INTELLIGENCE 5, 29-40 (1981)

Raising the IQ: The Ramey and Haskins Study

ARTHUR R. JENSEN
University of California, Berkeley

The Ramey and Haskins intervention experiment succeeded in producing IQ gains at three years of age averaging about one standard deviation in young children who were selected for being at risk for subnormal intellectual development. The study is examined in terms of its consistency with other findings, the heritability of IQ, the *g* aspect of IQ, the simplex pattern of longitudinal interage mental test score correlations, mother-child IQ correlations, and criteria for establishing educationally and socially significant gains in intelligence defined as *g* rather than as a score on a particular test. Narrow transfer of training from cognitive intervention techniques to IQ test performance in early childhood, rather than enhancement of the *g* factor itself, is hypothesized as a cause of the typical fadeout of early IQ gains in later childhood.

Despite more than half a century of repeated efforts by psychologists to improve the intelligence of children, particularly those in the lower quarter of the IQ distribution relative to those in the upper half of the distribution, strong evidence is still lacking as to whether or to what extent it can be done. Probably no other topic in the whole history of psychology has commanded such vast funds for research, especially in the past 20 years. Are the intelligence gains attained through experimental treatments as stable over time as nontreated IQ? Are the gains that may be statistically significant also of sufficient permanence and magnitude as to be individually or socially important? The evidence from those assiduous efforts, as best as I can see it today, warrants only the most cautious and tentative optimism concerning the capability of psychologists and educators permanently to raise the intelligence of humans by any nonbiologic techniques.

In order to think more clearly about the issues, two important distinctions should be kept in mind.

First, there is the distinction between the IQ or any specific test-score measurement of intelligence, on the one hand, and the general intelligence factor *g*, on the other. Performance on specific tests, and on some types of tests more than others, is more amenable to alteration through experimental treatments than is *g*, the general factor common to a wide diversity of cognitive tasks. It is badly misleading to view individual differences in

intelligence as consisting only of differences in the various specific items of knowledge and skill that comprise the contents of any particular IQ test. These content-specific features of tests are merely *vehicles* for the measurement of *g*. Performance on the vehicles is undoubtedly trainable. All organisms possessing a nervous system are capable of learning. But what is learned about the specific vehicles for measuring *g* does not itself constitute *g*. When one speaks of raising IQ, however, the implication (and hope) is that it is *g* which is being raised and not just performance on a particular test or others much like it. Intelligence test scores are really important only because of their many educationally, occupationally, and socially important *correlates*, and these are largely a result of the *g* factor in all manifestations of mental ability. There would be no point in raising intelligence if all it resulted in was higher scores on IQ tests. The real hope is that it would result in a higher level of performance on all the "real life" *correlates* of IQ, as would be the theoretically expected consequence of raising *g* itself.

Second, we should take note of the distinction between intelligence gains that are significant for the individual and gain that is significant for a population. A relatively small IQ gain, say, 5 IQ points, would not be of much tangible consequence to an individual and may hardly seem worth creating the special conditions that may be needed to bring it about. I wouldn't give five dollars to have five more points added to my IQ, whatever it may be. But, an average gain of even two or three IQ points (assuming, of course, it represents a gain in *g*) can be of great social consequence to a whole population, at least if the entire distribution is moved up the scale. Because of the normal distribution of mental ability, of course, a slight change in the population mean has marked effects on the proportions of the population that fall above a given high-cutting score on the IQ scale or below a given low-cutting score. Although a five-point IQ gain may be meaningless to an individual, in population terms a mean gain of five points would double the percentage of persons with IQs over 130 and would cut in half the percentage with IQs below 70. The educational, social, and economic consequences of such a change for a population could be tremendous. So, from a population standpoint, no one should belittle the potential importance of even a quite small IQ gain, provided that there is good reason to believe it is a permanent change in *g* rather than a short-lived enhancement of specific test knowledge and skills.

TYPICAL FINDINGS

There is now considerable concensus among workers in this field as to the typical findings from experimental attempts to raise intelligence. I am here excluding reports of the amelioration of abnormal developmental deficits in rare, freakish cases of extreme social isolation and of deplorable neglect in orphanages. Most studies that have tried to raise IQs have focused on

children from poor homes or on those whose IQs in later childhood are statistically predicted, on the basis of certain socioeconomic, racial, and parental characteristics, to fall into the lower half of the IQ distribution.

