DOCUMENT RESUME

ED 042 837 UD 010 540

AUTHOR	Goldstein, Ellen R.
TITLE	Principal Sources for the Study of the Mutability of Intelligence and the Epidemiology of Mild Mental Retardation. ERIC-IRCD Urban Disadvantaged Series, Number 19.
INSTITUTION	Columbia Univ., New York, N.Y. ERIC Clearinghouse on the Urban Disadvantaged.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	Sep 70
CONTRACT	OEC-0-9-420088-2327(010)
NOTE	71p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$3.65
DESCRIPTORS	*Annotated Bibliographies, Cognitive Development, Disease Control, Disease Rate, Early Childhood Education, Intelligence Differences, *Intelligence Level, Intelligence Tests, *Mental Retardation, Mental Tests, National Intelligence Norm, Negro Achievement, *Resource Materials, *Retarded Children, Verbal Development
IDENTIFIERS	England, Scotland, Sweden, United States

ABSTRACT

This extensively annotated bibliography supplements an article by Zana Stein and Mervyn Susser entitled "Mutability of Intelligence and Epidemiology of Mild Mental Retardation," appearing in "Review of Educational Research," Volume 40, Number 1, February 1970, pages 29-67. Resource materials listed focus on: the fate of National intelligence; cognitive and social changes in the feeble minded; intelligence trends in certain districts of England; mental and scholastic tests for retarded children; schooling and IQ, Negro intelligence and achievement; early education of the retarded and disadvantaged; verbal development of imbecile children; prevalence of mental deficiency in Sweden; Scottish intelligence surveys of eleven-year-old children; adult status of children with contrasting early experiences; growth and development of mongoloids in infancy and early childhood; intelligence increments in families of dull children; institutional residence and intellectual functioning; and, the intelligence of East Tennessee mountain children. (RJ)

ERIC-IRCD URBAN DISADVANTAGED SERIES

Number 19, September 1970

PRINCIPAL SOURCES FOR THE STUDY OF THE MUTABILITY OF INTELLIGENCE AND THE EPIDEMIOLOGY OF MILD MENTAL RETARDATION

Ellen R. Goldstein

School of Public Health and Administrative Medicine
Columbia University

ERIC INFORMATION RETRIEVAL CENTER ON THE DISADVANTAGED
Horace Mann-Lincoln Institute · Teachers College · Columbia University · New York, N.Y. 10027

This bibliography is one in the ERIC-IRCD Urban Disadvantaged Series produced by the ERIC Information Retrieval Center on the Disadvantaged under Contract OEC-0-9-420088-2327(010) between the U. S. Office of Education and Teachers College, Columbia University. It supplements ERIC-IRCD Urban Disadvantaged Series, Number 18: Zena Stein and Mervyn Susser, Mutability of Intelligence and Epidemiology of Mild Mental Retardation, reprinted from Review of Educational Research, Vol. 40, February 1970.

The bibliography has been assigned the ERIC-IRCD accession number UD 010 540, and is available without cost from ERIC-IRCD during the period immediately following publication; thereafter, it can be obtained from:

ERIC Document Reproduction Service (EDRS)
National Cash Register Company
4936 Fairmont Avenue
Bethesda, Maryland 20014

In ordering from EDRS, use the document's assigned ED number. The bibliography will be abstracted in a monthly issue and announced in the semi-annual and annual indexes of Research in Education (RIE). Its ED order number will be listed in these indexes or can be obtained by writing to ERIC-IRCD.

September 1970

Cattell, R.B. "The Fate of National Intelligence: Test of a Thirteen-Year Prediction," Eugenics Review, 42: 136-148, 1950.

Cattell in England carried out two cross-sectional comparisons 13 years apart of the intelligence of ten-year-old schoolchildren in two cities, one highly industrial and one typically rural. In 1936 and again in 1949, he and his staff administered the Cattell Scale I, form A test plus a locally-used, non-verbal test, designed for children from eight to 11 years of age. In accordance with the growing concern about a decline in national intelligence, Cattell predicted that the mean intelligence of the population of ten year olds would decline by one point in ten years.

In 1936 Cattell attempted to locate for testing all ten-year-old children in schools in both cities. He finally found and administered the test to 2,873 children in 30 schools. Included were children in primary schools, private schools, preparatory schools, and special schools (schools for the physically and mentally defective as well as for delinquents). Probably excluded were children kept at home and thus not on the education registers, (for example retarded or defective children) and children absent the day of the test.

In addition to bias introduced by the sampling procedure, Cattell also failed to have well-defined age groups, since he tested some children who were younger and some who were older than the ten year olds because they attended the same classes.

In 1949, profiting by noting the poor sampling methods he employed the first time, Cattell again administered the same tests to all the ten-year-old children he could locate in the same two cities. This time, he found and tested 3,832 children in 53 schools and included children from some additional special and private schools.

The results of the comparison were not as Cattell had predicted. Though the improved sampling in the second testing program could have been expected to bias the results, the bias was against and not in favor of the result obtained. The mean I.Q. of the population rose from 100.487 in 1936 to 101.764 in 1949. The standard deviations remained rather constant at 21.97 and 20.22 respectively.

The following table reports the distribution of I.Q. scores in percentages:

I.Q.'S IN 1936 AND 1949 IN TERMS OF PERCENTAGES OF TOTAL GROUP

I.Q.	1936	1949
31-40	.03	.03
41-50	.35	.21
51-60	.85	.66
61-70	2.99	2.22
71-80	9.79	7.01
81-90	17.28	14.52
91-100	23.20	24.63
101-110	18.78	23.97
111-120	10.69	11.83
121-130	6.19	6.77
131-140	3.89	2.99
141-150	2.20	1.77
151-160	1.69	1.43
161-170	1.48	1.27
171-180	0.65	0.69
Median	97.61	99.80
Mean	100.49	101.76
SD	21.97	20.22

Looking at the scoring distribution, it can be seen that the major shift in scores has been toward the middle range I.Q.'s with low scorers in 1936 showing up as borderline and average scorers in 1949. The proportions of both low and high scorers decreased over the 13-year period. However, since the sample size was larger in 1949 and included more defective and backward children and low scorers, possibly the actual decrease in proportion of low scorers was even greater than that shown in the data.

In an attempt to investigate the shift, Cattell split both scoring distributions in half at the median. In 1936 the mean I.Q. of all children below the 50th percentile was 84.25. In 1949 the mean I.Q. below the median was 86.83. Thus, low scorers (all those below the 50th percentile) showed a mean gain of 2.6 I.Q. points. Using the same comparison for high scorers (those scoring above the median) showed that the mean I.Q. stayed relatively constant--i.e., 116.73 in 1936 and 116.70 in 1949.

Cattell concluded that on the basis of this data, the predicted fall in national intelligence was not borne out. Probably the growth of the national educational system, increased test sophistication, and general economic prosperity between the two test administrations accounted for the slight rise and offset the fall.

Clarke, A.D.B., Clarke, A.M. and Reiman, S. "Cognitive and Social Changes in the Feeble Minded: Three Further Studies," British Journal of Psychology, 49: 144-187, 1958.

In England, Clarke, Clarke & Reiman reported that in a six year study of a cohort of institutionalized high-grade retarded adolescents, I.Q. gains (Wechsler, Form I) were made over time. Though all of the subjects of the investigation had come from socially and economically disadvantaged homes, Clarke and Clarke (1953) found that those coming from the worst homes (i.e., those characterized as having the poorest psychological environments as compared to those coming from less adverse circumstances) made the greatest gains after removal to the institution.

Before analysis of their findings, two points need to be mentioned. First, since we know that, in general, institutionalized young children often come from families which have damaged structure and in which there is little cohesiveness, it can be assumed that these children when compared to retarded children living outside of institutions, have had social and psychological experiences which set them apart from others, making comparison difficult.

Secondly, institutions for the retarded have been shown to have sterile, unstimulating, and de-personalizing environments which generally act to depress normal growth and development. Therefore, it would be difficult to assume that the institution, in itself, could be the source of stimulation for I.Q. change found by Clarke & Clarke.

Taking account of these two points, Clarke & Clarke suggested that the gains shown over time by their sample of institutionalized retarded adolescents indicated recovery from the effects of early deprivation. This recovery, moreover, was facilitated by removal from the previous adverse environment, rather than by any attributes of the new environment.

Clarke & Clarke found that not only did severity of preinstitutional environment relate to the degree of improvement, but also gains took place in adolescence after the ages when intellectual growth was commonly assumed to have ceased. (See Table 3.1.4 in the text) They decided to retest their study sample two years later to see if there had been any further gain, and if found, if the gain could still be related to severity of background. They were aware that those of the original sample who could still be found in the institution (28 out of 59) were probably the least satisfactory members of the group, since about half of the original sample had been discharged as no longer being mentally deficient. However, the remaining group could be expected to give a minimal estimate of I.Q. change over a six-year period.

The results of retesting showed that those from the most adverse homes had gained a total of 16.2 points in the six-year period, while they had gained 9.7 points in the initial two years of the six-year period. Thus, gains were indeed continuous. Those from the less adverse homes had also gained, however their total gain was less marked: 10.2 points in six years, and 4.1 points in the initial two years of the six-year period. This comparison showed

that, for those from the most adverse homes, initial gains were the greatest in magnitude. This was evidence, Clarke & Clarke suggested, that intellectual gain was an immediate response to removal from an adverse home, and the continuing gain was evidence of a gradual fading of "intellectual scars."

