DOCUMENT RESUME

ED 046 976 TM 000 336

AUTHOR	Jensen, Arthur R.
TITLE	Do Schools Cheat Minority Children?
INSTITUTION	California Univ., Berkeley. Inst. of Human Learning.
SPONS AGENCY	Rand Corp., Santa Monica, Calif.
PUB DATE	Apr 70
NOTE	69p.; Paper presented in Seminar Series on Education, The Rand Corporation, Santa Monica, California, April 1970
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Ability Identification, Academic Ability, Academic Achievement, Caucasian Race, Comparative Analysis, Defacto Segregation, *Educational Disadvantagement, Educational Discrimination, *Educational Equality, Educational Opportunities, *Elementary Schools, Environmental Influences, Ethnic Groups, Mexican Americans, *Minority Group Children, Negro Students, Personality, *Racial Differences, Self Concept, Socioeconomic Background

ABSTRACT

Large representative samples of Negro and Mexican-American children from Kindergarten through 8th grade in largely de facto segregated schools were compared with white children in the same California school district on tests of mental abilities and scholastic achievement, personality inventories, and indices of socioeconomic and cultural disadvantage. It was found that when certain ability and background factors over which the schools have little or no influence are statistically controlled, there are no appreciable differences between the achievement of minority and majority pupils. Moreover, there was no evidence of a "cumulative deficit" from lower to higher grade levels between the mean achievements of minority and majority pupils when the differences were measured in standard deviation units. It is concluded that these schools do not cheat minority students in terms of conventional criteria. But it might be concluded that minority children are, in fact, cheated if it were shown that their ability patterns require different instructional approaches to optimize their scholastic learning. Marked differences, not only in overall level of ability but also in the pattern of abilities, were found among all three ethnic groups. (Author/GS)

DO SCHOOLS CHEAT MINORITY CHILDREN?

Arthur R. Jensen
Institute of Human Learning
University of California, Berkeley

Paper Delivered in the Seminar Series on Education
The Rand Corporation
Santa Monica, California
April, 1970

Do Schools Cheat Minority Children?

Arthur R. Jensen

University of California, Berkeley

Americans' faith in education is tangibly substantiated in the fact that the American people now invest in educational institutions annually almost as much as all other nations combined. In the past two decades educational spending nationwide has increased fivefold while personal consumption merely doubled. Since World War II school enrollments have increased 88 percent, while school expenditures (in constant dollars) increased 350 percent. While employment in private industry increased 38 percent, it increased 203 percent in public education. With such an abundant outlay for education, the question naturally arises whether the benefits are equitably distributed to all segments of our population. A keystone of public education is the promise that no child should be denied the opportunity to fulfill his educational potential, regardless of his national, ethnic, or socioeconomic background. When substantial inequalities in educational achievement are evident between large segments of the population nominally sharing the same educational system, serious questions are raised, and rightly so. Numerous attempts have been and are being made to find the answers to the inequities in the benefits of education. In California the chief subpopulation differences in scholastic attainments involve majority-minority differences, the minorities in this case being Negroes and Mexican-Americans.

The causes of educational inequalities, in terms both of input and output, cannot be discussed very fruitfully in general terms. There are considerable regional and local differences in educational expenditures and facilities and in their distribution within local districts. In assessing the existence and degree of educational inequities, we must get down to specific cases. That is what is intended in this paper. We shall take a rather close look at some of the questions and answers involved in assessing inequalities within a single school system which serves three subpopulations: a majority group, which we shall refer to as Anglos, and two sizeable minorities, Negroes and Mexican-Americans. Before going into the details of this study, however, a few more general points should be reviewed.

School Comparisons of Academic Achievement

The now famous Coleman Report (Coleman, et al., 1966), which surveyed 645,000 pupils in more than 3,000 schools in all regions of the United States, found relatively minor differences in the measured characteristics of schools attended by different racial and ethnic groups but very great differences in their achievement levels. The Report also argued that when the social background and attitudes of students are held constant, per pupil expenditures, pupil-teacher ratio, school facilities and curricula show very little relation to achievement. The Report concluded ". . . that schools bring little influence to bear on a child's achievement that is independent of his background and general social context" (p. 325). A critical examination of this study by Bowles and Levin (1968) led them to the conclusion that Coleman's methodology could have resulted in an underestimation to some unknown degree of the extent of the relationship between school differences and pupil achievement. They also criticize the conclusion of the Coleman Report that,"There is a small positive effect of school integration on the reading and mathematics achievement of Negro pupils after differences in the socioeconomic background of the students are accounted for" (pp. 29-30). Bowles and Levin claim that ". . . the small residual statistical correlation between proportion white in the schools and Negro achievement is likely due, at least in part, to the fact that the proportion white in a school is a measure of otherwise inadequately controlled social background of the Negro student. Thus, we find that the conclusion that Negro achievement is positively associated with the proportion of fellow students who are white, once other influences are taken into account, is not supported by the evidence presented in the Report." Here then is one critique of the Coleman Report which suggests just the opposite of the most popularly held conceptions of what was proved by the Report. Bowles and Levin argue that school effects are probably larger than suggested by the study, and racial composition of the school *per se* is probably a more negligible factor than suggested in the Report's conclusions. A smaller-scale but statistically more thoroughly controlled study by Alan B. Wilson (1967) found that after controlling for other factors, the racial composition of the school had no significant direct association with Negro achievement, thus supporting the conclusion of Bowles and Levin, at least in the one California school district studied by Wilson.

But probably the most compelling argument for requiring racial balance in public schools is not the direct effect of a school's racial composition *per se*, but the fact that it could lead to a greater equalization of school facilities for majority and minority groups such that disadvantaged minorities would not be largely confined to schools with inferior resources. This may be a valid argument in some parts of the country, but one may justifiably question whether it is a cogent factor in California schools.

Consider the following evidence. A rather coarse-grained analysis of the relationship between the proportion of minority enrollment and certain school characteristics in California is made possible by the State Department of Education's recent publication of statistics on several scholastic variables for all school districts in the State. The present analysis, carried out by the writer, is based on only the total of 191 school districts in the ten counties of the greater Bay Area.[1]

The variables on which all school districts were ranked were: Grade 6 Reading Achievement, Grade 10 Reading, Grade 6 median IQ, Grade 10 median IQ, Proportion of Minority Enrollment, Per Pupil Expenditure, Teacher Salary, Teacher-Pupil Ratio (Grades 4-8), Number of Administrators per 100 Pupils, and General Purpose Tax Rate in the school district. The rank order correlations[2] among these variables for the 191 school districts are shown in Table 1. We see that minority enrollment has

Insert Table 1 about here

quite negligible correlations with all the school facility variables except number of administrators per 100 pupils (Variable 10), and this correlation is positive. On the other hand, there is a strong negative correlation between minority enrollment and the 6th and 10th grade Reading and IQ scores. This correlation matrix can be elucidated by factor analyzing it, thereby reducing it to three independent components which account for most of the variance (78%). This was accomplished by a varimax rotation of the first three principal components. The rotated factors are shown in Table 2. Factor I is scholastic aptitude (IQ),

Insert Table 2 about here

reading achievement and minority enrollment. Factor II represents the financial resources of the schools, with the highest loading on teacher

Table 1

Correlations (Spearman's ρ) Among Ten Educational Variables in 191 California School Districts (Decimals Omitted)

Variable	2	3	4	5	6	7	8	9	10
1. Grade 6 Reading	81	94	87	-73	23	21	18	18	-09
2. Grade 10 Reading		75	90	-70	08	06	02	-03	-06
3. Grade 6 IQ			85	-67	25	21	17	19	-08
4. Grade 10 IQ				-67	05	05	09	-13	00
5. Minority Enrollment					02	05	08	-10	17
6. Per Pupil Expenditure						35	53	42	47
7. Tax Rate							54	-06	24
8. Teacher Salary								18	45
9. Teacher/Pupil Ratio									01
10. No. Administrators/100									

Table 2

Rotated Factor Loadings for Ten Educational Variables

in 191 California School Districts

Variables	Factors I	Factors II	Factors III
1. Grade 6 Reading	.95	.12	.15
2. Grade 10 Reading	.92	.00	-.08
3. Grade 6 IQ	.92	.13	.17
4. Grade 10 IQ	.95	.06	-.17
5. Minority Enrollment	-.82	.19	-.09
6. Per Pupil Expenditure	.10	.67	.55
7. Tax Rate	.11	.75	-.15
8. Teacher Salary	.06	.83	.17
9. Teacher/Pupil Ratio	.03	.01	.96
10. No. of Administrators	-.13	.71	.01
Percent of Variance	42.0	22.8	13.6

salary. Factor III is teacher/pupil ratio and that part of per pupil expenditure not associated with Factor II. What this analysis shows most clearly is the absence of any appreciable correlation between the aptitude-achievement variables and the school district's financial outlay. If there were a substantial relationship between the financial resources and the reading achievement of the various school districts, the factors shown in Table 2 could not be so clearly separated. Note also that while minority enrollment has a negative correlation (-.82) with Factor I (IQ-Reading), it has a small positive correlation (+.19) with Factor II (expenditures). The negative correlation (-.09) between minority enrollment and Factor III indicates a slight disadvantage to districts with a high proportion of minorities in terms of average class size. Overall, these data suggest that there is no appreciable relationship between these particular school resources and minority enrollment, and if anything the correlation is in just the opposite direction to the popular belief that educational facilities are relatively inadequate in districts with a higher percentage of minority students.