It is found that the IQs of younger children (preschoolers) are more malleable than of older children. Those programs that begin intervention earliest (usually in infancy) and last longest (up to school age or beyond), have produced the largest gains.

IQ gains appear most marked during those early years that the kinds of test items used for assessing IQ gains would seem to allow for the most direct transfer of specific learning from the cognitive materials and intervention procedures that are applied to the experimental subjects.

Virtually without exception there is a partial or total fadeout of treatment-induced IQ gains. In later childhood, experimental subjects, who were initially accelerated, generally gravitate toward their normally expected level, as shown by a control group.

The effects of the experimental treatment on important cognitive correlates of the IQ are less pronounced and fade more rapidly, in some cases without leaving any residual trace of the treatment. Probably the most intensive and prolonged intervention study ever attempted, extending from early infancy to school age, is the highly publicized study of high-risk Black ghetto children in Milwaukee by Heber and Garber (1973). The early average gain in IQ of the experimental over the control group was so impressive -- close to 30 IQ points – that Ellis Page (1972) referred to it as the "Miracle in Milwaukee." The children are now well along in elementary school, and at the latest report (Clarke & Clarke, 1979, pp. 224–225) the mean IQ has declined some 20 points. The experimental group still has a higher IQ than the control group, but there is no difference between the experimental and control subjects in reading ability, which is probably most crucial for scholastic success. In the general school population, reading comprehension is highly *g* loaded. We have found that it is more highly correlated with IQ than any other area of scholastic performance included in the complete battery of Stanford Achievement Tests. That the extraordinary intervention provided in the Milwaukee project so strikingly raised IQs without showing any residual effect on reading achievement leads me to suspect that the cognitive skills inculcated by the treatment program displayed only the relatively narrow transfer typical of trained skills, rather than the broadly general cognitive ability that characterizes *g*.

IMPLICATIONS OF HERITABILITY ANALYSIS

In case there is still any confusion about the implications of the heritability of intelligence for interventions intended to raise intelligence, or vice versa, I should sketch out precisely what these implications are and the logic on which they are based. The two most common misconceptions are (1) that a high

heritability of intelligence absolutely precludes the possibility of altering intelligence appreciably by *any* kind of environmental intervention; (2) that successful attempts to raise intelligence through environmental intervention is evidence against the heritability of intelligence.

In quantitive genetics, broad heritability (h^2_B) is defined as the proportion of the total variance in phenotypes (e.g., IQ scores) attributable to variation in genotypes, i.e., all the genetic factors that influence the development of the phenotypes. This can be expressed as $h^2_B = \sigma^2_G / \sigma^2_P$, where σ^2_G and σ^2_P are the variances of genotypes and phenotypes, respectively. Broad heritability can also be thought of as the squared correlation between genotype and phenotype.

There is now a consensus among most experts that the heritability of intelligence is substantial. Strangely, the popular media and even a few psychologists have tried to create the impression that the substantial heritability of IQ is a far-out or radical opinion held by only a few offbeat scientists in the face of the overwhelming opposition of their colleagues. Nothing could be further from the truth. The overwhelming majority of textbooks in psychology and genetics that I have surveyed on this topic agree that genetic factors are predominant in IQ variance. The typical position is expressed in the latest *Encyclopedia Britannica* (1975, Vol. 8, p. 1148):

> Concerning the extent of genetic determination in human intelligence, most investigations have yielded heritability estimates between 70-80 percent. Since such values are relative to the populations studied and to the method of estimation, some disagreement should be expected. It seems most unlikely, however, that genotype contributes less than 50 percent of the variability and it is conceivable that the figure is closer to 80 percent.

The proportion of nongenetic or environmental variance in a trait is the reliability of the trait measurement minus the broad heritability, i.e., r_{xx}-h^2_B. If the broad heritability of IQ is .70 and test reliability is .95, and if the σ of IQ in the population is 15, the standard deviation of environmental effects on IQ would be $\sigma(r_{xx}\text{-}h^2_B)^{1/2}$, or $15(.95-.70)^{1/2} = 7.5$ IQ points. This means that a shift of one standard deviation in the sum total of the environmental factors that influence IQ should alter the IQ by 7.5 points.