In 1956, 32 available members of the original group were retested to see if the gains in I.Q. shown could be attributed to length of time interval between testings. Clarke & Clarke found that those from the most adverse homes had made their greatest gains initially in the total study period as shown by a previous investigation, while those from the less adverse homes made steady progress over the entire time period. The Clarkes concluded:

> "Thus, the difference between the gains made by both the groups is maximal at the beginning of the period in which increments have been studied, and subsequently because the growth rate of the bad home group tends to decelerate sharply...while the residual group maintains a steady progress, the difference between the average gains decreases."*

The next part of the research was conducted to find out if special environmental stimulation would speed and aid the "natural process of recovery" indicated by the gains over time. Twenty-one patients from two small rehabilitation units at the institution were retested after a period of one year of special treatment. During the year all had had at least six months of vocational training and six months of work in the community. Though they made gains,

*(p.148) (Clarke, Clarke & Reiman)

these were not greater than those made under normal hospital conditions. Once again, those from the most adverse backgrounds made the greatest gains (a difference of 3.7 points between the group from very bad homes and the group from less adverse homes). However, the difference between gains made as a result of time spent in the institution and those made as a result of the special stimulation was clearly small. Thus, Clarke & Clarke concluded that I.Q. was not affected in a positive way by the quality of the current environment, but rather that gain in I.Q. was most likely a direct result of removal from a negative environment and recovery from previous damage.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D. and York, R.L. Equality of Educational Opportunity. Washington, D.C.: U.S. Dept. H.E.W., Office of Education, 1966.

In 1966 Coleman and colleagues evaluated the results of the Head Start program which had been conducted throughout the United States the summer before. Disadvantaged children in various parts of the country attended summer preschool classes offered in public schools, community centers, etc. and were given special tutoring and preparation for regular elementary school. A year after admission to first grade, the children were tested and their verbal and non-verbal performance was compared with two groups of controls (matched for age, sex, race and region of the country) who could be classified as:
a) non-participants attending the same school as Head Start participants, and
b) non-participants living in communities where Head Start was not available.
The matched groups were introduced to control (in the analysis) for selection based on parental interest, area, and socio-economic status. The investigators recognized the possibility that: (1) parents of Head Start children were self-selected volunteers and were likely to have been highly motivated for their children's education; (2) selection by geographic area could also have entered into comparisons because the area in which Head Start was offered may have contained more children with background deficiencies than other areas; and (3) areas where Head Start was offered probably contained more children from lower socio-economic family backgrounds than areas where Head Start was not offered. To offset this, comparisons were made between participants

and non-participants in the same schools.

The following table summarizes the results of ability tests* by race, participation, and metropolitan - non-metropolitan geographic area for the entire country. (In the original Coleman et al report the reader can refer to Table 8.12.1 for complete data on regional variation.)

TABLE 4

Entire Country Group	Verbal		Non Verbal		Total Number of cases	
	Negroes	Whites	Negroes	Whites	Negroes	Whites
Metropolitan						
Head Start not available - couldn't participate	16.78	19.40	19.48	27.00	2,706	2,908
Head Start available - didn't participate	15.78	18.80	16.83	24.92	1,056	655
Head Start available - participated	15.91	18.50	16.13	23.45	3,011	996
Non-Metropolitan						
Head Start not available - couldn't participate	16.11	19.09	17.61	25.97	2,665	2,637
Head Start available - didn't participate	15.43	18.66	16.04	24.16	1,660	1,637
Head Start available - participated	15.50	17.85	16.04	22.00	3,001	1,994

*Tests derived from Inter-American Tests of General Ability.

The table shows that (1) Head Start participants of either racial group, black or white, did not score differently than non-participants of the same

racial group from the same schools. (2) Children of both racial groups from areas where Head Start was not offered show a definite advantage over children living where the program was offered. (3) Black children scored lower than white children in every comparison. (4) Children living in metropolitan areas did slightly better than those in non-metropolitan areas. In summary, then, the first year of the Head Start program was an overwhelming success!

One background variable which differentiated the children was prior educational experience, or specifically, kindergarten attendance. It was found that those children who had attended kindergarten prior to the Head Start summer program scored higher than those who had not. Possibly parents who enrolled their children in kindergarten might also have tended to enroll them in Head Start, and children who would benefit from a kindergarten experience would probably also benefit from Head Start.

Analysis by region showed that participating black children in the South scored higher than all non-participants in Head Start areas. However, their scores were not consistently higher than all participants in non-Head Start areas. Educational motivation differed for white and black participants and non-participants. Among black children, Head Start participants had a higher "index of favorable responses" to education than non-participants from the same schools. Among white children, however, participants had fewer responses than white non-participants. The difference was not explained. The white/black observed difference shows that the process of selection, and the impact of the program, or both, were not the same for the two groups.

Emmett, W.G. "The Trend of Intelligence in Certain Districts of England," Population Studies, 3: 324-337, 1950.

The trend of the intelligence level of the population of Great Britain between 1935 and 1947 was also described by Emmett. Utilizing data on the performance of 11-year-old boys and girls on the Moray House Test (the same test employed by the Scottish Council) Emmett showed that I.Q. scores fluctuated with environmental change. The results of an analysis of change in I.Q. scores between 1935 and 1947 demonstrated that, when scores of boys and girls were combined to yield a total score, a slight, non-significant fall over time was recorded. However, when the scores of boys and girls were separated, boys showed a significant fall in I.Q. while girls showed an equal and significant rise.

INTERMEDIATE STATISTICS OF THE INQUIRY*

Local Education Authority	Number of Children		Interval Bet. tests (yrs.)	Rise in Raw Score (difference bet. score on 1st and 2nd tests)
Boys:	1st Test	2nd Test		
Birmingham	7,065	6,686	9.02	-1.305
Darlington	593	495	10.81	-3.645
Northumberland	3,450	2,669	11.59	-0.914
Preston	844	746	9.99	-3.188
Smethwick	527	422	7.97	-3.635
Wigan	553	542	8.01	+1.622
Wolverhampton	1,140	1,078	7.61	-2.693
TOTAL	14,172	12,638		
			Rise in $\bar{X}$ I.Q. per year =	0.0921
Girls:				
Birmingham	6,705	6,430	9.02	+1.050
Darlington	561	542	10.81	-0.674
Northumberland	3,439	2,642	11.59	+3.452
Preston	870	718	9.99	-0.134
Smethwick	489	405	7.97	+3.442
Wigan	521	510	8.01	+3.963
Wolverhampton	1,139	1,012	7.61	-0.591
TOTAL	13,724	12,259		
			Rise in I.Q. per year	+0.0921

*Adapted from Emmett - 1950

To account for the changes in performance, Emmett suggested that the following factors may have been influential:

1. Schooling - The children taking the test in 1947 had had different school experiences than those taking it in 1935. Due to the onset of the war in the early part of the decade, schooling for most children was often interrupted to close the schools, pupils were evacuated from danger areas, and the caliber of the teachers varied. This experiential difference might have played a role in the amount of preparation as well as the amount of motivation of the pupils of the later testing program.

2. Sex differences - Emmett had found that girls did better in 1947 than they did in 1935, while boys did better in 1935 than they did in 1947. In 1935 boys and girls did equally well. However, since 1940 there had been a marked superiority of girls to boys, specifically in verbal intelligence, English ability and arithmetic. Emmett suggested that since the relative performance of boys and girls in the 12-year interval had shifted, the differences clearly were not innate.

In 1947, the Moray House Verbal Intelligence Test was given to groups of children aged 15 to 18 who were attending secondary schools. Comparing their performance at these later ages in 1947 with their earlier performance at age 11 years, Emmett found that though boys and girls differed appreciably at age 15 to 18, there were no marked differences between them at age 11. Testing at the later ages showed the superiority of boys over girls.

Emmett postulated that the superiority of girls at age 11, shown in the results of the second testing program in 1947, was transitory and could be attributed to a different environmental effect upon each sex. He suggested that World War II produced a relaxation of discipline and that this, in combination with other environmental factors, may have brought about changes in the behavior of school boys which indirectly affected I.Q. scores.

Gordon, H. "Mental and Scholastic Tests Among Retarded Children," Educational Pamphlet No. 44. London: Board of Education, 1923.

Gordon in England administered intelligence and scholastic tests to samples of canal-boat and gypsy children to find out if their performance would reflect their poor records of school attendance. Canal-boat childrer generally attended school about once a month for one or two days (about 5% of the school year). They lived with their families on canal boats, and they did not mingle socially outside of canal-boat people. The gypsies on the other hand, though leading nomadic lives, had slightly better attendance records (35% of the school year), though they tended not to be registered with the Education authorities unless "caught".

The tests used were the Stanford Revision of the Binet-Simon (with some revisions of terms and extensions) and various standardized scholastic tests (Reading, Adding and Subtracting). The names of 76 canal-boat children between the ages of five and 14 were taken from records in various schools in England. Since it was difficult to assemble these children at one time in school, testing was spread out over a long period of time in many different schools, presenting problems of interpretation concerning the comparability of the children.

The findings indicated low intelligence and a decrease in intelligence with an increase of age, even within families. The following table taken from Gordon (1923) presents the obtained results:

Average Mental* & Educational Ratios of Those Who Could Do All the Educational Tests

	N	Av. Age	Mental Ratio*	Educ. Ratio
Boys	14	8y 9m	77.6	70.8
Girls	22	9y 10m	67.6	70.1
B & G	36	--	71.5	70.4

* I.Q.

The mental ratio, as indicated, is the same as intelligence quotient (mental age divided by chronological age, multiplied by 100). The Educational Ratio is, similarly, the average of scores on 3 tests: reading, adding and subtracting.