Since this analysis is based on data in which the smallest unit for analysis is the school district, it permits no inference concerning the allocation of educational resources to the various schools, which probably differ in minority enrollments, within the districts. A similar analysis could be performed within a district, using the individual schools as the unit of analysis, but different indices of a school's resources would have to be used, since there would be relatively little variance on such variables as teacher salary and per pupil expenditure <u>within</u> any given school district. More fine-grained indices of the school's specific educational facilities should be included. In any case, the first and most obvious step in assessing the equality of educational facilities is to make a direct examination of the facilities, per pupil expenditures, etc. The recreational, hygienic, safety, and aesthetic aspects of the school plant should be considered no less than those facilities deemed to have more direct educational consequences, such as pupil/teacher ratio and special services.

The Misuse of National and Statewide Norms

School boards, the public, and the press commonly misuse the published and statewide norms on standardized achievement tests. Schools and districts are compared against "norms," which are intended to represent national or state averages, as if achieving a close approximation to the norms, if not exceeding them, should be the primary goal of every school system. Deviation from the norm, above or below, is commonly regarded as a credit or a discredit to the particular school system. The fallacy in this, of course, is the fact that the average level of scholastic achievement is a community is highly predictable from a number of the community's characteristics over which the local schools have no control whatsoever. Thorndike (1951), for example, correlated average IQ and an average scholastic achievement index (based on half a million children) with 24 census variables for a wide range of communities, large and small, urban and rural. Eleven of the correlations were significant at the 1 percent level. Census variables with the highest correlation with IQ and achievement were educational level of the adult population (.43), home ownership (.39), quality and cost of housing (.33), proportion of native-born whites (.28), rate of female employment (.26), and proportion of professional workers (.28). In a multiple correlation these census variables predicted IQ and achievement between .55 and .60. Essentially the same picture is revealed in many other similar studies (Wiseman, 1964, Chapter IV). A school's or district's deviation from the mean achievement predicted from a multiple regression equation based on a host of community characteristics would, therefore, make much more sense than a mere comparison of the school's average with national or state norms.

Majority-Minority Comparisons Within a School District

Even when a school district has equalized the educational facilities in all of its schools in terms of physical plant amenities, teacher salaries and qualifications, per pupil expenditures, teacher/pupil ratios, special services, curriculum, and the like, the question may still be asked whether majority-minority differences in scholastic achievement are a product of more subtle and less tangible factors operating in the school situation. We have in mind, for example, such factors as racial and socioeconomic composition of the school, differential teacher attitudes and expectancies in relation to majority and minority pupils. Is there any way we can assess the degree to which schools afford unequal educational advantages to majority and minority pupils over and above what can easily be reckoned in terms of pupil expenditures and the like?

I have tried to answer this question as best as I believe it can be answered with the psychometric and statistical methodology now available and with the rather modest resources within the financial means of most school systems. Although it would be impossible to present all the technical details and results of this study within the limits of this paper, it is possible to indicate some of the methods and the most relevant results they have yielded.

The study was conducted in 1970 in a fairly large (35 schools) elementary school district of California. This school district was ideal for this kind of study for four main reasons: (1) the district's school population has substantial proportions of Negro (13%) and Mexican-American (20%) students; (2) the majority (Anglo) population is very close to statewide and national norms for Anglos in IQ, for both mean and standard deviation, and the same is true for the two minority groups in relation to norms for their respective populations in the U. S.; (3) the schools are largely de facto segregated due to rather widely spaced residential clustering of the three ethnic groups, and (4) the district had made a thorough effort to provide equal educational facilities in all of its schools, if anything, favoring those schools with the largest minority enrollments to whom additional federal and state funds were allocated for special compensatory programs.

Large representative samples totalling 28 percent of the school population from grades K through 8 were selected for study. A total of 6,619 children were tested; more or less equal numbers were tested at each grade. The three main ethnic classifications were Anglo (N = 2453), Mexican-American (N = 2263), and Negro (N = 1853). Approximately half the sample (selected randomly with the classroom as the unit of selection) were tested by a small staff of specially trained testers, and half were tested by their regular classroom teachers. Because of the large sample sizes the tester vs. teacher results often differ significantly but do not differ appreciably or systematically except that the results of teacher administered tests consistently have somewhat greater variance and lower reliability which would tend to attenuate intercorrelations among measures and lessen the statistical significance of group differences. Parallel analyses for testers and teachers were run on all the data, which were combined when there were no significant or systematic differences between the two forms of testing. For the sake of simplicity in the present summary only the tester results are reported here when the two sets of data were not combined.

Rationale of the Study

In terms of this study one can think of the educational process as being analogous to an industrial production process in which raw materials ("input") are converted to a specified product ("output"). The output will be a function both of the input and of the effectiveness of the process by means of which the input is converted into output. In the case of schooling, the input is what the child brings with him to school by way of his abilities, attitudes, prior learning, cultural background, and personality characteristics relevant to learning in the classroom. The school itself has relatively little, if any, control over these input variables. The school, however, can have considerable influence on one variable -- prior learning -- for children who are already somewhere along the educational path, and if the school's instructional program is deficient for some children, the deficiencies in prior learning in earlier grades should show up increasingly in later grades as a cumulating deficit in scholastic achievement.

Whatever else one may say about it, schooling is essentially a process whereby children are helped to acquire certain skills, which are the output of the system. The effectiveness of the process can be judged, among other ways, in terms of the relationship between input and output. Meaningful comparisons cannot be made between the output (scholastic achievement) of different pupils, classes, schools, or school districts without reference to the input variables. The main purpose of the present study is the comparison of the outputs, i.e., educational achievements, of three categories of pupils -- Anglo, Negro, and Mexican-American -- when these groups are statistically equated on the input variables. In this way we can make some judgment concerning the relative efficiency of the educational process for each of the three groups. The adequacy of the statistical equating of the groups in terms of input depends upon a judicious selection of instruments for measuring the input variables. The chief aims in selecting the input control variables are (1) to represent the domain of educationally relevant abilities, personality, and home background factors as broadly as feasible, and (2) to include only those ability and background variables which are not explicitly taught by the schools or are not under direct control of the schools. That is to say, they should represent the raw materials that the schools have to work with. The output, on the other hand, should represent objective measures of those skills which it is the school's specific purpose to teach. These are best measured by standardized tests of scholastic achievement.

The input variables can be classified into three categories: (1) ability or general aptitude tests, (2) motivation, personality, and school-related attitudes, and (3) environmental background variables reflecting socioeconomic atatus, parental education, and general cultural advantages.

Input Variables

Ability Tests

Lorge-Thorndike Intelligence Tests. This is a nationally standardized group-administered test of general intelligence. In the normative sample, which waa intended to be representative of the nation's school population, the test haa a mean IQ of 100 and a standard deviation of 16. It ia generally acknowledged to be one of the best paper-and-pencil tests of general intelligence.

The Manual of the Lorge-Thorndike Test states that the test was designed to measure reasoning ability. It does not test proficiency in specific skills taught in school, although the verbal tests, from

Grade 4 and above, depend upon reading ability. The reading level required, however, is intentionally kept considerably below the level of reasoning required for correctly answering the test questions. Thus the test is essentially a test of reasoning and not of reading ability, which is to say that it should have more of its variance in common with nonverbal tests of reasoning ability than with tests of reading *per se*.

The tests for Grades K-3 do not depend at all upon reading ability but make use exclusively of pictorial items. The tests for Grades 4-8 consist of two parts, Verbal (V) and Nonverbal (NV). They are scored separately and the raw score on each is converted to an IQ, with a normative mean of 100 and SD of 16. The chief advantage of keeping the two scores separate is that the Nonverbal IQ does not overestimate or underestimate the child's general level of intellectual ability because of specific skills or disabilities in reading. The Nonverbal IQ, however, correlates almost as highly with a test of reading comprehension as does the Verbal IQ, because all three tests depend primarily upon reasoning ability and not upon reading *per se*. For example, in the 4th Grade sample, the correlation between the Lorge-Thorndike Verbal and Nonverbal IQs is .70. The correlation between Verbal IQ and the Paragraph Meaning Subtest of the Standard Achievement Test is .52. The correlation between the Nonverbal IQ and Paragraph Meaning is .47. Now we can ask: What is the correlation of Verbal IQ and Paragraph Meaning when the effects of Nonverbal IQ are partialled out, that is, are held constant? The partial correlation between Verbal IQ and Paragraph Meaning (holding Nonverbal IQ constant) is only .29.

The following forms of the Lorge-Thorndike Intelligence Tests were used:

Level 1, Form B.		Grades K-1.
Level 2, Form B.		Grades 2-3.
Level 3, Form B.	Verbal and Nonverbal.	Grades 4-6.
Level 4, Form B.	Verbal and Nonverbal.	Grades 7-8.

Figure Copying Test. The Figure Copying Test was given in Grades K-6. Beyond Grade 6 too large a proportion of children obtain the maximum possible score (30) for the test to be useful in making group comparisons. In fact, by Grades 5 and 6 group differences are very probably underestimated by this test, since a larger proportion of the higher-scoring group will obtain the maximum score and this "ceiling" effect will prevent the group's full range of ability from being represented. The ceiling effect consequently spuriously depresses the group's mean and reduces the variance (or standard deviation). Nevertheless, this test is extremely valuable for group comparisons because it is one of the least culture-loaded tests available and successful performance on the test is known to be significantly related to readiness for the scholastic tasks of the primary grades, especially reading readiness.