The more important point, however, is that the estimate of the nongenetic variance used in making this calculation refers only to those environmental factors, whatever they may be, that actually contribute to IQ variance in the particular population at that time. It does not include all *possible,* but presently nonexistent, environmental influences on IQ, and it does not include all *possible,* but presently nonexistent, combinations or interactions of presently existing environmental influences. In other words, r_{xx}-h^2_B reflects only the variance resulting from the actually existing environmental variation in the population. Environmental interventions that consist only of

reallocating the existing environmental variation to different individuals will have their effects limited by the heritability of trait. But even with heritability in the range of .70 to .80, the magnitude of environmental effects can be considerable. With a standard deviation of 7.5 IQ points, for example, and assuming that existing environmental effects on IQ are normally distributed (for which there is good evidence), the total range of environmental influences would be about six σ, or 45 IQ points. Intervention that produces IQ changes within that range is not in the least incompatible with present estimates of the broad heritability of IQ.

The real problem, however, has been in bringing the environmental influences on IQ under experimental control. Even though evidence on the genetic analysis of IQ leaves considerable latitude for nongenetic influences, psychologists have not yet discovered more than a fraction of the nongenetic factors that contribute to IQ variance or how they can be experimentally harnessed to raise IQ markedly and permanently. Although it may come as a surprise to many psychologists, at present, we know more about the genetics of IQ than we know about environmental influences on IQ, except for extreme deprivations and traumas that are too rare to contribute importantly to the IQ variance of the general population. My hunch is that the nongenetic variance in IQ is the result of such a myriad of microenvironmental events as to make it extremely difficult, if not impossible, to bring more than a small fraction of these influences under experimental control. The results of all such attempts to date would seem to be consistent with this interpretation.

One other empirical finding regarding the heritability of IQ is highly relevant to early intervention studies. For IQ, as for height and weight, the correlation between genotype and phenotype is only slight in infancy and early childhood, but increases gradually as the child grows up. The phenotypic expression of the genotype steadily increases with increasing age, up to maturity. This phenomenon, probably more than anything else, is responsible for the typical "fading out" of the effects of early childhood environmental interventions, as the children approach puberty.

Finally, regarding the heritability of IQ, I wish to point out what I now consider the impropriety of any longer citing the purported empirical studies of the late Sir Cyril Burt, particularly his reported estimates of IQ heritability based on monozygotic twins reared apart. I have elsewhere (Jensen, 1974a) pointed out apparent errors and inconsistencies in Burt's empirical papers. More recently, Burt's biographer has assembled compelling evidence from personal documents that much of Burt's data, and certainly that on separated MZ twins, was fraudulent (Hearnshaw, 1979). As damaging as these sad revelations are to Burt's once great reputation, however, they do not importantly affect any of our conclusions about the heritability of IQ, which are now based on much other evidence.

THE RAMEY AND HASKINS STUDY

Ramey and Haskins (1980) have shown that a group of infants at risk for subnormal intellectual growth could be brought, through intensive daycare environmental intervention designed to stimulate intellectual growth, up to a mean Stanford-Binet IQ of 95.15, SD = 14.42, at 36 months of age, as compared to a proper control group, with a mean IQ of 80.60, SD = 14.87. The IQ difference of approximately one standard deviation between the experimental (E) and control (C) groups is highly significant (t = 3.51, df = 50, $p < .001$). A significant E–C mean difference of 8.62 S–B IQ points was also found at 24 months of age. At ages 6, 9, 12, and 18 months the E and C groups differed only slightly and inconsistently on the Bayley index of Psychomotor Development and on the index of Mental Development, except at 18 months, when the C group mean for the first time falls significantly below the E group mean by 13.34 points or a difference of 0.90σ.

It is an important fact, in this experiment, that the C group as well as the E group both received social work services, and pediatric care; most important, both were provided with adequate nutrition throughout the course of the study. The significant test score differences between the E and C groups at 18 months, 24 months, and 36 months, therefore, can be more confidently attributed to the special daycare treatment.