It can be seen from the table that only 36 children out of 76 were able to do all three scholastic tests. The average I.Q. of these 36 was 71.5. The average I.Q. of the remaining 40 who could not do the scholastic tests was 67.9. The average of the total group of 76 canal-boat children was 69.6.

The differences shown between the performance of boys and girls could probably be attributed, according to Gordon, to a difference in age. It was found that the older the child, the lower his intelligence, and girls tended to be about a year older than boys.

Educational ratio was found to be dependent on school attendance. Moreover, I.Q. correlated quite well with educational ratio (.715) and therefore I.Q. was probably also related to attendance, Gordon concluded. What else could account for the low intellectual performance of canal-boat children, known to be poor attenders, unless it could be shown that the majority were mentally defective? More than likely, the intelligence tests were dependent on school attainments or mental exercises given in school. An indication that

the I.Q. scores were not reflections of heredity was the high negative correlation found between chronological age and I.Q. even of children within the same family. That is, the older the child, the lower was his I.Q. even among siblings living together. This would seem to indicate that the lack of schooling in an environment unfavorable to intellectual growth had a depressing effect upon performance on standardized I.Q. tests.

Gordon located 82 gypsy children between the ages of five and 14 in four different schools. Gypsies had more social contacts than canal-boat children and were generally more gregarious and were better at dealing with people.

Although slightly above the canal-boat children in school attendance, educational ratio and I.Q., they were still below the average 88% of all schoolchildren in England. Gordon found a decrease in I.Q. with increase of chronological age, a high positive correlation between I.Q. and school attendance, and a high positive correlation between I.Q. and Educational ratio.

Sixty out of 82 gypsies could do the scholastic tests, a larger proportion than the canal-boat children. The average I.Q. of those 60 was 75.4 and their Educational ratio was 77.4. The average I.Q. of the total group (N=82) was 74.5. The difference between boys and girls was very slight and Gordon attributed this to the fact they were the same age.

No Educational Ratios were computed for the 22 gypsies who could not

do the scholastic tests. The following table shows the results of I.Q. testing:

	N	Average Age	Average i.Q.
Boys	36	9 yr. 7 mo.	74.9
Girls	46	9 yr. 11 mo.	74.4
Boys & Girls	82	9 yr. 9 mo.	74.5

The correlation between I.Q. and chronological age was -.430. This relationship is quite weak compared to the stronger correlation of -.755 for canal-boat children. Gordon attributes this to the relatively greater amount of schooling of the gypsy children. The correlation between I.Q. and Educational ratio was .784, a relationship that was strong for the gypsies as it was for canal-boat children (.715), indicating that schooling and scholastic attainment was related to performance on intelligence tests.

The results of Gordon's investigation were suggestive, not conclusive. Due to the smallness of the samples and the possible unrepresentativeness of the children tested, generalizability is limited. It must be kept in mind that many canal-boat and gypsy children probably were not registered in schools, and therefore were not seen by Gordon. Furthermore, he only tested those in attendance at the time he was doing the study and probably missed a great number who were off with their families at the time. However, though there were methodological flaws, Gordon's finding of I.Q. related to low scholastic ratio and poor attendance, and the decrease of I.Q. as age increases (see

Table 2.1) is supportive of the findings of studies of disadvantaged children in the U.S. and lends support to the hypothesis that measured intelligence is affected by cultural and environmental forces.

Husēn, T. "The Influence of Schooling on the I.Q.," Studies in Individual Differences, eds. J. J. Jenkins, and D.G. Patterson. New York: Appleton-Century-Crofts, 1961, pp. 677-93.

Husēn in Sweden did an historical prospective cohort study of the intelligence of man tested at induction into military service in 1948 who had been tested ten years before in 1938. Though he had almost complete data for the same group of men tested at age ten and again at age 20 using the same intelligence test, Husēn's study was beset with methodological problems and his results must be interpreted with caution. The hypothesis that he attempted to prove was that schooling had a positive effect upon the results of intelligence tests.

In 1938 the "entire" school population (1,549 children) born in 1928 who were in third grade in Malmo primary schools were given a group intelligence test. The test was used nationally and consisted of time-limited scales and three verbal subtests. Then in 1948 all of the 20 year-old males examined at induction for military service were retested with the same basic test. Husēn was able to collect data on 85 percent of the original group of schoolboys and was left with I.Q. scores of 613 boys tested in 1938 and in 1948 (out of a total of 722 boys tested in 1938).

The main finding was that a rise in I.Q. took place as the level of attained schooling increased. Those who went beyond primary school and received secondary education had higher I.Q.'s in 1948 than those who stopped after primary school, even though their I.Q.'s in 1938 were similar.

The following table illustrates the change in I.Q. (or class intervals of I.Q.) for those above average as a function of schooling:

*Frequency of Ascent or Descent into Another Class Interval Between 1938 and 1948 in Various Groups of Schooling (Investigated range of I.Q. 105-134)

I.Q. position in 1948 compared with 1938	Primary School (N=123)	Junior Secondary school with leaving certificate (N=57)	Senior Secondary School and matriculation (N=62)
higher interval	5	18	44
same interval	34	56	43
lower interval	61	26	13

(*adapted from Husēn, p. 689)

This table, as well as Table 2.2 in the text illustrate a regression effect. Those subjects with only primary school education, who scored low in 1938, tended to score higher in 1948, while the high scorers in 1938 tended to score lower in 1948 (see Table 2.2). However, the regression effect is counterbalanced by the effect of schooling, according to Husēn, for no such pattern is shown for those with secondary schooling and above. This is shown in another table devised by Husēn:

I.Q. 1938 Interval	Number (in %) of S's who descended to a lower interval between 1938 and 1948. Primary School	Secondary School
95-104	25	--
105-114	56	14
115-124	70	26

This study had serious flaws and though Husēn subjected the data to many different statistical procedures, the findings are tenuous and much could probably be learned by focusing on the design rather than the results.

First of all, the sampling method employed is vulnerable to selection bias. Husēn only had data already collected and therefore he had no way of knowing whether the men were comparable initially on attributes other than I.Q. score. He could not control for such factors as socio-economic status and other such variables. A prior random matching would have solved this problem, but unfortunately Husēn had to settle for an posteriori approach.

Secondly, longitudinal studies are beset with regression effects and Husēn had no way of adequately controlling for this.

Third, testing and retesting after a ten-year interval using the same test, but varying the subtests resulted in a correlation of +.72 between the two test score distributions. The errors not accounted for may have produced some of the differences.

Fourth, a look into the past presents the problem of historical effects on cohorts. The effect of an improved environment over time cannot be measured by examining past data.

Fifth, Husēn was only able to collect data on 85 percent of the men tested in 1938. He did not state whether the 85 percent was comparable to the 15 percent, and moreover, whether the 85 percent was representative of the population of 20 year-old men in Malmo who had been tested at age 10.

Kennedy, W.A., Van De Riet, V. and White, J.C., Jr. "A Normative Sample of Intelligence and Achievement of Negro Elementary School Children in the Southeastern United States." Monographs of the Society for Research in Child Development, 28 (6, Serial No. 90) 1963.

Kennedy and his associates did a cross-sectional survey of the I.Q.'s of a stratified random sample of 1,800 black elementary school children in five southeastern states of the United States. Since it was known that black children, especially those living in the South, tended to have lower mean I.Q.'s than white children (on I.Q. tests standardized on white samples) Kennedy intended to establish a separate set of I.Q. norms for this relatively 'disadvantaged' population.

The sample children were in the first through sixth grades in three counties of each state. One county in each state was urban, another was rural, and one was metropolitan. One hundred and twenty students were selected at random from each county.

The tests, administered by the investigator and his students, were the 1960 Revision of the Stanford-Binet and Form W of the 1957 Revision of the California Achievements Test.

The mean obtained by the black children on the Binet Test was 80.7. The standard deviation was 12.4. The original mean of the 1937 Revision sample tested by Terman & Merrill was 101.8 and the standard deviation was 16.4.

The results by age can be seen in the following table:

STANFORD-BINET I.Q. MEAN AND STANDARD DEVIATION BY CHRONOLOGICAL AGE-TOTAL GROUP

AGE	NUMBER	I.Q. MEAN	SD
5 years	19	86.00	6.40
6 years	227	84.43	12.48
7 years	243	81.71	11.80
8 years	302	80.88	12.04
9 years	281	80.10	12.08
10 years	299	80.10	12.68
11 years	279	80.63	11.98
12 years	109	75.48	11.34
13 years	30	65.23	10.45
14 years	9	66.11	7.35
15 years	1	58.00	
16 years	1	51.00	
ALL AGES	1800	80.71	12.48

According to the table, there was a decrease in I.Q. corresponding to an increase in chronological age. However, when mean I.Q.'s and standard deviations were calculated for children of each grade in school, no such trend was shown.

STANFORD - BINET I.Q. MEAN AND
STANDARD DEVIATIONS BY GRADE (TOTAL SAMPLE)

GRADE	NUMBER	MEAN	SD
FIRST	300	81.81	12.43
SECOND	300	80.38	12.79
THIRD	300	80.25	11.90
FOURTH	300	79.65	11.83
FIFTH	300	81.20	12.60
SIXTH	300	80.77	12.87

As mentioned in the text, Schaeffer has suggested that children of five and six years of age are young to be in the early grades and are probably relatively brighter; and children of 12 years and over are generally old to be in sixth grade and are probably duller or are repeaters. Thus at the middle ages of eight to eleven Kennedy's data show no variation of I.Q. with age.

Data on the California Achievement test had to be interpreted carefully because of the problems encountered in carrying out the research. The test was administered at various times throughout the year instead of on one fixed day. Secondly, some data on 432 out of 1800 children were not obtainable and, therefore, their scores were not included in final analysis. In sum, then, Kennedy's data as a whole must be interpreted with caution.