The Figure Copying Test was developed at the Gesell Institute of Child Study at Yale University as a means for measuring developmental readiness for the traditional school learning tasks of the primary grades. The test consists of the ten geometric forms shown in Figure 1, arranged in order of difficulty, which the child must simply copy, each on a separate sheet of paper. The test involves no memory factor,

---

Insert Figure 1 about here

---

since the figure to be copied is before the child at all times. The test is administered without time limit, although most children finish

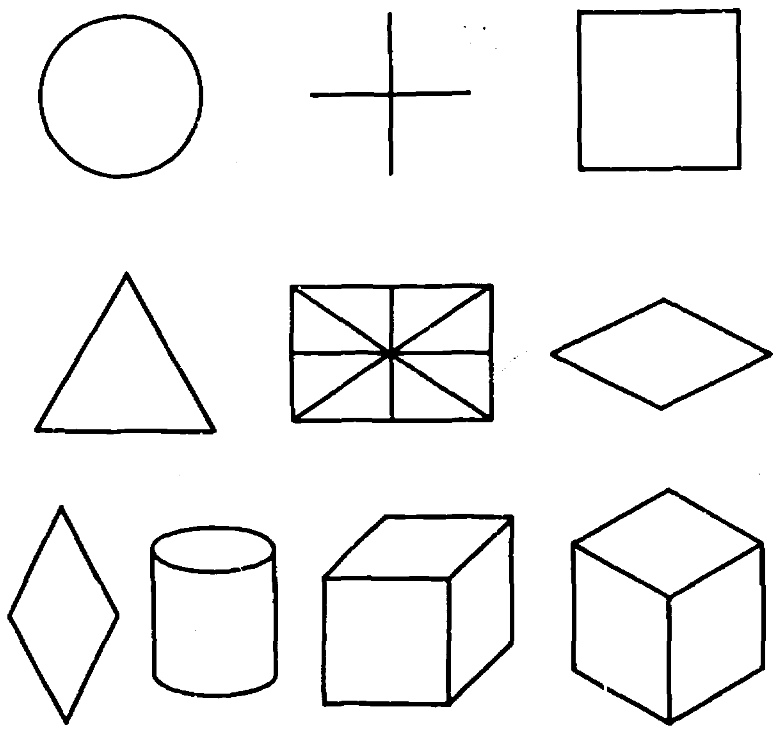

Fig. 1. The ten simple geometric forms used in the Figure Copying Test. In the actual test booklet each figure is presented singly in the top half of a 5-1/2" x 8-1/2" sheet. The circle is 1-3/4" in diameter.

in 10 to 15 minutes. The test is best regarded as a developmental scale of mental ability. It correlates substantially with other IQ tests, but it is considerably less culture-loaded than most usual IQ tests. It is primarily a measure of general cognitive development and not just of perceptual-motor ability. Children taking the test are urged to attempt to copy every figure.

Each of the ten figures is scored on a 3 point scale going from 1 (low)to 3 (high). (A score of zero is given in the rare instance when no attempt has been made to copy a particular figure.) A score of 1 is given if an attempt is made but the child's drawing completely fails to resemble the model. A score of 2 is given if there is fair resemblance to the model -- the figure need not be perfect but it must be easily recognizable as the model which the child has attempted to copy. A score of 3 is given for an attempt which duplicates the figure in all its essential characteristics -- this is an essentially adult level of performance. Since there are ten figures in all, the possible range of scores goes from 10 to 30 (or 0 to 30 if zeros are counted, but this is rare, since virtually all subjects attempt all ten figures).

The high level of motivation maintained by this test is indicated by the fact that the minimum score obtained in each group at each grade level increases systematically with grade level. This suggests that all children were making an attempt to perform in accordance with the instructions. Another indication that can be seen from the test booklets is that virtually 100 percent of the children in every ethnic group at every grade level attempted to copy every figure. The attempts, even when unsuccessful, usually show considerable effort, as indicated by redrawing the figure, erasures and drawing over the figure repeatedly in order to improve its likeness to the model. It is also noteworthy about this test that normal children are generally not successful in drawing figures beyond their mental age level and that special instructions and coaching on the drawing of these figures hardly improves the child's performance. This test, in other words, is not very susceptible to training, but measures some fundamental aspects of mental development. The diagnostic significance of this test has been explicated extensively in School Readiness (Harper & Row, 1967, pp. 63-129) by Drs. Frances L. Ilg and Louise Bates Awes of the Gesell Institute of Child Development at Yale University.

Raven's Progressive Matrices. This nonverbal reasoning test, devised in England, is intended to be a pure measure of *g*, the general factor common to all intelligence tests. It is a highly reliable measure of reasoning ability, quite free of the influence of special abilities, such as verbal or numerical facility. It is probably the most culture-free test of general intelligence yet devised by psychologists. The test mainly gets at the ability to grasp relationships; it does not depend upon specific acquired information as do tests of vocabulary, general information, etc. The test, which is group administered, begins with problems that are so easy that all children by third grade can catch on and solve the problems even without instructions.

Two forms of the test were used. The Colored Progressive Matrices, which is the children's form, was used in grades 3 to 6. This test is appropriate even for kindergarten children, but to insure that all children tested could go through the first several problems without difficulty, giving them a chance to catch on easily and experience success in the early part of the test, we used this test only from the

3rd grade and above. The Colored Matrices consist of 36 matrix problems which are administered without time limit. Children are encouraged to attempt all problems. There is no penalty for guessing.

The Standard Progressive Matrices were used in Grades 7 and 8. These begin as easily as the colored matrices but advance in difficulty more rapidly and go up to a level appropriate for average adults. There are 60 matrix problems in all, and the subjects are encouraged to attempt all of them, without penalty for guessing.

Listening-Attention Test. In the Listening-Attention Test the child is presented with an answer sheet containing 100 pairs of digits in sets of 10. The child listens to a tape recording which speaks one digit every two seconds. The child is required to put an X over the one digit in each pair which has been heard on the tape recorder. The purpose of this test is to determine the extent to which the child is able to pay attention to numbers spoken on a tape recorder, to keep his place in the test, and to make the appropriate responses to what he hears from moment to moment. Low scores on this test indicate that the subject is not yet ready to take the Memory for Numbers test which immediately follows it. High scores on the Listening-Attention Test indicate that the subject has the prerequisite skills for taking the digit span (Memory for Numbers) test. The Listening-Attention Test thus is intended as a means for detecting students who, for whatever reason, are unable to hear and to respond to numbers read over a tape recorder. The test itself makes no demands on the child's memory, but only on his ability for listening, paying attention, and responding appropriately -- all prerequisites for the digit memory test that follows.

It has been found in previous studies using the Listening-Attention Test that the vast majority of subjects from Grade 2 and above obtain perfect scores; the median score is 100, and the lower quartile rarely goes below 95. This means that nearly all subjects have the prerequisite skills for the Memory for Numbers test to yield a valid measure of the subjects' short-term memory ability.

Memory for Numbers Test. The Memory for Numbers test is a measure of digit span, or more generally, short-term memory. It consists of three parts. Each part consists of six series of digits going from four digits in a series up to nine digits in a series. The digit series are presented on a tape recording on which the digits are spoken clearly by a male voice at the rate of precisely one digit per second. The subjects write down as many digits as they can recall at the conclusion of each series, which is signaled by a "bong." Each part of the test is preceded by a short practice test of three digit series in order to permit the tester to determine whether the child has understood the instructions, etc. The practice test also serves to familiarize the subject with the procedure of each of the subtests. The first subtest is labeled Immediate Recall (I). Here the subject is instructed to recall the series immedistely after the last digit has been spoken on the tape recorder. The second subtest consists of Delayed Recall (D). Here the subject is instructed not to write down his response until after ten seconds have elapsed after the last digit has been spoken. The ten-second interval is marked by audible clicks of a metronome and is terminated by the sound of a bong which signals the child to write his response. The Delayed Recall condition invariably results in some retention decrement. The third subtest is the repeated series test, in which the digit series is repeated three times prior to recall; the subject then recalls the series immediately after the last digit in the series has been presented. Again, recall is signaled by a bong. Each repetition of the series is separated by a tone with a duration of one second. The repeated series almost invariably results in greater recall than the single series. This test is very culture fair for children in second grade and beyond and who know their numerals and are capable of listening and paying attention, as indicated by the Listening-Attention Test. The maximum score on any one of the subtests is 39, that is the sum of the digit series from four through nine.

Motivational and Personality Tests

Speed and Persistence Test (Making X's). The Making X's Test is intended as an assessment of test-taking motivation. It gives an indication of the subject's willingness to comply with instructions in a group testing situation and to mobilize effort in following those instructions for a brief period of time. The test involves no intellectual component, although for young children it probably involves some perceptual-motor skills component, as reflected by increasing mean scores as a function of age between grades 1 to 5. The wide range of individual differences among children at any one grade level would seem to reflect mainly general motivation and test-taking attitudes in a group situation. The test also serves partly as an index of classroom morale, and it can be entered as a moderator variable into correlational analyses with other ability and achievement tests. Children who do very poorly on this test, it can be suspected, are likely not to put out their maximum effort on ability tests given in a group situation and therefore their scores are not likely to reflect their "true" level of ability.

The Making X's Test consists of two parts. On Part I the subject is asked simply to make X's in a series of squares for a period of 90 seconds. In this part the instructions say nothing about speed. They merely instruct the child to make X's. The maximum possible score on Part I is 150, since there are 150 squares provided in which the child can make X's. After a 2-minute rest period the child turns the page of the test booklet to Part II. Here the child is instructed to show how much better he can perform than he did on Part I and to work as rapidly as possible. The child is again given 90 seconds to make as many X's as he can in the 150 boxes provided. The gain in score from Part I to Part II reflects both a practice effect and an increase in motivation or effort as a result of the motivating instructions, i.e., instructions to work as rapidly as possible.

Ethnic and social-class group differences on this test are generally smaller than on any other test, with the exception of the Listening-Attention Test (on which there are almost no group <u>or</u> individual differences).