It is expecially noteworthy that nutritional supplements did not prevent an IQ deficit in the C group, whose mean IQ of 80.60 at 36 months is about the same as their mothers' mean IQ of 81.36. It seems doubtful, therefore, that the IQ deficit in this population is significantly attributable to nutritional factors.

Finally, it should be noted that Ramey and Haskins employed about 300 different curriculum activities intended to enhance cognitive development. One wonders to what extent some of these activities resemble the *vehicles* for the measurement of g employed in the Bayley and Stanford-Binet scales. The Year II tests of the Stanford-Binet consist of three-hole form board, delayed response to a small object hidden under one of three boxes, identifying parts of the body on a paper doll, building a tower of four blocks, and picture vocabulary of common objects. The Year III tests include stringing beads, building a "bridge" with three blocks, copying a circle, and drawing a vertical line. It is hard to imagine how a preschool program aimed at stimulating cognitive development could avoid providing practice in skills that, although perhaps not identical to those in the Stanford-Binet, would not result in a narrow transfer-of-training enhancement of Stanford-Binet test performance. One would like to see the results on other types of g-loaded tests with quite different item content.

I am presently developing a number of very simple reaction time "tests" which are g loaded, but involve such minimal practice effects that performance levels can hardly be ascribed to transfer effects from similar prior experiences (Jensen, 1979). The same thing is true of the demonstrable g

component of average evoked brain potentials (Hendrickson & Hendrickson, 1980). These types of measurements could be useful in intervention studies, for determining whether the enhancement of tested IQ is paralleled by changes in the reaction time and evoked potential correlates of the *g* component of IQ that we find in untreated samples. The usefulness of these techniques, however, will depend on some preliminary parametric studies of these types of measurements on preschool children. But this has not yet been done.

CUMULATIVE DEFICIT AND THE DISCONTINUITY OF SCALES

Ramey and Haskins refer to the phenomenon of "cumulative deficit," i.e., the gradual decline of IQ throughout the school years in children from poor backgrounds. One way to detect this effect is by comparing the IQs of younger and older siblings. Because the expected genotypic values of full siblings are completely independent of birth order, any average IQ difference between younger and older siblings, assuming proper age standardization of the test, must be environmental. "Cumulative deficit" would show up as a lower mean IQ of older as compared with younger siblings. Using this method on all of the full siblings of elementary school age in an entire California school district, I found no evidence of cumulative deficit in verbal or nonverbal IQ in the white population and a very slight, but statistically significant, cumulative deficit in verbal IQ, but not in nonverbal IQ, in the Black population—although verbal IQ was higher than nonverbal IQ in the Black population—(Jensen, 1974b). When the same method was later applied to whites and Blacks in the rural South, however, whites showed no sign of cumulative deficit in either verbal or nonverbal IQ, but Blacks showed a significant and substantial decline in both verbal and nonverbal IQ, amounting to about one IQ point per year between the ages of 5 and 16 (Jensen, 1977). The Blacks in that community were about as disadvantaged as any that could be found anywhere in the United States. Thus decline in the IQs of severely disadvantaged Black children throughout the school years is an authentic phenomenon.

But is it the same phenomenon that we see in the much younger group in the Ramey and Haskins study? Their samples, too, consist entirely of Black children from very poor homes. And in their control group, we see a rather sharp decline in scores beginning at 18 months on the Bayley Mental Development index. At 24 months of age, the control group's mean Stanford-Binet IQ is 83.96. Bayley (1965) has shown that large representative samples of Black infants do not differ, on the average, from representative samples of white infants, in scores on the Bayley Mental Development index before 15 months of age. Black infants significantly outperform whites on the Psychomotor Development index. Bayley expresses the belief that Black

infants are innately precocious in psychomotor development (Bayley, 1965, pp. 408–409). As the factorial composition of the infant tests changes at higher age levels, becoming more cognitive, particularly with the transition to the Stanford-Binet at 2 years, the Black children's raw scores show a lessened rate of increase relative to the white norms. This seeming decline of psychomotor precocity by Black infants as the test items become more cognitively complex appears to me to be a quite different developmental phenomenon than the cumulative deficit seen in the IQs of school-age Black children, where the IQ tests at every age level are highly *g*-loaded and possess quite high predictive validity for later IQ.