Kennedy, W.A. "A Follow-up Normative Study of Negro Intelligence and Achievement," Monographs of the Society for Research in Child Development, 34, (126), 1969.

In 1965 Kennedy did a follow-up study of children living in the state of Florida who had been part of the original sample of 1,800 children of five Southeastern states. In the 1960 study Kennedy had found no significant differences in mean I.Q.'s between the states, so the sample of 360 in Florida was considered to be representative of the entire sample of 1,800. All 360 children were still found to be attending racially segregated, all black schools.

Data for comparison, using the same tests (Stanford-Binet and California Achievements Tests) was obtained for 312 out of the 360 children. Some children had dropped out of school during the four year interval, others had missed at least one of the tests.

The results of the second testing program showed no significant difference between the sample means over the four years. The mean in 1960 was 79.2 and in 1965 it was 79.4. There was a slight difference, however, between the standard deviation in 1960 (12.6) and that in 1965 (14.3) indicating slightly greater variability.

Kennedy found a decline in mean I.Q. across chronological age, but no decline in the individual child's mean I.Q. as he grew older. Thus, the decline across chronological age (as found in the 1960 study) as Schafer (1965) had suggested, was a sampling artifact.

The drops in achievement, as measured by the California Achievement Test, did seem to be significant. There was, according to Kennedy, a gradual loss of ground as children moved upward through the grades over the four years.

Kennedy located and tested, in 1965, ten out of the 24 school dropouts. As expected, their mean achievement level was much lower than that of the 1965 sample of 312. Thus, most probably the addition of their test scores would have brought the Binet and California Achievement means down.

In sum, a fall-off in academic achievement with age was observed in this sample of black elementary and high school children in the South. Socio-economic variables are inextricably tied to academic performance and it is obvious that black children in Southern schools have sub-standard living conditions and attend sub-standard schools.

Kennedy suggests that an important finding of this study, which is both optimistic and pessimistic, is the stability of the I.Q. of these children. Further analysis would seem to be needed, however, for results of change for each age cohort were not given, and this may have proved illuminating.

Kirk, S.A. Early Education of the Mentally Retarded. Chicago: University of Illinois Press, 1958.

Kirk devised a study in which four groups of mentally retarded children would be tested and observed over a four-year period. Three to six-year-old children with I.Q.'s ranging from 40 to 80 were to make up two sets of groups, one set living in their own homes (The Community Experimental and Community Control Groups) and the other set living in an institution (The Institution Experimental and Institution Control Groups).

The Community Experimental group consisted of 28 children residing in their own homes and attending a pre-school program from 9 A.M. to 3 P.M. daily for about three years. They were followed up for one year when, after pre-school, at age six they attended regular compulsory school. The Community Contrast group, a non-randomized group of 26 children who also lived with their own families, attended no pre-school classes and were tested along with the Experimental subjects.

The Institution Experimental group was composed of 15 "feeble-minded" institutionalized children attending pre-school classes from 9 A.M. to 3 P.M. daily in the institution in which they lived. At age six they attended an institution school and were followed for one year. Since a matched control group of subjects living at the institution at the same time was impossible to construct, Kirk searched the records in the same institution and those of one nearby and came up with test scores of 22 children whose progress he compared

with the institution Experimental group.

Kirk also had difficulty finding subjects for his Community Contrast group, since retarded children who have never attended regular school are often unknown to State and local authorities. The task of finding suitable matches for an already chosen Experimental group was very difficult, and Kirk finally had to select a contrast group of 26 children from among siblings of the experimental subjects and referrals from physicians. The control groups, therefore, cannot be assumed to have been comparable to the members of the Experimental group at the outset since selection procedures were not random.

Kirk characterized the pre-school as including: a) a nursery school environment, with a teacher-to-pupil ratio varying from one to four or one to five, and b) clinical individual tutoring. The formal aspects of this milieu encouraged close, personalized attention for the children in an atmosphere of intellectual stimulation. Testing was done at nine-month intervals and included various personality and social measures in addition to the Stanford and Kuhlmann revisions of the Binet test.

The results of I.Q. testing, as shown in Table 3.2, indicate that in the Community study, the experimental pre-school children made significant gains in I.Q. and maintained these gains at follow-up (after a year of regular school) at age seven. The contrast group, on the other hand, did not show any gains from first to last test at age six. However, at follow-up, at age seven, after they had attended regular school for one year, the contrast group did

show gain in I.Q.

The children in the Institution Experimental group showed significant gains in I.Q. which also lasted until age seven after a year of regular attendance in the Institution. The contrast group, however, showed a decline in I.Q. during the experimental period. At follow-up, after a year of regular school, no further loss was shown by the contrast children, indicating that perhaps the decline was arrested by institution school attendance.

Kirk attempted to account for the difference between the Community and Institution Contrast groups at follow-up in terms of social interaction. He suggested that the children in the community who started school at age six had the opportunity to interact with and be stimulated by normal children of their own age. The children in the Institution, on the other hand, only encountered other retarded children and did not have the example nor the challenges set up by normal peers. Another factor to be considered, however, was the possibility of initial differences between the group living at home and the one living in the Institution which might also account for some of the difference shown.

In order to test an original hypothesis relating to the differential effect which a pre-school program for retarded children might demonstrate, for children with organic etiology as compared to those with no organic etiology from psychologically deprived homes, Kirk compared the performance of children whose retardation was ascribed to an organic cause with the remainder.

The results of this comparison are shown below:

	Children with Organic Etiology* (Total) (N=14)		Children w/ No-Organic Etiology (Total) (N-29)	
	Community	Institution	Community	Institution
N showing gain	4	3	17	6
N showing loss	3	4	4	2
TOTAL:	7	7	21	8

*not brain damage

Analysis by etiology showed that 14 with "organic etiology" and eight of the 29 who had no known organic problems came from "culturally deprived homes." Kirk suggested that some of the mentally retarded children with organic etiology made gains as a result of change from total immersion in an impoverished home situation to a new milieu (i.e. pre-school in the community, or institution). As Tizard (1964) has said, there is little reason to think that a handicapped child profits less than another from food, love, playmates and learning situations. The gains shown by children with organic problems could have been attributed to the special stimulation of pre-school and regular school in the same manner as for the non-organically defective children.

The chief limitation of Kirk's experiment was the non-random method of selection of the study sample. His findings indicate that mentally retarded children living at home or in an institution can be stimulated to make developmental and intellectual gains by a special school program. Moreover, entering a regular school program as late as the age of six seemed not too late to stimulate the development of children who did not attend pre-school.

Klaus, R.A. and Gray, S.W. "The Early Training Project for Disadvantaged Children. A Report After Five Years." Monographs of the Society for Research in Child Development 33 (4), 1968.

The Early Training Project, developed by Klaus and Gray, was a five-year study in which a stratified random sample of a cohort of children born in 1958, randomly assigned to groups, were subjected to different treatment conditions. The goal of the program was to offset potential educational retardation. It was hoped that by stimulating attitudes relating to school achievement, as well as aptitudes for achievement, children who, without intervention would probably become intellectually backward over time, would be adequately prepared to offset this fall and meet the school experience with success.

The first group, T1, (consisting of 22 children) experienced a ten week pre-school session for three summers and had concurrently three years of weekly home meetings with a special teacher assigned to that group. The second experimental group, T2 (with 21 members) had two summers of a ten-week, pre-school session (started one year later than group T1), and two years of concurrent weekly meetings with a teacher. Two control groups, a "local" (18 children) and a "distal" (27 children) group, did not receive any pre-school training nor home visits. The local group resided in the same community as the experimental groups, while the distal group was located in a similar city 60 miles away.

TABLE 3.3

MEAN STANFORD BINET IQ SCORES FOR THE FOUR TREATMENT GROUPS AT EACH ADMINISTRATION

	(N = 19) T1	(N = 19) T2	(N = 18) T3	(N = 24) T4
Date of Administration	I.Q.	I.Q.	I.Q.	I.Q.
May 1962	87.6	92.5	85.4	86.9
-Intervention began for T1-				
August 1962	102.0	92.3	88.2	88.2
May 1963	96.4	94.8	89.6	87.4
-Intervention began for T2-				
August 1963	97.1	97.5	87.6	85.8
August 1964	95.8	96.6	82.9	80.8
-Intervention ceased; Regular school began-				
August 1965	98.1	99.7	91.4	89.4
June 1966	91.2	96.0	87.9	84.8

In addition to a daily schedule of varied pre-school activities, a schedule of positive and negative reinforcement was selectively applied to elicit attitudes and aptitudes for school achievement. The high adult-to-child ratio (one to four) facilitated this procedure, for desired behavior could be

rewarded immediately. Moreover, a large amount of verbal interaction took place as well as the giving and receiving of attention and affection.

Attitudes which were selectively elicited in the study were defined by the authors as: (1) achievement motivation, 2) persistence, 3) delay of gratification, 4) interest in school-like activities, and 5) identification with achieving role models, all of which are ostensibly inherent in the middle-class culture. Aptitudes which were fostered were: 1) perceptual development, 2) concept formation, and 3) language development.