<u>Eysenck Personality Inventory-Junior</u>. The EPI-Junior is the children's form of the EPI for adults. It is a questionnaire designed to measure the two factors of personality which have been found to account for most of the variance in the personality domain -- Extraversion and Neuroticism. The Extraversion (E) scale represents the continuum of social extraversion-introversion. High scores reflect sociability, outgoingness and carefreeness. The Neuroticism (N) scale reflects emotional instability, anxiety proneness, and the tendency to develop neurotic symptoms under stress. The Lie (L) scale is merely a validity detector consisting of a number of items which are very rarely answered in the keyed direction by the vast majority of subjects. A high score on L indicates that the subject is "faking good" or is answering the questionnaire items more or less at random, either intentionally or as a result of insufficient

comprehension of the items. Naïveté is also reflected in elevated L scores, and it is probably mainly this factor which causes a decrease in L scores as children mature.

The EPI scales were included in the present study as a control variable because previous studies had shown the E and N scales to predict a small but significant part of the variance in scholastic performance. Because of the reading level required by the EPI, it was not given below the 4th grade.

Student Self-Report. This 21-item self-report inventory was composed mainly of items in the self concept inventory used by James Coleman in his study, Equality of Educational Opportunity. It reveals the student's attitudes toward school, toward himself as a student, and other attitudes affecting motivation and self-esteem. The questionnaire was administered by the classroom teachers in grades 4 through 8. Because of the reading level required, it was not administered below grade 4.

Background Information

The Home Index. This is a 24-item questionnaire about the home environment, devised by Harrison Gough (1949). It is a sensitive composite index of the socioeconomic level of the child's family. Factor analysis of past data by Gough has shown that the 24 items fall into 4 categories, each of which can be scored as a separate scale. Part I (Items 6, 7, 8, 9, 10, 15, 16, 23) reflects primarily the educational level of the parents. Part II (Items 1, 2, 3, 4, 5, 13, 20, 24) reflects material possessions in the home. Part III (Items 17, 18, 21, 22) reflects degree of parental participation in middle or upper-middle class social and civic activities. Part IV (Items 11 and 19) relates to formal exposure to music and other arts.

<u>Output Variables -- Scholastic Achievement</u>

<u>Stanford Achievement Tests.</u> Scholastic achievement was assessed by means of the so-called "partial battery" of the Stanford Achievement Tests, consisting of the following subtests: Word Meaning, Paragraph Meaning, Spelling, Word Study Skills, Language (grammar), Arithmetic Computation, Arithmetic Concepts, and Arithmetic Applications. The Stanford Achievement battery was administered in grades 1 through 8.

<u>Distinction Between Aptitude and Achievement</u>

Can we justify the separation of our tests into two categories, ability or aptitude tests versus scholastic achievement tests, and then regard the former as <u>input</u> and the latter as <u>output</u>? Do not intelligence or aptitude tests also measure learning or achievement? The answer to this question is far from simple, but I believe there are at least six kinds of evidence which justify a psychological distinction between intelligence tests and achievement tests:

(1) <u>Breadth of Learning Sampled.</u> The most obvious difference between tests of intelligence and of achievement is the breadth of the domains sampled by the tests. Achievement tests sample very narrowly from the most specifically taught skills in the traditional curriculum, emphasizing particularly the 3 R's. Achievement test items are samples of the particular skills that children are specifically taught in school. Since these skills are quite explicitly defined and the criteria of their attainment are fairly clear to teachers and parents, children can be taught and can be given practice on these skills to shape their performance up to the desired criterion. Because of the circumscribed nature of many of the basic scholastic skills, the pupil's specific weaknesses can be identified and remedied. The skills or learning sampled by an intelligence test, on the other hand, represent achieve-

ments of a much broader nature. Intelligence test items are sampled from such a very wide range of potential experiences that the idea of teaching intelligence, as compared with teaching, say, reading or arithmetic, is practically nonsensical. Even direct coaching and practice on a particular intelligence test raises individual's scores on the average by only five to ten points; and some tests, especially those referred to as "culture fair," seem to be hardly amenable to the effects of coaching and practice. The average five year old, for example, can copy a circle or a square without any trouble, but try to teach him to copy a diamond and see how far he gets! Wait until he is seven years old and he will have no trouble copying the diamond without any need for instruction. Even vocabulary is very unsusceptible to enlargement by direct practice aimed at increasing vocabulary. This is part of the reason why vocabulary tests are regarded as such good measures of general intelligence and always have a high _g_ loading in factor analyses of various types of intelligence tests. The items in a vocabulary test are sampled from such an enormously large pool of potential items that the number that can be acquired by specific study and practice is only a small proportion of the total, so that few if any are likely to appear in any given vocabulary test. Furthermore, persons seem to retain only those words which fill some conceptual "slot" or need in their own mental structures. A new word encountered for the first time which fills such a conceptual "slot" is picked up and retained seemingly without conscious effort, and will "pop" into mind again when the conceptual need for it arises, even though in the meantime the word may not have been encountered for many months or even years. If there is no conceptual slot needing to be filled, that is to say, no meaning for the individual which the word serves to symbolize, it is very difficult to make the definition of the word stick in the individual's memory, and even after repeated drill, it will quickly fade beyond retrieval, as when a student memorizes a long list of foreign words in order to pass his foreign language exam for the Ph.D. Since intelligence tests get at the learning that occurs in the total life experiences of the individual, it is a more general and more valid measure of his learning potential than are scholastic achievement tests. It should come as no surprise that there is a substantial correlation between the two classes of tests, since both measure learning or achievement, one in a broad sphere, the other in a much narrower sphere. In a culturally more or less homogeneous population the broader based measure called intelligence is more generally representative of the individual's learning capacities and is more stable over time than the more specific acquisitions of knowledge and skill classed as scholastic achievement.

(2) Equivalence of Diverse Tests. One of the most impressive characteristics of intelligence tests is the great diversity of means by which essentially the same ability (or abilities) can be measured. Tests having very diverse forms, such as vocabulary, block designs, matrices, number series, "odd-man out," figure copying, verbal analogies, and other kinds of problems ca[illegible] .1 serve as intelligence tests yielding more or less equivalent results because of their high intercorrelations. All of these types of tests have high loadings on the *g* factor, which, as Wechsler (1958, p. 121) has said, ". . . involves broad mental organization; it is independent of the modality or contextual structure from which it is elicited; *g* cannot be exclusively identified with any single intellectual ability and for this reason cannot be described in concrete operational terms." We can accurately define *g* only in terms of certain

mathematical operations; in Wechsler's words "*g* is a measure of a collective communality which necessarily emerges from the intercorrelation of any broad sample of mental abilities" (p. 123).

Assessment of scholastic achievement, on the other hand, depends upon tests of narrowly specific acquired skills -- reading, spelling, arithmetic operations, and the like. The forms by means of which one can test any one of these scholastic skills are very limited indeed. This is not to say that there is not a general factor common to all tests of scholastic achievement, but this general factor common to all the tests seems to be quite indistinguishable from the *g* factor of intelligence tests. Achievement tests, however, usually do not have as high *g* loadings as intelligence tests but have higher loadings on group factors such as verbal and numerical ability factors and they also contain more task-specific variance. It is always possible to make achievement tests correlate more highly with intelligence tests by requiring students to reason, to use data provided, and to apply their factual knowledge to the solution of new problems. More than just the mastery of factual information, intelligence is the ability to apply this information in new and different ways. With increasing grade level, achievement tests have more and more variance in common with tests of *g*. For example, once the basic skills in reading have been acquired, reading achievement tests must increasingly measure the student's comprehension of more and more complex selections rather than the simpler processes of word recognition, decoding, etc. And thus at higher grades, tests of reading comprehension, for those children who have already mastered the basic skills, become more or less indistinguishable in factorial composition from the so-called tests of verbal intelligence. Similarly, tests of mechanical arithemtic (arithmetic computation) have less correlation with *g* than tests of arithmetic thought problems, such as the Arithmetic Concepts and Arithmetic Applications subtests of the Stanford Achievement battery. Accordingly, most indices of scholastic performance increasingly reflect general intelligence as children progress in school. We found in our study, for example, that up to grade 6, verbal and nonverbal intelligence tests could be factorially separated, with the scholastic achievement tests lining up on the same factor with verbal intelligence. But beyond grade six both the verbal and nonverbal tests, along with all the scholastic achievement tests, amalgamated into a single large general factor which no form of factor rotation could separate into smaller components distinguishable as verbal intelligence vs. nonverbal intelligence vs. scholastic achievement. By grades 7 and 8 the Lorge-Thorndike Nonverbal IQ and Raven's Progressive Matrices are hardly distinguishable in their factor composition from the tests of scholastic achievement. At the same time it is important to recognize that the Lorge-Thorndike Nonverbal IQ and Raven's Matrices are not measuring scholastic attainment *per se*, as demonstrated by the fact that totally illiterate and unschooled persons can obtain high scores on these tests. Burt (1961), for example, reported the case of separated identical twins with widely differing educational attainments (elementary school education versus a University degree), who differed by only one IQ point on the Progressive Matrices (127 vs. 128).