A very large common factor is measured by IQ tests given at all ages beyond age 5, as can be shown by extracting the first principal factor from the matrix of correlations between test scores obtained on the same individuals at different ages from 5 to 18 years. The infant tests below 2 years of age, however, are not highly loaded on the general factor common to IQ tests obtained in later childhood. Infant tests before age 2 or 3 years measure something largely different from the *g* factor that constitutes the largest part of the variance in the IQs of older children. According to Bloom's (1964) analysis of the relevant evidence, mental test scores at age 2 to 3 predict only about 10 to 20 percent of the IQ variance in the late teens. This predictive validity is for samples of children who have not been subjected to any experimental treatment designed to accelerate mental development. I do not know of any evidence on the validity of early test scores for predicting later IQ in children who have been given special treatment to enhance their cognitive development prior to obtaining the early test scores. My hunch is that the early scores would have even lower validity for predicting later IQ than is found in untreated children. This hunch is based on my suspicion that the observed enhancement of early IQ by intervention programs is not an enhancement of the *g* factor that largely predicts later IQ, but of directly teachable test-specific skills that transfer rather narrowly to other forms of cognitive tasks.

The later residue of early intervention of the kind described by Ramey and Haskins might well consist more of improving the noncognitive aspects of social development, attitudes, and adjustment to school than of heightened IQs or academic achievements per se.

INTER-AGE CORRELATIONS

There is a good deal of evidence that the matrix of intercorrelations among mental test scores obtained at various ages in longitudinal studies conforms quite closely to a Guttman simplex (Jensen, 1973, pp. 79–93). That is, the intercorrelations for temporally adjacent tests are higher, and there is a regular decrease in the correlations as the interval between the tests increases. This distinctive pattern of inter-age correlations during the period of growth

holds for physical measurements, such as height and weight, as well as for mental measurements and scholastic achievement. Also, the intercorrelations among the Bayley Mental Development scores obtained in a large sample ($N > 200$) at 6, 9, 12, 18, and 24 months conforms perfectly to the simplex pattern that generally characterizes growth measurements (Wilson & Harping, 1972).

It should be interesting to see whether the typical simplex pattern shows up in the Ramey and Haskins data. They have provided the inter-age correlations for the E and C groups, as shown in Table 1. The standard error of the correlation coefficient is quite large for relatively small samples, and therefore very close conformity to the simplex pattern can hardly be expected from these data. Nevertheless, the E group's correlations do, in fact, fit the pattern quite well, which means that the experimental treatment has not seriously distorted the typical correlations among the tests at various ages. Surprisingly, it is the C group that shows the less typical pattern of intercorrelations, which jointly conforms to a simplex only by virtue of its first-order off-diagonal correlations being larger than the rest. The E and C matrices are not entirely dissimilar, however; the correlation between the 15 pairs of E and C correlations is .70. The mean r for the E group is .53 as compared with .39 for the C group. Thus it appears that the experimental treatment has resulted in greater temporal consistency of individual differences in mental growth rates. But this is only suggestive. I would not place too much emphasis on it, because a statistical test of significance of the overall difference between the E and C correlation matrices shows that the difference falls far short of significance. (A chi squared test of the difference between two correlation matrices is described by Jensen, 1980, pp. 449–450.)

TABLE 1
Inter-age Correlations of Bayley MDI and Binet IQ Scores for Experimental (above diagonal) and Control (below diagonal) Groups

Test and Age	MDI at Age				Binet at Age	
	6	9	12	18	24	36
MDI 6		.63	.66	.49	.41	.39
MDI 9	.28		.58	.24	.19	.38
MDI 12	.53	.46		.61	.52	.49
MDI 18	.20	-.01	.42		.78	.65
Binet 24	.42	.31	.41	.50		.64
Binet 36	.17	.24	.36	.56	.74	

Experimental Group (above diagonal), $N = 26$: $r > .381$, $p < .05$. $r > .487$, $p < .01$.
Control Group (below diagonal), $N = 25$: $r > .388$, $p < .05$. $r > .496$, $p < .01$.

