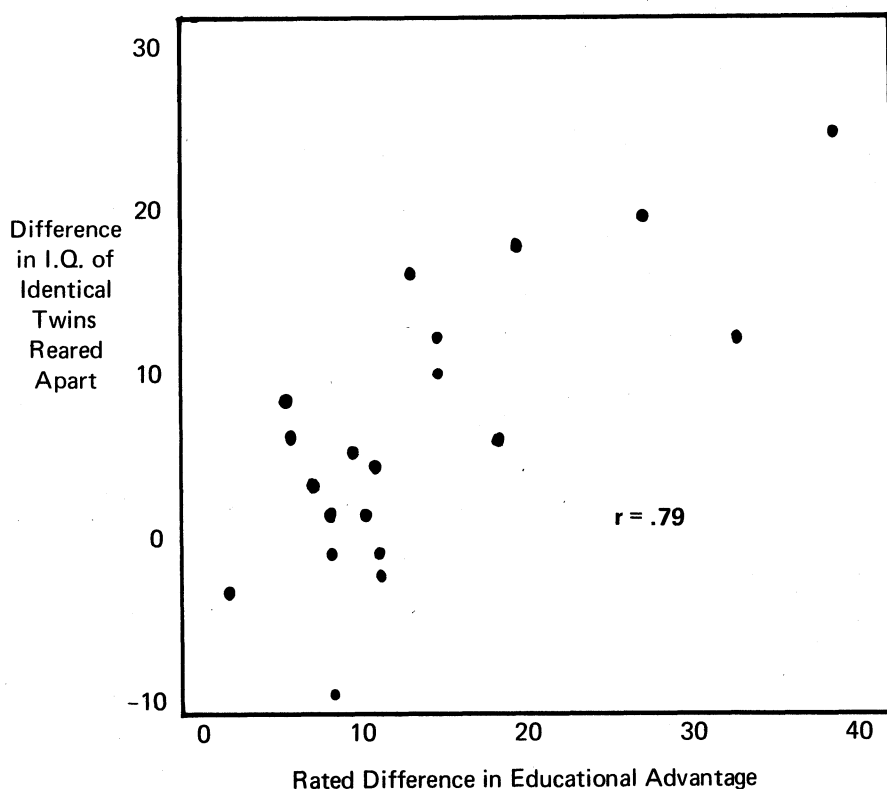
Figure 1. Scatterplot Showing Relationship Between Differences in I.Q. and Educational Advantage of 20 Identical Twins Reared Apart. (Based on H. H. Newman, F. N. Freeman, and K. J. Holzinger, *Twins: A Study of Heredity and Environment.* Chicago: University of Chicago Press, 1937.)

the similarity of the environments in which the twins were reared. As the similarity becomes lower, the correlation becomes lower.

Jensen has shown that the absolute differences in I.Q. of identical twins reared apart fall into a distribution that "closely approximates the chi distribution,"[13] which would result from taking the absolute difference between pairs of values drawn at random from a normal distribution. This means in turn that environmental effects on the I.Q., as represented by co-twin differences, are normally distributed. Since all of the 122 pairs of identical twins reared apart that were involved in the four studies reviewed by Jensen were drawn from English, Danish, and North American Caucasian populations, this finding seems reasonable. At least, there are no obvious historical or sociological factors that would interfere with the operation of the large number of independent, relatively equally influential factors that make any set of data, including these environmental differences, fall into a normal distribution.

But such a normal distribution is not tantamount to a finding that the I.Q. differences are merely "chance" or "uncaused," attributable entirely to errors

of measurement. For Table 1 and Figure 1 show that the I.Q. differences are obviously associated with environment differences. Thus it is difficult to see the point of Jensen's citing the normality of distribution in relation to the frequently cited case of Gladys (I.Q. 92) and Helen (I.Q. 116) in the study by Newman *et al.* (This case is represented in Figure 1 by the dot in the upper right-hand corner.) These twins "had markedly different health histories as children. . . . Gladys did not go beyond third grade in school, while Helen obtained a B.A. degree from a good college. . . ."[14] Since these are exactly the kinds of health histories and educational opportunities that comprise environmental differences, especially those between Negroes and whites, they should be regarded as supporting an environmentalist view of racial differences in I.Q. The Gladys-Helen I.Q. difference is rare; the Gladys-Helen environmental difference is also rare. To have both rare events occurring *together* is evidence not of randomness but of strong nonchance association.