(3) Heritability of Intelligence and Scholastic Achievement. Another distinguishable characteristic between intelligence and achievement tests is the difference between the heritability values generally found for intelligence and achievement measures. Heritability is a technical term in quantitative genetics referring to the proportion of test score

variance (or any phenotypic variance) attributable to genetic factors. Determinations of the heritability of intelligence test scores range from about .60 to .90, with average values around .70 to .80 (Jensen, 1969). This means that some 70 to 80 percent of the variance in IQs in the European and North American Caucasian population in which these studies have been made is attributable to genetic variance, and only 20 to 30 percent is attributable to nongenetic or environmental variability. The best evidence now available shows a somewhat different picture for measures of scholastic achievement, which on the average have much lower heritability. A review of all twin studies in which heritability was determined by the same methods for intelligence tests and for achievement tests shows an average heritability of .80 for the former and of only .40 for the latter (Jensen, 1967). It is likely that scholastic measures increase in heritability with increasing grade level and that the simpler skills such as reading, spelling, and mechanical arithmetic have lower heritability than the more complex processes such as reading comprehension and arithmetic applications. The reason is quite easy to understand. Simple circumscribed skills can be more easily taught, drilled, and assessed and the degree of their mastery for any individual will be largely a function of the amount of time he spends in being taught and in practicing the skill. Thus children with quite different learning abilities can be shaped up to perform more or less equally in these elemental skills. If Johnny has trouble with his reading or arithmetic or spelling his parents may give him extra tutoring so that he can more nearly approximate the performance of his brighter brother. Siblings in the same family differ considerably less in scholastic achievement than in intelligence. Conversely, identical twins reared apart differ much more in scholastic achievement than in intelligence.

From these facts we conclude that environmental factors make a larger contribution to individual differences in achievement than in intelligence as measured by standard tests.

(4) Maturational Aspects of Intelligence. An important characteristic of the best intelligence test items is that they clearly fall along an age scale. Items are thus "naturally" ordered in difficulty. The Figure Copying Test (see Fig. 1) is a good example. Ability to succeed on a more difficult item in the age scale is not functionally dependent upon success on previous items in the sense that the easier item is a prerequisite component of the more difficult item. By contrast, skill in short division is a component of skill in long division. The age differential for some tasks such as figure copying and the Piagetian conservation tests is so marked as to suggest that they depend upon the sequential maturation of hierarchical neural processes (Jensen, in press). Teaching of the skills before the necessary maturation has occurred is often practically impossible, but after the child has reached a certain age successful performance of the skill occurs without any specific training or practice. The items in scholastic achievement tests do not show this characteristic. For successful performance, the subject must have received explicit instruction in the specific subject matter of the test. The teachability of scholastic subjects is much more obvious than of the kinds of materials that constitute most intelligence tests and especially nonverbal tests.

Cumulative Deficit and the Progressive Achievement Gap

The concept of "cumulative deficit" is fundamental in the assessment of majority-minority differences in educational progress. Cumulative deficit is actually an hypothetical concept intended to explain an observable phenomenon which can be called the "progressive achievement gap" or PAG for short. When two groups show an increasing divergence between their mean scores on tests, there is potential evidence of a PAG. The notion of cumulative deficit attributes the increasing difference between the groups' means to the cumulative effects of scholastic learning such that deficiencies at earlier stages make for greater deficiencies at later stages. If Johnny fails to master addition by the second grade he will be worse off in multiplication in the third grade, and still worse off in division in the fourth grade, and so on. Thus the progressive achievement gap between Johnny and those children who adequately learn each prerequisite for the next educational step is seen as a cumulative deficit. There may be other reasons as well for the PAG, such as differential rates of mental maturation, the changing factorial composition of scholastic tasks such that somewhat different mental abilities are called for at different ages, disillusionment and waning motivation for school work, and so on. Therefore I prefer the term "progressive achievement gap" because it refers to an observable effect and is neutral with respect to its causes.

Absolute and Relative PAG. When the achievement gap is measured in raw score units or in grade scale or age scale units, it is called absolute. For example, we read in the Coleman Report (1966, p. 273) that in the metropolitan areas of the northwest region of the U. S. ". . . the lag of Negro scores (in Verbal ability) in terms of years behind grade level is progressively greater. At grade 6, the average Negro is approximately 1 1/2 years behind the average white. At grade 9, he is approximately 2 1/4 years behind that of the average white. At grade 12, he is approximately 3 1/4 years behind the average white."

When the achievement difference between groups is expressed in standard deviation units, it is called <u>relative</u>. That is to say, the difference is relative to the variation within the criterion group. The Coleman Report, referring to the findings quoted above, goes on to state: "A similar result holds for Negroes in all regions, despite the constant difference in number of standard deviations." Although the absolute white-Negro difference increases with grade in school, the relative difference does not. The Coleman Report states: "Thus in one sense it is meaningful to say the Negroes in the metropolitan North-east are the same distance below the whites at these three grades -- that is, relative to the dispersion of the whites themselves." The Report illustrates this in pointing out that at grade 6 about 15 percent of whites are one standard deviation, or 1 1/2 years, behind the white average; at grade 12, 15 percent of the whites are one standard deviation, or three and a quarter years behind the white average.

It is of course the absolute progressive achievement gap which is observed by teachers and parents, and it becomes increasingly obvious at each higher grade level. But statistically the proper basis for comparing the achievement differences between various subgroups of the school population is in terms of the relative difference, that is, in standard deviation units, called sigma (σ) units for short.

Except in the Southern Regions of the U. S., the Coleman study found a more or less constant difference of approximately one sigma (based on whites in the metropolitan Northeast) between whites and Negroes in Verbal Ability, Reading Comprehension, and Math Achievement. In other words, there was no progressive achievement gap in regions outside the South. In the Southern Regions, there is evidence for a PAG from grade 6 to 12 when the sigma unit is based on the metropolitan Northeast. For example, in the nonmetropolitan South, the mean Negro-white differences

(Verbal Ability) in sigma units are 1.5, 1.7, and 1.9 for grades 6, 9, and 12, respectively. The corresponding number of grade levels that the Southern Negroes lag behind at grades 6, 9, and 12 are 2.5, 3.9, and 5.2 (Coleman, 1966, p. 274). The causes of this progressive achievement gap in the South are not definitely known. Contributing factors could be an actual cumulative deficit in educational skills, true subpopulation differences in the developmental growth rates of the mental abilities relevant to school learning, and selective migration of families of abler students out of the rural South, causing an increasing cumulation of poor students in the higher grades.

Cross-Sectional vs. Longitudinal PAG. Selective migration, student turnover related to adult employment trends, and other factors contributing to changes in the characteristics of the school population may produce a spurious PAG when this is measured by comparisons between grade levels at a single cross section in time. The Coleman Report's grade comparisons are cross sectional. But where there is no reason to suspect systematic regional population changes, cross sectional data should yield approximately the same picture as longitudinal data, which are obtained by repeated testing of the same children at different grades. Longitudinal data provide the least questionable basis for measuring the PAG. Cross sectional achievement data can be made less questionable if there are also socioeconomic ratings on the groups being compared. The lack of any grade-to-grade decrement on the socioeconomic index adds weight to the conclusion that the PAG is not an artifact of the population's characteristics differing across grade levels. (This type of control was used in the present study reported in the following section.)

Another way of looking at the PAG is in terms of the percentage of variance in individual achievement scores accounted for by the mean achievement level of schools or districts. If there is an achievement decrement for, say, a minority group across grade levels, and if the decrement is a result of school influences, then we should expect an increasing correlation between individual students' achievement scores and the school averages. In the data of the Coleman Report, this correlation (expressed as the percentage of variance in individual scores accounted for by the school average) for "verbal achievement" does not change appreciably from the beginning of the first school year up to the 12th grade. The school average for verbal achievement is as highly correlated with individual verbal achievement at the beginning of grade 1 as at grade 12. If the schools themselves contributed to the deficit, one should expect an increasing percentage of the total individual variance to be accounted for by the school average with increasing grade level. But no evidence was found that this state of affairs exists. The percent of total variance in individual verbal achievement accounted for by the mean score of the school, at grades 12 and 1 is as follows (Coleman, et al., 1966, p. 296):

	Grade	
Group	12	1
Negro, South	22.54	23.21
Negro, North	10.92	10.63
White, South	10.11	18.64
White, North	7.84	11.07

Progressive Achievement Gap in a California School District

We searched for evidence of a PAG in our data in several ways, which can be only briefly summarized here. Separate analyses for each of the achievement tests did not reveal any striking differences in PAG, so the results can be combined without distortion of the essential results.

Mean Sigma Differences. The mean difference in sigma (standard deviation) units, based on the white group, by which Negro and Mexican-American pupils fall below the white group at each grade from 1 to 8 is shown in Table 3. The first three columns show the sample sizes on which the sigma differences are based. The sigma differences (i.e.,

- - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - -

σ below white mean) for Negroes and Mexican-Americans shown in columns 4 and 5 is the average of all the Stanford Achievement Tests given in each grade. Note that there is a reliable and systematic increase in the sigma difference from grade 1 to grade 3, for both Negro and Mexican groups, after which there is no further systematic change in achievement gap. The mean gap over all grades is .66σ for the Negroes and .55σ for the Mexicans. By comparison, look at columns 6 and 7, which show the mean sigma differences for those nonverbal ability tests in our battery which do not depend in any way upon reading skill and the content of which is not taught in school; this is the average sigma difference for the Lorge-Thorndike Nonverbal IQ, Figure Coyping, and Raven's Progressive Matrices. We see that the sigma differences show a slight upward trend from the lower to the higher grades. Furthermore, the sigma differences are very significantly larger for the nonverbal intelligence tests than for the scholastic achievement tests in the case of Negroes (1.08σ for nonverbal intelligence vs. 0.66 for achievement). The Mexicans show only a slight difference between their sigma decrement in nonverbal ability and in scholastic achievement (0.63 vs. 0.55). If we can regard

Table 3

Number of White Sigma Units by which Minority Group Means Fall Below the White Mean

	Sample Size (N)			Stanford Achievement Tests		Nonverbal Intelligence		Home Index (SES)		Adjusted Achievement Means	
Grade	White	Negro	Mexican	Negro	Mexican	Negro	Mexican	Negro	Mexican	Negro	Mexican
1	285	218	258	.25	.34	1.07	.53	---	---	-.09	.15
2	229	162	250	.57	.37	1.03	.70	---	---	.15	.06
3	281	207	241	.83	.68	0.98	.53	.58	1.13	.11	.05
4	237	189	239	.69	.59	0.95	.48	.38	1.18	.17	.15
5	242	198	211	.75	.54	1.05	.62	.70	1.18	.21	.10
6	219	169	218	.84	.69	1.23	.67	.47	1.36	.09	.02
7	388	262	305	.71	.57	1.13	.72	.71	1.36	.07	.08
8	356	289	303	.64	.62	1.18	.79	.77	1.34	.06	.08
Mean				.66	.55	1.08	.63	.60	1.26	.10	.09

these nonverbal tests as indices of extrascholastic learning ability, it appears then that these Negro children do relatively better in scholastic learning as measured by the Stanford Achievement Tests than in the extrascholastic learning assessed by the nonverbal battery. In this sense, the Negro pupils, as compared with the Mexican pupils, are "over-achievers," although the Negroes' absolute level of scholastic performance is 0.11σ below the Mexicans'. For the Negro group especially, the school can be regarded as an equalizing influence: Negro pupils are closer to white pupils in scholastic achievement than in nonscholastic, nonverbal abilities. The mean Negro-white scholastic achievement difference is only 61 percent as great as the nonverbal IQ difference. This finding is exactly the opposite of popular belief. The white vs. Mexican achievement difference is 87 percent as great as the nonverbal IQ difference.