Testing using the Stanford-Binet and Wechsler Intelligence Scale for children (WISC), among others, was done at the following intervals:

May 1962 - (before the pre-school program)

August 1962 - (after the first summer experience for group T1)

May 1963 - (after the first year of weekly meetings for T1 and the initiation of the summer experience for group T2)

August 1963 - (after the second summer for group T1 and the first summer for group T2)

August 1964 - (after the 3rd summer for group T1 and the second summer for group T2)

August 1965 - (after the first year of regular school for groups T1 and T2)

June 1966 - (after the second year of school for groups T1 and T2)

The results show that the experimental group children scored significantly higher than the controls on the intelligence tests at the end of the first year

of the program. (see Table 3.3 in the text) In fact, for group T1, the greatest improvement was shown initially, after the first summer session (August 1962). For group T2, similarly, the greatest improvement was shown in August 1963 after that group's first summer session. The gain shown by the experimental groups during the program exceeded their own performance before and after the program.

The control groups, on the other hand, showed decline in I.Q. over time. Their performance in August 1964 was significantly below that of August 1962, May 1963 and August 1963. Interestingly, however, after a year in the first grade they showed a positive leap in performance on the Binet and WISC tests, which exceeded the gain shown by the experimental groups after their year of first grade. Presumably, as the authors suggested, the school situation had at least the initial effect of producing gains. This finding, like Kirk's (1958) suggests that the age of six (at school entry) is not too late to alter a child's intellectual performance. However, the gains are not sustained and in this study, all four groups showed a decline in mean I.Q. by the end of the second year of regular school.

The findings of Klaus & Gray suggest several possibilities: 1) The rise shown after entry into the program (after the start of intervention) is not sustained after the program ends. Thus, there is something inherent in the treatment situation which produces gain. 2) The gain shown at entry to regular school is similar to that shown on entry to the program; however, it is not

sustained as long. Thus, there is something inherent in entry to the school situation (or a school-like situation) which produces gains, but the gains are not held as long in regular school as they are in the special program. Probably the personalized attention resulting from the close adult-to-child ratio in the pre-school program is responsible, at least in part, for the relative advantage.

Klaus & Gray investigated the family background of their study sample and ranked the children on the following grouped variables: 1) housing 2) education of parents 3) occupation of parents, and 4) family income. The rank order, from low to high was T1, T4, T3 and T2. The group with the poorest home background made the sharpest response to stimulation, a result reminiscent of that discussed in the research of Kirk and the Clarkes. Group T4, the distal control group, the second most disadvantaged group, might similarly have profited from the pre-school experience.

Generalizations are limited by the following: First, Group T2, which came from the most advantaged group of families in the study, had the highest mean I.Q. initially. The fact that the groups happened to be not comparable in composition (i.e., socioeconomically) though selection was based on random allocation, weakens support for the findings. Second, treatment was administered by different persons, and teacher variables present problems of reliability.

One of the strongest points made by the authors was that pre-school intervention needs to be continuous and long-term, if demonstrated gains are to be sustained. The data indicate that the leaps in I.Q. shown after the introduction of intervention are followed by gradual decline over time--a finding not uncommon to studies of disadvantaged groups in general.

Klineberg, O. Negro Intelligence and Selective Migration. New York: Columbia University Press, 1938.

The main portion of Klineberg's 1935 study is a set of cross-sectional comparisons of the intelligence test performance of black migrant children living in New York City for varying durations of time. During the time period 1931 to 1932, Klineberg and his associates tested over 3,000 black children in Harlem public schools, using the National Intelligence, the Pintner-Paterson, and the Stanford-Binet tests.

Klineberg's hypothesis was that black Southern migrant children who had come to New York City with their families would show a steady improvement in I.Q. with duration of stay in New York. To show that the gains demonstrated by the children were not due to selective migration of brighter black children, Klineberg examined the school records of migrant children in three large southern cities: Birmingham, Alabama; Nashville, Tennessee; and Charleston, South Carolina. A comparison of the school marks of migrants and non-migrants showed no consistent differences between the school performance of the groups.

In the New York City schools, Klineberg and his co-workers tested different groups of children living in the city for varying durations of time and compared the mean test scores among the migrant groups (i.e., those residing in New York for 1, 2, 3, 4 and 5 or more years) as well as between migrant groups and the New York City born children. The total study sample consisted

of 3,081 ten and twelve-year-old boys and girls in Harlem public schools between the years 1931 and 1932.

Three different examiners administered the National Intelligence Test (a group test) at three different times. The highest mean scores of all three test groups were made by the New York City born children, and the next highest by those migrant children living in New York City the longest (4 to 5 years). Though girls tended to score higher than boys, Klineberg noted that the differences were not "statistically reliable." One of the examiners found that the mean I.Q.'s for the boys who had just arrived, and those who had stayed five years were 64.43 and 76.72 respectively. For girls, the means rose from 63.66 for those who just arrived to 96.58 for those in New York for five years.

All testing using the Stanford-Binet, the only individual test used in the study, showed that those with the longest exposure to New York City schools had the highest I.Q.'s. The mean difference between those first arriving in New York and those in Harlem schools for four years was 4.7 points.

The Pintner-Paterson testing yielded inconsistent results. One examiner found gains in scores with duration of time in Harlem schools and the other found no gains. Klineberg attributed the discrepant findings of the Pintner-Paterson to the fact that the test does not measure that which is affected by environment, or schooling. The National Intelligence and Stanford-Binet tests are more heavily loaded with items measuring verbal ability. The New York

City schools, he contended, provide opportunity for improvement in verbal skills and this improvement is indicated by higher I.Q. scores.

One of the major weaknesses of Klineberg's design was that the groups of children may not have been comparable. By testing different children living in New York for different lengths of time, the differences found may have been attributable to inherent differences among the groups of children. One could also question the validity of Klineberg's use of relative school marks of Southern children to test the selective migration hypothesis. This measure does not adequately answer the question of whether successive waves of migrants were brighter than others.

Lee, E.S. "Negro Intelligence and Selective Migration: A Philadelphia Test of the Klineberg Hypothesis," American Sociological Review, 16: 227-233, 1951.

The 1951 study by Lee was a retrospective observation of a cohort of black children born in the South who had migrated with their parents to Philadelphia, compared with a cohort of black children born in Philadelphia. By successively testing all of the study children from the grade in which they entered the Philadelphia school system until grade 9, Lee was able to observe a steady gain in the I.Q. scores of migrants with duration of stay in Philadelphia, so that by grade 9, the I.Q.'s of the migrants approached, or were comparable to, the I.Q.'s of their comparison group.

Lee randomly selected nine junior and senior high schools in Philadelphia in which the proportion of black children varied from 23 to 92 percent. He obtained the records of all ninth grade children in those schools who had come to Philadelphia from Southern states (south of the District of Columbia and east of the Mississippi River). In addition, he chose a 20 percent systematic random sample of Philadelphia-born black children in the same classes and divided them into two groups: those who had attended kindergarten and those who had not. The migrants were divided according to the grade in which they had entered the Philadelphia school system. The range was from grade one to grade nine.

The tests used were the Philadelphia Tests of Mental and Verbal Ability, a battery of group intelligence tests, standardized on Philadelphia school children and similar to the Otis series. The table of results (Table 5.1 in the

text) shows the number of children in each group who were tested. Out of a total of 1,234 migrants whose records were studied, 292 were excluded for having missed one or more tests in the series. In addition, out of 962 Philadelphia-born blacks, 326 were excluded. These exclusions represent a great loss of information. However, Lee reported that though these exclusions may have affected intergroup comparisons, they did not disturb the patterns shown by the individual groups over time.

The table shows a continuous improvement in test scores for each of the cohorts of black migrant children as their length of residence in Philadelphia increased. The I.Q.'s of the children at entrance (ostensibly immediately following the migration) are comparable for every year group, but it can be seen that the longer the duration of stay, the greater the growth (or the more room there is for it). Age at entrance, therefore, does not seem to be as critical a variable for I.Q. gain as does length of exposure.

Differences between the migrant group entering in grade one and the Philadelphia-born children, who did not attend kindergarten, were reported to be significant until the last two tests (grades 6B and 9A) at which point the gap between the groups narrowed and differences were not significant.

Lee suggests that the gain in scores for the migrant children could not have been due to increasing familiarity with the tests, for gains were shown only by migrants and not for Philadelphia-born children, all of whom were tested the same number of times.

Moreover, Lee obtained data for the same children on the Chicago Tests of Primary Mental Abilities and the Minnesota Paper Form Board Test, both of which also showed improvement for the migrants with duration of residence in Philadelphia. Moreover, the results of the Chicago Tests indicated that no one factor or ability alone accounted for the rise in I.Q., since there was improvement over time for the migrants on every sub-test except 'memory ability'.

Lee did not establish the comparability of the groups and possibly, the Philadelphia-born were better off socio-economically than the migrants. This weakness of his study, as well as that caused by the high rate of exclusions from the final analysis, seriously weakened the reliability of the comparison between the Philadelphia-born and the migrants. However, as Lee has suggested, the general pattern of improvement over time for the migrants was clearly demonstrated and was likely to be the result of better schooling in the North.

Lyle, J.G. "The Effect of an Institution Environment upon the Verbal Development of Imbecile Children. III The Brooklands Residential Family Unit," Journal of Mental Deficiency Research, 4: 14-23, 1960.

Tizard & Lyle in London carried out a before-and-after experiment using a matched group of 32 children, half of whom were removed from a conventional mental deficiency institution and placed in a smaller residential house, and the other half left in the institution. The hypothesis tested was that imbecile children (IQ 25-50) between the ages of five and ten years could show significant gains in verbal ability if placed in a small setting in which there was much individualized attention, and intimate relations between adults and children.

Selection for the experiment was based on the following criteria: Children were capable of being tested non-verbally, had no hearing defects, no physical disabilities, and did not exceed the upper and lower age limits of 4 1/2 to 10 1/2 years. Sixteen pairs of children who met the criteria were matched for chronological age, sex, non-verbal and verbal intelligence, and where possible (in eleven cases) for mongolism. Sixteen of the children were then randomly allocated to the residential unit, Brooklands, and sixteen to a control group in the institution.