Relevance to Race Differences

How is all of the foregoing related to

the problem of interpreting the 15-point difference between the mean I.Q.'s of black and white Americans? The heritability of I.Q. among whites is one thing, but extending the conclusions to Negroes in the United States is another. Herrnstein put it this way: "Although there are scraps of evidence for a genetic component in the black-white difference, the overwhelming case is for believing that American blacks have been at an environmental disadvantage."[15]

We have seen that, given large enough environmental differences, identical twins reared apart can also exhibit I.Q. differences of 15 points or more. The question thus becomes, Are the environmental differences between Negroes and whites in the United States also large enough to produce average I.Q. differences of about 15 points?

The answer is that no one knows. We have not been given the necessary information about the way in which environment differences were measured in the studies by Newman *et al.* and Burt. Even if we had much more detailed information of this kind, we would then need to apply the same or comparable techniques to the measurement of the environments of representative samples of Negroes and whites in the United States.

Suppose such measurements revealed differences with magnitudes like those experienced by the identical twins represented by dots in the upper right-hand corner of Figure 1 or by dots of the same kind in scatterplots based on Burt's data. Then we would have evidence that Negroes and whites would differ in I.Q. about as much as they do even if they had identical kinds of I.Q.-determining genetic compositions.

Although we lack the data necessary for environment comparisons of this kind, we can offer some reasonable conjectures based on the history of the two races, especially the blacks, in the United States. American history since 1700 has designed and executed a massive experiment in which radical manipulations of the environment constituted the experimental treatment. One substantial fraction of the population was enslaved, literally, not figuratively. Then, after being freed, it was subjected to an elaborate, pervasive, systematic, and rigorously enforced set of social, political, economic, and educational discriminations. The treatment operated so as to impair the fabric of that fraction's familial and educational life. The experimental group was deprived of books and access to opportunities to hear standard

English. Its workers were kept so physically tired by hard labor that they seldom could find energy for self-educative activities demanding intellectual effort. The experimental fraction was insulted, impoverished, made fearful, and instilled with self-hatred. In short, it would be difficult for psychologists, using what research has yielded concerning factors affecting cognitive functioning and development, to plan an environment better designed to harm the average intelligence of an experimental group consisting of about a 10% sample of the nation's population.

Unfortunately for the validity of the experiment, its design had a basic flaw: The subjects were not randomly assigned to the alternative treatments. Instead, the experimental treatment was confounded with the variable of race. All members of the experimental group were Negroes. Hence it has been impossible to determine, at least on the basis of the logic of experimental design, whether the resulting differences in the numerous dependent variables, including I.Q., should be attributed to the treatment or to the race of the subjects.

Inasmuch as the experiment was flawed, it would seem to be bad research work to compound that error with another. If we cannot be sure that the educational and economic inequalities of Negroes result from the grievous experimental treatment to which they have been subjected, should we leap to the conclusion that it was their genetic makeup? For it should be recalled that all of the research on identical twins reared apart, and almost all of that on the other relationship and rearing combinations, has been done with white subjects only. Hence that research has dealt with only the environmental variances to which whites are subjected. We have no way of being sure that those environmental variances have been large enough to embrace, at the low end of the scale, the environments to which Negroes have been subjected. It thus becomes dubious in the extreme to conclude that the I.Q. heritability values found for whites only would be found for whites and blacks together. And it becomes correspondingly dubious, therefore, to conclude that Negroes cannot be helped, through better education, to achieve educational and economic equality.