Is there any systematic grade trend in our indices of socioeconomic status and home environment? Columns 8 and 9 show the sigma differences below the white group on the composite score of Gough's Home Index, which assesses parental educational and occupational level, physical amenities, cultural advantages, and community involvement. (The Home Index was not used below grade 3.) There is a slight, but not highly regular, upward trend in these sigma differences for both Negro and Mexican groups, as if the students in the higher grades come from somewhat poorer backgrounds. Despite this, the sigmas for scholastic achievement (unlike the nonverbal ability tests) do not show any systematic increase from grade 3 to 8. Note also that on the Home Index the Mexicans, on the average, are further below the Negroes than the Negroes are below the whites. Moreover, the percentage of the Mexican children whose parents speak only English at home is 19.7 percent as compared with 96.5 percent for whites and 98.2 percent for Negroes. In 14.2 percent of the Mexican homes Spanish or other foreign language is spoken exclusively, as compared with 1.1 percent for whites and 0.5 percent for Negroes.

Covariance Adjustments of Achievement Scores. The next step of our analysis consists of obtaining covariance adjusted means on all the achievement tests, using all the ability tests[3], along with sex and age in months, as the covariance controls. What this procedure shows, in effect, is the mean score on the achievement tests ("output") that would be obtained by the three ethnic groups if they were equated on the ability tests ("input"). Although it is beyond the scope of this paper to explain in mathematical detail just how this kind of covariance adjustment is accomplished, a few words of explanation are in order to remove any mystery that may seem to exist for those who have not studied or used this statistical technique. A simplified illustration will give the reader some notion of what is involved.

The simplest possible illustration consists of two groups, say, Negro and white, who are given two tests, say, an IQ test and an achievement test. What we wish to find out is: what would be the mean achievement scores of the Negro and white groups if they were equated on IQ? What we must determine, in statistical terminology, is the "covariance adjusted mean" achievement for each group. It is defined mathematically as

$$\hat{Y} = \overline{Y}_G - b\,(\overline{X}_G - \overline{X}_{..})$$

In terms of our example,

$\hat{Y}_N$ = adjusted mean achievement score of Negro group

$\bar{Y}_N$ = raw mean achievement score of Negro group

$\bar{X}_N$ = mean IQ of Negro group

$\bar{X}_{..}$ = mean IQ of Negro and white groups combined, i.e., total mean IQ.

b = the regression coefficient of Y on X, i.e., of achievement on IQ for both groups combined. The regression coefficient is the slope of the regression line. It is $r_{xy}\frac{\sigma_y}{\sigma_x}$, where, r_{xy} is the correlation between the two variables, X and Y (or IQ and achievement) and σ_x and σ_y are the standard deviations of these variables.

The situation can be pictured as follows:

- - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - -

For the sake of graphic clarity, this is a greatly exaggerated picture. The so-called regression line is the one straight line about which the squared deviations of all scores are a minimum. Thus, every individual score plays a part in determining the position and slope of the regression line. It is the one best-fitting line to the data of <u>all</u> the subjects in both groups. Although the mean raw achievement scores differ markedly for Negroes and whites in this illustration, we see that each group falls only slightly off the common regression line; in this example, the white mean is above the line and the Negro mean is below. The <u>adjusted</u> means for the two groups consist of the grand mean plus (or minus) the deviation of the particular group's mean from the regression line. If the means of both groups fall exactly on the common regression line, the adjusted means will be exactly the

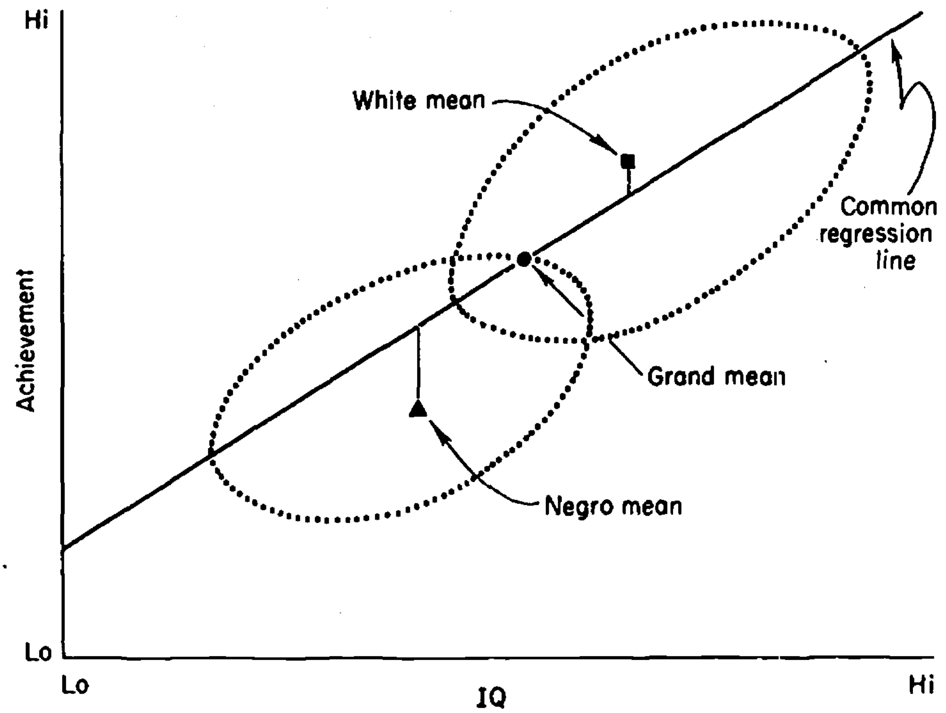

Fig. 2. Simplified correlation scatter diagram illustrating the regression of achievement on IQ and the covariance adjustment of hypothetical white and Negro achievement means.

same and are equal to the grand mean. If there is zero correlation between the input (IQ) and output (achievement) variables, then the regression line will be perfectly horizontal and parallel to the base line, and the adjusted means will consequently be exactly the same as the raw (or unadjusted) means. In the above example, the white adjusted mean would be slightly higher than the Negro adjusted mean, because the white mean is above the regression line and the Negro below. The regression line can be thought of as predicting the most probable achievement score for any given IQ. If the correlation between IQ and achievement were perfect, one could predict achievement from IQ exactly, and vice versa.

The situation is essentially the same for adjusting the means of 3 or more groups, and one can easily picture another group placed in the above illustration. It is much more difficult to picture the situation when more than 2 variables are involved. In this illustration, we have one output variable (achievement) and only one input variable (IQ). It is possible to have 2 or 3 or more input variables. If there are 2, then the situation would have to be pictured in three dimensions. The common regression line would no longer be a line on a 2-dimensional surface but would become a plane in a 3-dimensional cube, and we would be adjusting our means in terms of their deviations from the surface of this 2-dimensional plane. If we go to 3 input variables the situation can no longer be pictured, since we would have to deal with a "hyper plane" in 4-dimensional space. Four input variables require a 5-dimensional space, and so on. Although the problem can no longer be pictured graphically beyond 2 input variables, it can be solved mathematically for any number of input variables (although the point of diminishing returns is rapidly reached). For the sample sizes and the number of input variables used in the present study, the mathematical computations would be virtually impossible without the aid of a high speed computer.

Columns 10 and 11 of Table 3 show the sigma difference by which the Negro and Mexican covariance adjusted mean falls below that of the white group. These differences are quite small for both Negroes and Mexicans (averaging 0.10 and 0.09, respectively), and they show no systematic trend with grade level. In other words, when the minority groups are statistically equated with the majority (white) group on the ability test variables, their achievement, on the average, is less than 0.1 sigma below that of the white group. On an IQ scale that would be equivalent to 1.5 points, a very small difference indeed. The adjusted decrement is statistically significant, however, which raises the question of why it should differ significantly from zero at all. The reason could be actual differences between minority and majority schools in the effectiveness of instruction, or incomplete measurement of all the input variables relevant to scholastic learning, or some lack of what is called homogeneity of regression for the three ethnic groups, which works against the covariance adjustment. We know the latter factor is involved to some extent, and some combination of all of them are most likely involved. But taken all together, the fact that the majority-minority difference in mean adjusted achievement scores is still less than 0.1σ means the direct contribution of the schools to the difference must be even smaller than this, if existent at all. Surely it is of practically negligible magnitude.