The test used was the Minnesota Preschool Scale of Intelligence (Form A) which contains separate verbal and non-verbal sub-tests. However, four pairs of the younger imbeciles had to be tested with the Merrill-Palmer test of non-verbal intelligence because their mental ages were below the lower

limit of the Minnesota (i.e. below MA 2 1/2 years).

Since it was found in another larger investigation of 194 imbeciles (Lyle, 1960)* that sex was not significantly related to achievement on various verbal tests, the investigators realized that mongolism could have been used as a criterion of matching instead of sex. Therefore, although there were only 11 pairs who could be matched by sex and clinical condition, all 16 pairs could be matched by clinical condition alone (there were now eight mongols in each group). For all 16 pairs, test scores were available at the beginning of the experiment and after 18 months.

Differences in improvement scores after 18 months (N=16).

Test	P (one-tailed)	In Favor of
Non-verbal intelligence (Minnesota C-scores)	Not significant	Brooklands
Verbal Intelligence (Minnesota C-scores)	.005	Brooklands

(Table adapted from Lyle, JG. 1960 p.20.)

This table as well as table 3.1.2 in the text shows that the Brooklands group developed verbally at a greater rate than the control group. The gains as shown are objective indices of improvement. However, the differential improvement in verbal M.A. based on "C scores" (Table 3.1.2) should be treated with caution. The Minnesota Preschool Scale is calibrated in C-scores

*Lyle, JF 1960 The Effect of an Institution Environment upon the verbal development of imbecile children. II Speech and Language. Journal of Mental Deficiency Research 4, 1.

(units of equal interval), and in the test manual C-score values have been extrapolated below the lower limit of the test (18 months).

In addition to showing improvement in verbal ability there were other behavioral changes in the Brooklands group. These changes were in the areas of social and emotional maturity.

Limitations on the generalizability of the study stem from the fact that: 1) the total sample size is small, and 2) the children selected for the study are not fully representative of children living in institutions, since only children free of physical handicaps or known sensory defects, and those suitable for testing were selected. However, Lyle contends that the experimental and control groups were ostensibly comparable and could be considered representative of the healthy, non-handicapped, testable imbeciles at the institution. Moreover, the experiment demonstrated that moving imbecile children to a more informal setting with a small adult-to-child ratio, even where there was no structured educational program, could accelerate their verbal and social development.

Rayner, S. "An Investigation of the Change in Prevalence of Mental Deficiency in Sweden," Hereditas--Genenetiskt Arkiv, 51: 297-314, 1964.

Rayner in Sweden did a retrospective analysis over the period 1944 to 1962 of rates of rejection for mental deficiency of potential military inductees. The dependent variable used by Rayner was rejection based on I.Q. and clinical judgement. The geographical areas compared were an isolated parish named Hallnas, the region containing Hallnas called 108, and the entire country of Sweden.

The test used was one devised by Military Psychologists in Stockholm and purportedly measured both verbal ability and performance. The same test was used throughout the period between 1944 and 1962, throughout the entire country.

Table 1.4 in the text shows the percentages of men judged to be mentally defective and hence rejected from the army, during three time periods. The criterion of rejection was a diagnosis of incapacity to bear arms. All those scoring less than 75 on the intelligence test were examined by a military doctor who made the ultimate decision of whether to accept or reject.

Rayner found that the percent of mental deficiency in Hallnas had declined between 1944 and 1962. The differences between the rates of mental deficiency in Hallnas and the rest of the region (108) were statistically significant until 1954. He attributed this decline to a breaking up of isolates (such as Hallnas) and a more favorable combination of polygenes in the population. However, when he showed the rates for the rest of the region as well as for the rest of

the country, it could be seen that the percentage of mental deficiency declined in both as well as in Hallnas. Rayner noted that the difference between the rates for Hallnas and the entire country are statistically significant; however, he did not indicate whether the differences between 108 and the country were significant. Judging by the closeness of the figures it would seem safe to suggest that the differences would not be significant and that the percent of mental deficiency in army inductees has declined throughout the country.

Since the period 1944 to 1962 seems too short for genetic alterations to be produced in the population, a more plausible explanation would be that there has been some widespread environmental change accounting for the decline. Inspection of the figures for Hallnas in 1944-48 shows an abnormally large percent of mental defectives in that period compared to the rest of the region as well as for the entire country. Very possibly the large proportion (20.2%) can be attributed to a cohort effect, or a generational effect, indicating that some environmental factor or factors were responsible for the largeness of the number. Unfortunately no data was available for other isolated parishes during the same time period for comparison.

A plausible explanation for the decrease in mental deficiency among potential army inductees in Sweden relates to the general social and economic improvement which the country experienced. More and better jobs made available, more adequate roads and transportation facilities, a better system of education, and improved and more comprehensive medical care, can be hypothesized

to have produced a general rise in the intellectual level of the Swedish population.

There are several sources of bias that must be considered in this study. First of all, Rayner reported (in personal communication) that the intelligence test was continually readjusted over the years to assure its reliability and validity. What effect this readjustment had on the total score as an objective index of intelligence is questionable. Secondly, Rayner reported that the final decision to accept or reject was made by a doctor and thus there was some uncertainty in diagnosis. He noted that, due to increasing technical demands etc., there has been a tendency to reject more dull inductees for military service. Both of these sources of bias would seem to affect the data in a direction unfavorable to the outcome, and therefore they lend further support to the interpretation that there has been a real decline in mental deficiency.

Scottish Council for Research in Education. The Trend of Scottish Intelligence: A Comparison of the 1947 and 1932 Surveys of the Intelligence of Eleven-Year Old Pupils. London: University of London Press, 1949.

In 1932 the Scottish Council for Research in Education administered intelligence tests to all eleven-year-old children in Scotland who could be located. The Mental Survey Group Test was given to all the children and the 1916 Stanford Revision of the Binet test was given individually to a random sample of 1,000. By using both types of tests, the Council hoped to validate the written against the individual test.

A total of 96,028 children were entered on the register of the Scottish Education Department and 91 percent or 87,498 were actually located and tested. The mean score for the group was 34,457 out of a possible total of 76 points. 19.77 percent were low scorers (scored below 19 points) and 17.70 percent were high scorers (scored 50 points or more).

The performance of low scorers is the focus of this review. A sex difference was found; 19.61 percent of the boys scored low, while only 17.91 percent of the girls were low scorers. These rates are under-estimates, since some physically and mentally handicapped children who were either unable to take the test or were unknown to education authorities, were excluded.

Because the results of the Binet testing were not based upon "true" random sampling, a better sample was selected in 1935 and 874 eleven-year-old children were administered the Binet test. The mean I. Q. for this sample was 100.11, the boys scoring slightly higher than the girls.

In 1947, fifteen years after the first testing program in 1932, the same age-group of Scottish children (11 years old born in 1936) was selected to test the hypothesis of declining national intelligence, prompted by the growing concern over increased population growth among the lower socioeconomic strata. The same procedures were used and the same group test was administered to all children registered with the Education Department who could be located in and out of school. Out of a total of 76,330 children registered, 93 percent or 70,805 eleven-year-olds were tested. The remaining seven percent, it was determined, were either absent the day of the test or were disqualified for reasons of health. In this testing program a special effort was made to test handicapped children, because the Council realized that omitting these children would reduce the generalizability of their findings. Therefore, a greater proportion of low scorers were included in the 1947 sample than had been included in 1932.

Another difference between the two testing programs, besides greater attention to problems of sampling, was that a 'long' sociological schedule was filled in by parents and teachers of a random sample of the 1936 cohort of tested children, In addition to the 'short'sociological schedule included with the group tests in both 1932 and 1947. The short schedule, filled in by teachers, asked about the child's birth rank, number of siblings taking the test, number of schools previously attended, number of times previously tested, date of birth and residence. The extra information provided by the 'long' schedule was: father's occupation, date of mother's birth,number of

rooms and persons at home, weight and height of child, medical record, with whom the child resided, and whether or not he had been evacuated during the war.

Between 1932 and 1947 the percent of low scorers on the group test decreased significantly from 18.77 percent to 16.5 percent. The decrease in the total proportion of low scorers could be attributed largely to the decrease in the proportion of low-scoring girls. Low-scoring girls decreased from 17.91 percent to 14.02 percent, while low-scoring boys decreased from 19.61 to 18.91. The percent of high scorers changed also, from 17.1 percent to 22.9 percent. This change in a positive direction suggested that the decrease in the percent of low scorers could not be explained wholly by the phenomenon of statistical regression to the mean. The mean group test score increased from 34.457 to 36.688 and, the mean for girls was slightly higher than that for boys.

The Binet test, as mentioned above, was administered to a random sample of the total eleven-year-old population in 1935 and in 1947 in order to calibrate the group test. The Terman-Merrill Revision was used in 1947 while the Stanford Revision was used in 1935. The investigators attempted to standardize the results for comparison by means of a mathematical procedure based on correlations between the two tests. The mean for 1,215 children given the Terman-Merrill Revision in 1947 was 102,523, two points higher than the mean of the group which took the test in 1935. However, after statistical analysis, it was found that the difference between the means was not significant.

The Council concluded that the intelligence level of the population, if changing at all, was probably on the rise. The years between 1932 and 1947

brought many educational, social and economic advances to the country. Specifically, educational facilities broadened and improved, more children were given opportunity for formal education, and more teachers and better teaching methods became available.