MOTHER-CHILD CORRELATION

The correlation between parent and offspring (r_{PO}) is complexly determined. It is a function of the narrow heritability (h^2_N) of the trait and the degree of assortative mating for the trait, as indexed by the correlation between parents (r_{PP}). The narrow heritability is that proportion of the total genetic variance that "breeds true" and accounts for the resemblance between parents and offspring. (Broad heritability, on the other hand, includes variance due to genetic dominance and other gene interactions which cause offspring to differ from their parents.) Using the symbols just explained, in quantitative genetics the (single) parent-(single) offspring correlation is $r_{PO} = 1/2\, h^2_N(1 + r_{PP})$. Empirical estimates of h^2_N are around .60, and estimates of r_{PP} are about .45 (Jensen, 1978), which, in the preceding equation, yields a parent-offspring correlation of .435. Empirically determined parent-child correlations average close to .50, however, because of common environmental correlation between parents and children. Adopted children's IQ shows correlations of about .30 to .40 with their biological mother's IQ. The parent-child correlation, however, is generally lower in early childhood and gradually increases to .40 or .50 in the early teens. By their late teens, the correlation of adopted children's IQs with the intelligence levels of their adoptive parents is practically nil. Scarr (1980) reports that in a study of children adopted at birth only 1 percent of their IQ variance in adolescence can be accounted for by a multiple regression equation which includes both the adoptive parents' IQs, the adoptive parents' education, the adoptive father's occupation, and the adoptive family's income.

The mother-child correlations (at 36 months) in the Ramey and Haskins study are .43 for the C group and -.05 for the E group. The correlation of .43 is well within the range of mother-child correlations reported in the literature. Again, because of the small samples, we cannot put too much emphasis on the difference between the correlations of .43 and -.05, which is not statistically significant ($z = 1.73$; $p = .084$).

The mother-child correlation of -.05 seems puzzling, however, in view of those studies of children adopted soon after birth that show their later IQs to correlate .30 to .40 with their biological mothers' IQs, even though the children have had no postnatal contacts with their biological parents. I have checked the present data for the usual statistical features that can attenuate correlations. The standard deviations of the E and C mothers' IQs are 11.57 and 12.15, a nonsignificant difference. The standard deviations of the E and C children's IQs (at 36 months) are 14.42 and 14.87. (It is interesting to note that the presumably quite uniform experimental treatment has not resulted in a decrease in the variability of IQ in the E group compared to the C group.) So the E vs. C difference in correlations cannot be ascribed to a restricted range of talent in the E group mothers or children. To see whether the effects of outliers on the Pearson *r* could be responsible, I computed the Spearman rank

order correlation between mother-child in the E group; it is -.01. (The mother -child rank order correlation for IQ at 24 months is -.13.) Moreover, these low correlations are unlikely to be due to the insufficient reliability of the children's Binet IQs, which show a rank order correlation of .58 (and a Pearson *r* of .64) between ages 24 and 36 months.

The near-zero mother-child IQ correlation in the E group could be interpreted as seriously undermining the construct validity of the Binet IQ as a measure of intelligence (that is, *g*) in the E group. A substantial parent-child correlation has become an important aspect of the construct validity of the IQ. A failure to demonstrate a significant parent-child correlation, especially when there is considerable uniformity of environmental treatment for all the children, raises the question of whether the IQ test is actually a measure of the general intelligence factor. The E group's acquisition of certain overlearned skills which transfer to easy Binet test items may depend on what I term Level I, or rote-learning ability, that is largely independent of Level II ability, or *g*. At least, in trying to understand the apparent wiping out of the mother-child correlation by the experimental treatment, it seems to me that the most obvious hypothesis is that the treatment has altered the factor composition of the test such that it reflects more of some other factors than of the *g* factor. But it is the *g* factor that is mainly responsible for the great importance accorded to the IQ. IQ test scores that have been largely divested of their *g* loading and therefore represent mostly task-specific variance will not, I suspect, predict the educationally, occupationally, and socially important criteria that are the only point of trying to raise a child's IQ. It is these correlates of IQ that are important, of course, and not particular test scores themselves. And it is largely because of that part of the IQ test score variance that can be identififed as *g* that the IQ correlates with so many other important variables and gains its construct validity as a measure of general intelligence.