When the personality variables (the Junior Eysenck Personality

Inventory) and the four scales of the Home Index are also included with the ability variables in obtaining covariance adjusted means, the ethnic differences in scholastic achievement are wiped out almost entirely. Two-thirds of the majority-minority differences (for various achievement subtests at various grades) are not significant at the 5 percent level and are less than 0.1σ. The adjusted mean differences between ethnic groups are smaller than the grade-to-grade sigma differences within ethnic groups. From this analysis, then, the school's contribution to ethnic achievement differences must be regarded as nil. If the input variables themselves are strongly influenced by the school to the disadvantage of the minority children, we should expect to find a greater sigma difference for nonverbal IQ at grade 8 than at Kindergarten. In the present study Negroes are 1.11σ below whites in nonverbal IQ in Kindergarten as compared with 1.17σ in Grades 7 and 8 -- a trivial difference. Mexican children are 0.98σ below whites in nonverbal IQ at Kindergarten and .88σ below at grades 7 and 8. Thus the minority children begin school at least as far below the majority children in nonverbal ability as they are by grades 7 and 8. The schools have not depressed the ability level of minority children relative to the majority, but neither have they done anything to raise it. Differences in verbal IQ are slightly more likely to reflect the effects of schooling, and we note that in grades 7 and 8 Negroes are 1.00σ below the white mean and Mexicans are 0.90σ below.

Paired Ethnic Group Differences. The maximum discrimination that we can make between the three ethnic groups in terms of all of our "input" variables (ability tests, personality inventories, and socioeconomic indexes) is achieved by means of the multiple point-biserial correlation coefficient. The product-moment correlation obtained between a continuous variable (e.g., IQ) and a quantized (dichotomous) variable (e.g., male vs. female, where male = 1 and female = 0) is called a point-biserial correlation (r_{pbs}). Mathematically it is defined as:

$$r_{pbs} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_t} \sqrt{pq}$$

where $\bar{X}_1$ and $\bar{X}_2$ = means of groups 1 and 2

σ_t = standard deviation of total (i.e., groups 1 and 2 combined)

p and q = proportions of total sample in groups 1 and 2, respectively. ($p + q = 1.00$)

It is also possible to compute r_{pbs} in the same manner that one computes the Pearson product-moment correlation between any two continuous variables, except that the dichotomous variable is quantized by assigning 0 and 1 to its two categories. It is also possible to obtain a multiple point-biserial correlation, which gives the maximum possible correlation between the quantized variable and the best weighted combination of a number of "predictor" variables. The multiple correlation thus represents the maximum degree of discrimination that can be achieved between the two categories of the quantized variable by means of the particular set of predictor variables. Since the multiple correlation capitalizes upon sampling error (chance deviations from population values) to achieve the maximum value of the correlation, it is spuriously inflated by a degree that is inversely proportional to the sample size and the number of variables correlated. For this reason, the obtained multiple correlation should be "shrunken" down to its estimated population value (i.e., its value if there were no sampling error). The method for doing this is given in most statistics textbooks (e.g., Guilford, 1956, pp. 398-399). All the multiple correlations reported here have thus been "shrunken" and therefore represent a conservative estimate of the amount of discrimination achieved between the ethnic groups by our battery of "input" tests.

When the sizes of the samples entering into the quantized variable are large and nearly equal, and when they have nearly equal standard deviations on the predictor variables, it is possible roughly to "translate" the point-biserial correlation into a linear mean distance in constant sigma units between the two categories of the quantized variable. Figure 3 shows the function relating the point-biserial correlation to the mean sigma difference ($\bar{d}$) between groups. The r_{pbs} can attain a value of 1.00 only if the variance within each group diminishes to zero.

Insert Figure 3 about here

Table 4 gives the multiple point-biserial correlations between each ethnic dichotomy and all the "input" variables -- first just the ability tests and second the ability tests plus the personality inventory and socioeconomic index. Note that the three groups are almost equally

Insert Table 4 about here

discriminable from one another in terms of the multiple correlation, especially after the personality and social background variables are added to the predictors. This is interesting, because it means that the two minority groups, though both are regarded as educationally and socioeconomically disadvantaged, actually differ from one another on

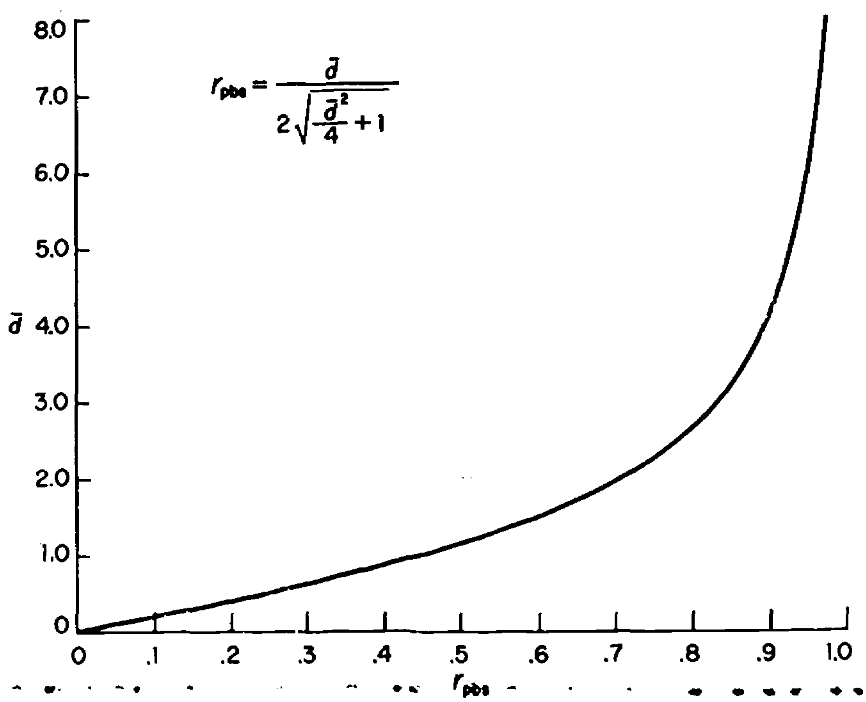

Fig. 3. The relationship between the point biserial correlation (r_{pbs}) and the mean difference ($\bar{d}$) between groups in sigma units on the continuous variable, assuming equal sigmas and equal Ns in the two groups.

Table 4

Point-Biserial Multiple Correlations for "Input" Variables

and Partial Correlation for "Output" with "Input" Held Constant

	"Input" All Ability Tests			"Input" Ability + Personality + Home Index			"Output" Stanford Achievement Minus All "Input" Variables		
Grade	W-N	W-M	M-N	W-N	W-M	M-N	W-N [2]	W-M	M-N
1	.49	.28	.29						
2	.54	.47	.37						
3	.54	.45	.35	.62	.59	.46	.06	.02	.07
4	.48	.38	.41	.55	.60	.55	.15	.07	.09
5	.47	.38	.27	.60	.59	.36	.13	.05	.11
6	.53	.47	.42	.69	.67	.59	.14	.11	.04
7	.52	.42	.26	.68	.70	.45	.09	-.04	.11
8	.57	.42	.43	.65	.66	.46	.06	-.02	-.07
Mean	.52	.41	.36	.63	.64	.48	.11	.05	.07

[1] Partial correlations of less than 0.10 are not significant at the 5 percent level.

[2] The quantized ethnic groups are White = 3, Mexican = 2, Negro = 1, so that for W-N and W-M positive correlations indicate higher achievement scores for the white group, and a positive correlation for M-N indicates higher scores for the Mexican group.

this composite of all input variables almost as much as each one differs from the majority group. The Negro and Mexican groups each differ from the majority group in a somewhat different way in terms of total pattern of scores, and they differ from one another almost as much. A factor analysis, shown in the next section, helps to reveal the ways in which the three groups differ from one another.

The last three columns in Table 4 show the correlation between each ethnic dichotomy and the Stanford Achievement Tests, with all the "input" variables partialed out, i.e., statistically held constant. These correlations represent the average contribution made to the ethnic discrimination by the Stanford Achievement Tests regarded independently of the "input" variables. It can be seen that these correlations are very small indeed. For the sample sizes used here, correlations of less than 0.10 can be regarded as statistically nonsignificant at the 5 percent level. The proportion of the total variance between the ethnic groups that is accounted for by the achievement tests is represented by the square of the correlation coefficient. Applied to the partial correlations for the Achievement Tests in Table 4, this shows how trifling are the ethnic group achievement differences after the ethnic group differences on the input variablea have been controlled.

Factor Analysis of All Variables. A factor analysis (varimax rotation of the principal components having Eigenvalues greater than 1) was carried out at each grade level on all test variables obtained at that grade level plus three others: sex, chronological age in months, and welfare status of the parent (whether receiving welfare aid to dependent children). The latter variable was added to supplement the indices of socioeconomic status (the four scales of Gough's Home Index).

Since grades 4, 5, and 6 had all the measures (27 variables) and the same tests were used at each of these grades, they are the most suitable part of our total sample for factor analytic comparisons. The results are essentially the same at all grade levels, although because the personality inventory and the Home Index were not used in the primary grades, and the Figure Copying Test was not used beyond grade 6, not all of the factors that emerged at grades 4, 5, and 6 come out at one or another of the other grades. Moreover, because of the large number of variables entering into the analysis at grades 4-6, more small factors come out which, in a sense, "purify" the main factors by partialing out other irrelevant and minor sources of variance.

Factor analyses were performed first on the three ethnic groups separately to determine if essentially the same varimax factors emerged in each group. They did. All three groups yield the same factors, with only small differences in the loadings of various tests. This finding justifies combining all three groups for an overall factor analysis of the total student sample at each grade level. This was done. Eight factors with Eigenvalues greater than 1 emerged at grades 4, 5, and 6, accounting respectively for 67%, 66%, and 70% of the total variance.