Yet such a conclusion seems to be implicit in suggestions that eugenic or birth-deterrent measures should be taken to reduce these inequalities. Persons with low I.Q.'s are born ineducable and unemployable, it is reasoned, and hence become burdens to themselves and society. So it is better to keep them from being born. Just as couples likely to produce physically nonviable or seriously defective children are cautioned against having them, so persons with low I.Q.'s should be discouraged, perhaps with monetary compensations, from having children.

How reply to such a proposal? We simply do not know enough, and are much too unsure about what knowledge we do have, to establish an adequate scientific and moral basis for such a social policy. But we do know that I.Q. is not everything, in our society or any other. Other human qualities, as Michael Young's *Rise of the Meritocracy* made clear, are equally valuable to us and our fellow men. There are many special intellectual abilities, artistic and musical talents, and types of creativity that are missed by I.Q. tests.

Also, the I.Q.'s of children are not predictable enough from those of their parents to justify preventing anyone other than the most defective from having children. And such proposals overlook the regression effect which occurs when correlations are less than perfect. Such an effect makes extreme values of one variable go along with less extreme values of another variable imperfectly correlated with it. Regression effects, relating both the empirical and the statistical facts inherent in correlations of .5 between the I.Q.'s of parent and children,[16] make parents with I.Q.'s below 100 likely to have children with I.Q.'s closer to 100.

Finally, it seems likely that such a policy would drain resources away from educational and other efforts aimed at improving environmental influences on educational achievement and employability.

Better Educational R & D Needed

Educational and other environment-improvement approaches to the problems of racial inequality have not been given anything close to an adequate trial thus far in the United States. The doubts often expressed concerning the outcomes of Project Head Start are based on disregard of the definite evidence of positive results in substantial subsamples.[17] The pessimists also forget that the Head Start projects evaluated were begun in a hurry and so were inadequately planned. They were handicapped by makeshift staff and curriculum materials.

What would come from adequately planned and stably supported projects? If based on systematically researched experimental alternatives, such projects would produce, in the opinion of many behavioral scientists, solid and worthwhile improvements in the educational achievement of children from low-income homes. It is *not* true that "compensatory education has been tried and it apparently has failed." Compensatory education needs more research and better-supported tryouts over a period of decades, not merely a single enthusiastic Presidential administration. Its support needs to be made independent of partisan politics, at least as much as cancer and weapons research have been freed in this way. It needs better-trained and larger research forces, competent in the best methods and theories of the behavioral and, where relevant, the biological and physical sciences. It needs research not just on what and how to teach but on how to train teachers and the administrators of their schools. In short, the effort to eliminate racial inequality in American society needs better education and the research to make that better education possible.

---

1. Arthur R. Jensen, "How Much Can We Boost I.Q. and Scholastic Achievement?," *Harvard Educational Review*, Winter, 1969, pp. 1-123.

2. Arthur R. Jensen, "I.Q.'s of Identical Twins Reared Apart," *Behavioral Genetics*, 1970, pp. 133-46.

3. Jensen, *op. cit.*, 1969.

4. *Ibid.*, p. 52.

5. Cyril Burt, "The Genetic Determination of Differences in Intelligence: A Study Of Monozygotic Twins Reared Together and Apart," *British Journal of Psychology*, 1966, pp. 137-53.

6. Burt, *op. cit.*, p. 149.

7. Jensen, *op. cit.*, 1970, p. 140.

8. Burt, *op. cit.*, p. 140.

9. Jensen, *op. cit.*, 1970, p. 14.

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

11. Benjamin S. Bloom, *Stability and Change in Human Characteristics.* New York: Wiley, 1964, p. 70n.

12. Jensen, *op. cit.*, 1970, p. 137.

13. *Ibid.*, p. 141.

14. Jensen, *op. cit.*, 1970, p. 142.

15. Richard Herrnstein, "I.Q.," *The Atlantic*, October, 1971, p. 57.

16. L. Erlenmeyer-Kimling and L. F. Jarvik, "Genetics and Intelligence: A Review," *Science*, 1963, pp. 1477-79.

17. Westinghouse Learning Corporation and Ohio University, "The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development." Presented to the Office of Economic Opportunity, Washington, D.C., 1969.   ☐