Any intervention program that aims to enhance cognitive development and bases its claims of success of on a mean IQ difference between E and C groups must sooner or later come to grips with the problem of demonstrating that the E group's heightened IQs still have the same meaning as an index of general intelligence, with all the implied correlates of IQ, that has long been established for untreated groups of children. Fortunately, the experiment by Ramey and Haskins is still under way, so that future reports of its progress may eventually be able to provide evidence on this crucial point.

CONCLUSION

The Ramey and Haskins experiment, like other studies of early intervention to enhance cognitive development, shows singificant and substantial IQ gains at three years of age. The practical significance of the gain in IQ can be determined only by follow-up studies designed to assess the

resistance of the early gains to fadeout in later childhood, as has been the fate of previous studies, and to demonstrate that the experimentally induced IQ gains in fact reflect the broad generality of influence on intellectual performance and achievements that lend the IQ its construct validity as a measure of *g* or general ability.

It may seem surprising that these desired outcomes have not yet been conclusively demonstrated in the past two decades of large-scale, well-supported efforts to experimentally accelerate children's intellectual development. Without such a demonstration eventually, future historians of psychology may well liken this period to the era of alchemy in the history of chemistry. The analogy, however, is not completely negative. Although the alchemists failed in their primary aim—to find the "philosopher's stone" that could transmute base metals into gold—their experiments nevertheless contributed to the scientific advancement of chemistry.

REFERENCES

Bayley, N. Comparisons of mental and motor test scores for ages 1–15 months by sex, birth order, race, geographical location, and education of parents. *Child Development,* 1965, *36,* 379–411.

Bloom, B. S. *Stability and change in human characteristics.* New York: Wiley, 1964.

Clarke, A. M., & Clarke, A. D. B. *Early experience: Myth and evidence.* New York: Free Press, 1979.

Hearnshaw, L. S. *Cyril Burt, psychologist.* Ithaca, N.Y.: Cornell University Press, 1979.

Heber, R., & Garber, H. The Milwaukee project: A study of the use of family intervention to prevent cultural-familial mental retardation. In B. Z. Frielander, G. M. Sterritt, & G. E. Kirk (Eds.), *Exceptional infant* (Vol. 3). New York: Brunner/Mazel, 1979.

Hendrickson, D. E., & Hendrickson, A. E. The biological basis of individual difference in intelligence. *Journal of Personality and Individual Differences,* 1980, 1, 3–33.

Jensen, A. R. *Educability and group differences.* New York: Harper & Row, 1973.

Jensen, A. R. Kinship correlations reported by Sir Cyril Burt. *Behavior Genetics,* 1974a, 4, 1–28.

Jensen, A. R. Cumulative deficit: A testable hypothesis? *Developmental Psychology,* 1974b, 10, 996–1019.

Jensen, A. R. Cumulative deficit in IQ of blacks in the rural South. *Developmental Psychology,* 1977, 13, 184–191.

Jensen, A. R. Genetic and behavioral effects of nonrandom mating. In R. T. Osborne, C. E. Nobel, & N. Weyl (Eds.), *Human variation: The biophychology of age, race, and sex.* New York: Academic Press, 1978.

Jensen, A. R. *g*: Outmoded theory or unconquered frontier? *Creative Science & Technology,* 1979, 2, 16–29.

Jensen, A. R. *Bias in mental testing.* New York: Free Press, 1980.

Page, E. B. Miracle in Milwaukee: Raising the IQ. *Educational Researcher,* 1972, 1, 8–16.

Ramey, C. T., & Haskins, R. The modification of intelligence through early experience. *Intelligence,* 1981, 5, 43–57.

Scarr, S., & Yee, D. Heritability and educational policy: Genetic and environmental effects on IQ, aptitude, and achievement. *Educational Psychologist,* 1980, 15, 1–22.

Wilson, R. S., & Harping, E. B. Mental and motor development in infant twins. *Developmental Psychology,* 1972, 7, 277–287.