In addition to information gained on the intelligence level of the population, the Scottish surveys were able to relate a set of sociological variables to test scores. The findings pointed to the importance of social factors for intelligence. The table below* shows that both family size and intelligence varied by occupational class.

Class	Mean Family Size	Mean Test Score
I Professional and large employers	2.6	51.8
II Small employers and self-employed	3.1	42.7
III Salaried employees	2.5	47.7
IV Non-manual wage earners	3.1	43.6
V Skilled manual wage earners	3.6	37.2
VI Semi-skilled manual wage earners	4.3	33.2
VII Unskilled manual wage earners	4.6	31.1
VIII Farmers	3.9	36.2
IX Agricultural workers	4.3	32.3
ALL	3.8	33.7

*Table adopted from Maxwell, J. 1954, Intelligence, Fertility and the Future, Eugenics Quarterly I, 244-247.

The inverse relationship shown between mean intelligence scores and size

of family obtained on both the group and individual tests held within each occupational class.

Occupational class was related to some of the other variables studied. For example, size of family, occupancy rate of the home, age of mother, and migration history were all found to be related to occupational class. More concretely, families of low occupational class tended to be of larger size, to live in the most overcrowded homes, to have younger mothers, to live in rural areas, and to have low-scoring children.

The Council did not neglect to mention, in the final analysis, that beyond demographic variables, certain socio-psychologic variables seem worthwhile to consider for further study. Parents' ambition for the family, the standard of living aimed at, parental conception of responsibility toward their children, and other similar variables must surely have had a profound influence on the achievement levels of children, but, on these points they could provide no data.

Skeels, H.M. "Adult Status of Children with Contrasting Early Life Experiences," Monographs of the Society for Research in Child Development, 51(3), 1966.

One of the pioneering studies testing the hypothesis that intelligence is mutable was carried out in the early 1940's by Harold Skeels and colleagues at the Iowa Child Welfare Research Station. Skeels convinced Iowa state officials that it would be worthwhile to remove 13 mentally retarded children from an orphan home and place them instead as "house guests" in the care of older retarded girls in an institution. Skeels contended that the regimented, affectionless environment of the orphan home was depriving the children of stimulation which they could receive from the more personalized attention of the older girls. Though these older girls were also retarded, they were sufficiently ahead of the children to be able to provide "special mothering". Moreover, toys and games which could be placed on the wards would provide the opportunity for manipulative experience as well as social interaction. Skeels contended that this enriched environment, in comparison with the monotonous orphanage environment, would result in developmental gains for the children.

Since the "experiment" was designed to "rescue for normalcy"* no control group was selected in advance. However, after data on the 13 experimental children was collected and analyzed, Skeels found that because I.Q. tests were regularly given to all orphanage residents, a contrast group of orphan

*Skeels, p.9

children who had not been given any special stimulation over a period of time comparable to the experimental period could be selected from the records on file. The criteria for selection were that: a) the children had been given intelligence tests at the same age as the experimental group (under age two); b) they were still in residence at the orphanage at the age of four; c) they had never attended pre-school; and d) they had all been members of a control group of a previous study.

The 13 experimental group children consisting of three boys and ten girls had a mean age of 18.3 months. The contrast group consisting of eight boys and four girls had a mean of 16.6 months. All cases had been considered unsuitable for adoptive placement because they had been diagnosed as being mentally retarded. Birth, family, and medical histories were comparable. Most of the children in the study were illegitimate and all were born to families of low socio-economic and educational level. However, in the absence of random allocation, retrospective matching cannot rule out some built-in selection factor. For instance, the initial mean I.Q. scores of the contrast group were 24 points higher than those of the experimental subjects. Secondly, the groups were not matched for sex, the contrast group having more boys and the experimental group having more girls.

As shown in table 3.1.1, the experimental group showed a mean gain of 27.5 I.Q. points from first to last test (Kuhlmann-Binet). The range of individual gains was from 7 to 58 points, all gains being statistically significant. The length of the experimental period varied according to the rate of

development of each child. As soon as a clinical team judged that a particular child had attained "normal" mental development, the experiment ended and the child was considered for adoption. Out of the thirteen children in the experimental group, eleven were subsequently adopted and two, who had not shown large gains, remained in the institution.

The contrast group showed a significant mean decline of 26.2 I.Q. points. Because the initial I.Q.'s of the contrast group were higher than those of the experimental children, however, Skeels could not dismiss the possibility that the gains of the experimental group could be attributed to regression to the mean.

In a follow-up two and a half years later, the adopted children continued to show gains, while the two not adopted showed individual losses of 17 and 9 points. The contrast group, which had meanwhile been transferred from the orphan home to the state institution, showed a mean gain of 5.6 points. Most of these contrast group children had been moved into environments which were similar to that of the orphanage. However, a few did receive some special treatment in their new environments (i.e. some attended regular pre-school classes and one child took part in a pre-school experiment). These children, though older than the experimental group, showed intellectual gains during follow-up testing. It can be seen in the table that the contrast group children demonstrated gains during the post-experimental period.

Twenty-one years later, Skeels did a follow-up study of all the cases in his original sample. He was not able to measure their intelligence test performance, but he did find that all members of the experimental group were self-supporting, ranging in occupations from domestic worker to vocational counselor. The mean grade completed was 12.8 (excluding the two not adopted). Eleven were married.

Most members of the contrast group were not self-supporting, had not received education comparable to the experimental group, and still lived in various state institutions. Those who did work held unskilled jobs. Only one was married. This man had the most successful occupational career of the group. However, he had been the only member of the contrast group to have been included in a special mental growth stimulation study carried out by a student at the State University of Iowa during the follow-up period, at which time the child showed a gain of 22 I.Q. points.

The offspring of the former experimental group members had a mean I.Q. of 103.9. The range was from 86 to 125. No child tested below dull-normal. The only child of a contrast group member had, as it would be expected, a low I.Q. of 66.

One finding of this study was that the children in the experimental group who had the lowest initial I.Q.'s made the greatest gains, while those with the highest I.Q.'s in the contrast group showed the greatest losses. The gains appeared as spurts in development. It is difficult to attribute a cause to these

spurts because they suggest something besides the steady progress through learning and development. Although measurement errors and improved adjustment to repeated tests must be considered as possibly producing gains, the losses of the contrast group suggest otherwise. The most plausible interpretation is that the experimental group had normal intellectual potential which was held back at the outset by their environment. The contrast group, exposed to the same depressive effect of the orphanage environment, were never given the opportunity to attain normal rates of development.

General conclusions to be drawn from Skeels' study are limited by the fact that tests of intelligence at early ages rest heavily on social responses. The social responsiveness of infants is diminished by poverty of social experience, and it is therefore likely to be manifest in low performance on intelligence tests. Intelligence tests that rely less on social responsiveness might thus have shown less dramatic changes. The result cannot be generalized to older children under the same conditions.

The sustained intellectual and social gains shown by the experimental group in Skeels' last follow-up cannot be attributed solely to the initial experimental intervention. Because the children had been adopted their subsequent life experience was quite different from that of the contrast group. Their relatively "normal" lives and the opportunities of life in the open community must surely have contributed to their social adjustment and may well have contributed a continuing stimulus to their development.

Stedman, D.J. and Eichorn, D.H. "A Comparison of the Growth and Development of Institutionalized and Home-Reared Mongoloids During Infancy and Early Childhood," American Journal of Mental Deficiency 69: 391-401, 1964.

Bayley, N., Rhodes, L. and Gooch, B.A. "Comparison of the Growth and Development of Institutionalized and Home-Reared Mongoloids. A Follow-up Study." Prepublication Paper Submitted to the American Journal of Mental Deficiency, (1966).

Stedman and Eichorn were interested in the differential effects of hospital and home-rearing on the intellectual performance of children with Down's anomaly. They selected ten mongoloid infants between one and 4 1/2 months of age (who were free of physical abnormalities) and placed them in the Special Project Unit of a State Hospital. They then matched this experimental group by age, sex and physical condition with ten mongoloid infants who were living at home.

The hospitalized group lived in a ward of the hospital set up much like a nursery. There, the children were engaged in stimulating activities and were given the opportunity, for example, to play with toys and games, and to look through books. The adult-to-child ratio of 1 to 5 provided the chance for the children to receive affection and individual care. The surroundings were generally pleasant, quite unlike the usual ward environment in which mongoloid children were found.

At the time of testing, (using the Bayley 1933 Mental and 1936 Motor Scales and the Vineland test), the mean age of the hospital group was 28.8 months and that of the home-group was 27.3 months. The results of testing

showed that the children reared at home were superior to the hospital group in mental and social performance. However, differences between the groups on motor performance were not significant. The following table taken from Stedman and Eichorn presents their findings:

Mean Mental, Social and Motor Scores

Mental	Age Range	Hospital	Home	p
IQ	17-34 Mo.	37.3	52.1	.01
Social				
SQ	"	61.8	75.0	.02
Motor				
MQ	"	45.8	51.6	N.S.

The hospital group was found to be most retarded in language ability and skills in manipulating small objects.

Stedman and Eichorn concluded from their data that the hospital-reared group, despite the special nature of their ward environment, lacked the coaching and coaxing of parents and the example of normal peers that the home group had. This lack evidently acted to handicap the hospital group in relation to the home group.

A few years later, in 1966, Bayley and her colleagues attempted to test the same group of children originally studied by Stedman and Eichorn in 1964, in order to see if their performance had changed over time. All but three of the home-reared subjects were available for retesting at age five; two had

left the area and one had entered an institution. Retesting was carried out with the same tests plus the Binet (form L-M). For the ten hospital-reared children, the mean I.Q. declined 7.5 points from first to second testing. For the seven home-reared children, the mean I.Q. declined, during the same time period, a total of 14.4 points.

Since the hospital children had been found to be most deficient in language ability, a special intensive training program was instituted at the hospital soon after the second testing. The children spent two hours per day, five days a week in the special program, in addition to three hours per day playing in groups on the ward. By the third testing session a year later (at age six) all of the children had been attending nursery school, and the hospital group had been attending for three months, in addition, the intensive language training program.

Results of the third set of tests showed an increase in I.Q. scores for both hospital and home-groups. The greater gain, however, was shown by the hospital group. The mean gain of 6.5 points brought their level back to that shown by their initial performance at first testing even though they had shown a decline between first and second testing. The home group showed only a slight gain on the third test after nursery school attendance. However, this gain did not compensate for the initial loss shown between first and second testing, and the home-reared children, as a group, showed a decline in I.Q. over time. The results can be summarized by the following table:

Investigators	C.A.	HOSPITAL (N=10) I.Q.	Change	Total Change	HOME (N=7) I.Q.	Change	Total Change	p
Stedman & Elchorn (1964)	17-34 mo.	37.3			54.4			.01
			-7.5			-14.4		
Bayley et al (1966)	60 mo.	29.8		-1.0	40.0		-12.4	.001
			+6.5			+2.0		
Bayley et al (1966)	72 mo.	36.3			42.0			N.S.

The findings of this study indicated that, in general, the home (provided it is an adequate one) is superior to the hospital for promoting the mental development of mongoloid children. However, the introduction of an intensive language training program in the hospital can bring the performance level of hospital-reared children up to a level commensurate with that of children reared in their own home.

Generalizability of these results are limited by the following. First, neither of the studies presented information on the backgrounds of the children and possibly there were some initial differences between the groups. The discrepancy in I.Q.'s at the first testing (a difference of 17.1 points favoring the home-reared group) points up the possibility that the control group was not well-matched with the experimental group. Probably the decision to institutionalize a child rather than care for him at home is often based upon aspects of the home environment (i.e. socio-economic level of the family, structure

of the family unit, etc.) as well as upon qualities of the parents (i.e. health, age, mental ability, etc.). Therefore, we have no reason to assume that the institution and home-reared groups were initially comparable. Secondly, though the institution environment was described by the authors, no description was given of the environments of the home-reared group, making further comparisons difficult.

Stein, Z.A. and Susser, M. "Families of Dull Children." Part IV Increments in Intelligence. Journal of Mental Science, 106: 1311-1319, 1960.

Stein and Susser in England did a follow-up retest study of a stratified sample of 50 young adults, who had been classified as educationally subnormal in school. At the time of initial testing, the children had an average age of 11 and were examined with the Terman-Merrill Intelligence Test. At age 20-24, Stein and Susser administered the Wechsler Adult Intelligence Scale and corrected the Terman-Merrill so that it had a standard deviation equal to that of the WAIS (SD=15).

Based upon typologies developed in previous work, the authors divided the families of the young adults into sub-cultural categories. Six were noted to come from families characterized as "aspirant" and 40 from families which were "demotic."*

Sixteen of the young adults had neurological lesions (persumptive brain injury) and 30 were clinically normal.** Of the sixteen with neurological

*"Aspirant" families were those which were potentially socially mobile and in which one child in the family or one parent had attended grammar school or was in a non-manual occupation. "Demotic" families showed no signs of potential social mobility and contained no children or had no parents who had attended grammar school or who were in a non-manual occupation.

**Four cases were omitted from analysis, one who had not been tested by Stein and Susser and three others who had hearing defects.

lesions, six came from aspirant families and ten from demotic. All 30 of the clinically normal cases were from demotic familes.

Between the time of first testing at about age 11 (time of ascertainment as educationally sub-normal) and the time of follow-up at age 20 to 24, a greater number of clinically normal subjects than brain-injured subjects had made increments in intelligence. The following table shows I.Q. changes by clinical status:

I.Q. CHANGES BY CLINICAL STATUS: WECHSLER ADULT SCORE COMPARED WITH UNCORRECTED TERMAN MERRILL CHILDHOOD SCORE.

Clinical State	Increments	Unchanged	Decrements	Total Cases	Mean Change (I.Q. Points)	P
Normal	23	6	1	30	+8.3	.002
Brain Disorder	3	7	6	16	-1.2	

Differentiating the subjects by sub-cultural group showed that more cases from demotic families made increments in intelligence than cases from aspirant families. The following table presents the data.

I.Q. CHANGES BY SUB-CULTURAL GROUP AND CLINICAL STATUS: WECHSLER ADULT SCORE COMPARED WITH UNCORRECTED TERMAN-MERRILL CHILDHOOD SCORE

Subculture	Increments	Unchanged	Decrements	Total Cases	Mean Change
Demotic (clinically normal)	23	6	1	30	+8.3
Demotic (brain injured)	2	4	4	10	-1
Aspirant (all brain injured)	1	3	2	6	-1.7

As the data indicates, clinically normal cases who arise from demotic subcultures are capable of demonstrating increments in I.Q. Though the brain injured cases arose from both the demotic and aspirant sub-cultures, their gains were minimal and about the same in both groups. The clinically normal, however, were found only in the demotic group, and they did show gains. The evidence, though indirect, suggests that cultural environment plays a part in I.Q. change.

It seems improbable that the non-comparability of tests and measurement error could account wholly for the gain shown by demotic subjects. As the scores of the brain injured children did not change from first test to retest, the tests were probably comparable in these cases if not in all. Concerning measurement error, the authors suggest that regression to the mean would not account for the differences between those with and without lesions, although it could contribute to part of the increase shown by low scorers without lesions. Due to the small size of the total group, the generalizability of the results must be interpreted cautiously.

Sternlicht, M. and Siegel, L. "Institutional Residence and Intellectual Functioning," Journal of Mental Deficiency Research, 12: 119-127, 1968.

Sternlicht and Siegel did a longitudinal study of the effect of institutionalization on the intelligence test performance of mentally retarded persons. Their hypothesis was that institutionalization would produce a lowering of I.Q. scores due to the lack of adequate attention, affection and stimulation provided in the institution.

They divided a group of retarded persons soon after their admission to an institution into three categories by age, sex, and high or low I.Q. Thirty-two subjects were allotted to each age category ildren (age 5-11), adolescents (age 12-18) and adults (20-59). Then each age category was divided into two 16-subject groups according to sex. Finally, the groups were divided according to low and high I.Q. The final design was for there to be eight subjects in each cell, with four matched extra subjects to each group to be used in cases of attrition.

At the beginning and end of the four year experimental period, the children, aged 5 to 11 years, were tested with the 1960 Stanford-Binet test. Declines in I.Q. were observed from start to finish, and these declines were found to be statistically significant. (See Table 3.1.3 in the text.)

The members of the adolescent group were tested with the 1960 Stanford-Binet or alternatively, the WAIS. Though the trend in scores indicated a decline, the decline was not statistically significant.

The adult group showed relatively little change in I.Q. as measured by the 1960 Stanford-Binet or WAIS.

A Chi-square test was performed to see if there were any differences not likely to be due to chance between the performance of males and females on the tests. No statistically significant differences were found between male and female high I.Q. scorers and male and female low I.Q. scorers.

In general, the findings as shown in Table 3.1.3 indicate that retarded children are adversely affected by institutionalization whereas, the effect on retarded adolescents and adults is less marked. This has been demonstrated despite the fact that the high I.Q. children and half the low I.Q. children participated in an educational program in the institution. The program consisted of the teaching of some academic skills to the high I.Q. children, and some habit training to the low I.Q. children.

Wheeler, L.R. "A Comparative Study of the Intelligence of East Tennessee Mountain Children," Journal of Educational Psychology, 33: 321-334, 1942.

In 1942 Wheeler published the results of two cross-sectional studies of the intelligence of East Tennessee Mountain Children. During an initial testing program in 1930, 946 children in grades 1-8 in 21 mountain schools were given the Dearborn 1A and IIC Intelligence tests. In addition, 564 of these children in grades 3 to 8 were given the Illinois Intelligence Test. The median score of the Dearborn test was 82, that of the Illinois Test, 78. Dearborn scores showed a sharp decline with increasing chronological age: 95 at age six and 74 at age 16.

During the decade between 1930 and 1940, the East Tennessee mountain area was subject to many socio-cultural and economic improvements. In addition to the introduction of more jobs in the area, roads and transportation facilities improved, larger, better-equipped schools were built providing better teachers and educational materials. These changes were expected to have an impact upon the scholastic abilities of mountain children.

In 1940, Wheeler administered the same tests to a larger sample of children from a greater number of schools. He tested 2,946 children in the same 21 mountain schools and 19 others. He compared the I.Q. distributions of children in the 21 original schools with those of children in the 19 schools, found them comparable, and then combined them for his analysis. He noted that only nine percent of the mountain residents in 1940 were in-migrants to the mountains

since 1930, so he concluded that the populations in 1930 and 1940 were comparable.

The results of the 1940 testing showed that the median I.Q. score was 93 $\pm$ .25 as compared to 82 $\pm$ .40 in 1930. Moreover, the median I.Q.'s in 1940 were consistently higher at all ages. The differences were statistically significant at all ages except 16 (when there were a limited number of cases due to school leaving). Table 1.2 in the text shows the gradient of scores with chronological age for both groups.

Generalizability of Wheeler's findings are limited by the following gaps: he did not describe his sampling method in any detail and thus the representativeness of the samples could be questioned; and he did not mention children who did not show up for the test, were absent the day of the test, or children who were kept at home for various reasons.