The first principal component can be regarded as the general or *g* factor for this set of 27 variables. Table 5 shows the loadings of each of the 27 (or 25 in grades 7 and 8) variables on the first principal component in grades 4 to 6. The first principal component is the single

Insert Table 5 about here

most general factor accounting for more of the variance than any other

Table 5

Loadings of Variables on First Principal Component

for Grades 4 to 8 (Decimals Omitted)

	Variable	Grade 4	5	6	7	8
1.	Sex (M = 0, F = 1)	14	14	03	08	12
2.	Extraversion	25	28	46	33	24
3.	Neuroticism	00	-06	-21	-12	01
4.	Lie Scale	-17	-11	-19	-27	-39
5.	Home Index - 1	31	45	41	49	48
6.	Home Index - 2	29	30	34	41	45
7.	Home Index - 3	36	41	27	50	44
8.	Home Index - 4	29	43	28	47	40
9.	Aid to Dependent Children	-21	-43	-32	-31	-26
10.	Age in Months	-05	-09	-04	-04	-12
11.	Lorge-Thorndike Verbal IQ	85	88	85	88	87
12.	Lorge-Thorndike Nonverbal IQ	73	75	76	79	83
13.	Raven's Progressive Matrices	54	55	54	54	63
14.	Figure Copying	45	51	57	--	--
15.	Listening-Attention	11	19	21	06	12
16.	Memory - Immediate	45	40	36	27	32
17.	Memory - Repeat	44	33	24	25	27
18.	Memory - Delayed	43	41	41	25	27
19.	Making X's 1st Try	14	02	31	53	10
20.	Making X's 2nd Try	19	14	29	48	19
21.	SAT: Word Meaning	83	81	81	--	--
22.	SAT: Paragraph Meaning	80	79	89	86	83
23.	SAT: Spelling	75	76	78	73	73
24.	SAT: Language	83	84	87	78	75
25.	SAT: Arithmetic Computation	57	45	63	73	73
26.	SAT: Arithmetic Concepts	72	62	80	76	83
27.	SAT: Arithmetic Applications	77	71	82	72	71
	Percent of Variance	22	26	29	28	27

factor. It is most heavily loaded in the Stanford Achievement Tests and Verbal IQ. Inspection of the loadings of the other variables gives an indication of their correlation with this most general achievement factor.

The eight principal components were rotated to approximate simple structure by the varimax criterion. In grades 4, 5, and 6 four substantial and clear-cut factors emerged. The remaining factors serve mainly to pull out irrelevant variance from the main factors. The four main factors that emerge are:

Factor I. Scholastic Achievement and Verbal Intelligence.

Variables	Factor Loading		
	Gr. 4	Gr. 5	Gr. 6
Lorge-Thorndike Verbal IQ	.75	.75	.85
Word Meaning	.83	.69	.82
Paragraph Meaning	.83	.77	.89
Spelling	.82	.77	.81
Language	.82	.79	.86
Arithmetic Computation	.64	.58	.65
Arithmetic Concepts	.73	.69	.83
Arithmetic Applications	.77	.71	.85

Factor II. Nonverbal Intelligence.

Variables	Factor Loading		
	Gr. 4	Gr. 5	Gr. 6
Lorge-Thorndike Nonverbal IQ	.61	.57	.32
Raven's Progressive Matrices	.75	.75	.55
Figure Copying	.69	.68	.41

Factor III. Rote Memory Ability

Variables	Factor Loading Gr. 4	Gr. 5	Gr. 6
Memory Span - Immediate Recall	.85	.81	.77
Memory Span - Repeated Series	.85	.81	.86
Memory Span - Delayed Recall	.83	.79	.74

Factor IV. Socioeconomic Status.

Variables	Factor Loading Gr. 4	Gr. 5	Gr. 6
Home Index:			
1. Parental Education & Occupation	.75	.74	.77
2. Physical Amenities	.69	.77	.72
3. Community Participation	.66	.76	.75
4. Cultural Advantages	.66	.59	.66
Receives Welfare Aid to Dependent Children	-.40	-.34	-.46

The remaining four minor factors are (1) Speed, motivation, persistence as defined principally by the Making X's Test, (2) Neuroticism, (3) Extraversion, (4) Age in months. These variables, having their largest loadings on separate factors, are in effect partialed out of the major factors. The four major factors listed above are orthogonal, i.e., uncorrelated with one another, and each one is thus viewed as a "pure" measure of the particular factor in the sense that the effects of all the other factors are held constant.

Ethnic Group Comparisons of Factor Scores. The final step was to obtain factor scores for every student on each of these four main factors. For the total sample, within each grade, these factor scores are represented on a T-score scale, i.e., they have an overall mean of 50 and a standard deviation of 10. Table 6 shows the mean and standard deviation of the factor scores for each of the ethnic groups.

- - - - - - - - - - - - - -
Insert Table 6 about here
- - - - - - - - - - - - - -

Note that the ethnic group differences in Factor I do not show any systematic increase from grade 4 to 6, thus lending no support to the existence of a cumulative deficit in the minority groups. Analysis of variance was performed on the factor scores and Schaffé's method of contrasts was used for testing the statistical significance of the differences between the means of the various ethnic groups at each grade level. The results of these significance tests are shown in Table 7. We see that in Factor I (Verbal IQ and Scholastic Achievement)

- - - - - - - - - - - - - -
Insert Table 7 about here
- - - - - - - - - - - - - -

both minority groups are significantly below the majority group, and Negroes are significantly below the Mexican group except in grade 6, where the difference is in the same direction but falls short of significance.

On Factor II (Nonverbal Intelligence) Negroes fall significantly below whites and Mexicans at all grades, and the differences between Mexicans and whites are nonsignificant at all grades. It should be remembered that this nonverbal intelligence factor represents that part of the variance in the nonverbal tests which is not common to the verbal IQ and achievement tests or to the memory tests. The Mexican-white difference is significant on that part of the ability tests variance which has most in common with scholastic achievement and is represented in Factor I.

Table 6

Mean Varimax Factor Scores for Three Ethnic Groups

in Grades 4, 5, and 6

			Mean Factor Scores							
			I		II		III		IV	
			Verbal IQ & Achievement		Nonverbal IQ		Memory		Socioeconomic Status	
Grade	Group	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	White	113	55.2	10.7	51.6	8.1	51.6	9.4	53.8	10.3
4	Negro	129	47.1	6.5	44.6	8.9	51.0	11.2	51.7	7.9
	Mexican	145	49.5	8.5	51.0	9.3	48.1	7.7	43.6	7.8
	White	144	54.7	8.7	52.3	8.2	50.4	9.1	54.1	9.2
5	Negro	132	45.5	8.4	47.0	11.1	51.1	9.9	49.7	9.5
	Mexican	135	49.6	8.5	50.1	8.5	48.2	9.5	44.6	8.1
	White	131	55.0	8.8	50.9	7.2	50.7	8.8	53.8	9.4
6	Negro	124	47.1	8.3	44.1	10.5	50.5	9.9	51.5	8.0
	Mexican	126	49.1	9.3	51.0	8.7	48.0	10.2	42.5	7.5

Table 7

The Significance of Ethnic Group Differences in

Mean Factor Scores, by Scheffé's Method of Contrasts

		Factors			
		I	II	III	IV
Contrasts (Means)	Grade	Verbal IQ & Achievement	Nonverbal Intelligence	Memory	Socioeconomic Status
Negro - White	4	-**	-**	- n.s.	- n.s.
	5	-**	-**	+ n.s.	-**
	6	-**	-**	- n.s.	- n.s.
Mexican - White	4	-**	- n.s.	-*	-**
	5	-**	- n.s.	- n.s.	-**
	6	-**	+ n.s.	- n.s.	-**
Mexican - Negro	4	+*	+**	-*	-**
	5	+**	+*	-*	-**
	6	+ n.s.	+**	- n.s.	-**

*$p < .05$ n.s. = Not Significant

**$p < .01$

Factor III (Rote Memory) shows no significant differences between the Negro and white groups; the Mexican group is significantly below the white at grade 4 and below the Negro at grades 4 and 5. This finding is consistent with the findings of other studies that mean differences between groups of lower and middle socioeconomic status are smallest on tests of short-term memory and rote learning (Jensen, 1968).

Factor IV (socioeconomic status) shows relatively small differences between the Negro and white groups, while the Mexican group is significantly below the other two. Again, it should be realized that we are dealing here with "pure" factor scores which are independent of all the other variables. Thus Factor IV shows us the relative standing of the three ethnic groups in socioeconomic status when all the other variables are held constant. What these results indicate is that Negro and white children statistically equated for intelligence, achievement, and memory ability differ very little in socioeconomic status as measured by our indices, but that Mexican children, when equated on all other variables with white children or with Negro children, show a comparatively much poorer background than either the white or Negro groups. On the present measures, at least, the Mexicans must be regarded as much more environmentally disadvantaged than the Negroes, and this takes no account of the Mexican's bilingual problem. In view of this it is quite interesting that Mexican pupils on the average significantly exceed the Negro pupils in both verbal and nonverbal intelligence measures and in scholastic achievement.

Equality of Educational Opportunity: Uniformity or Diversity of Instruction?

The results of our analysis thus fai 'ail to support the hypothesis that the schools have discriminated unfavorably against minority pupils. When minority pupils are statistically equated with majority children for background and ability factors over which the schools have little or no control, the minority chidren perform scholastically about as well as the majority children. The notion that poor scholastic achievement is partly a result of the pupil's ethnic minority status per se, implying discriminatory schooling, is thus throughly falsified by the present study. This does not imply that the same results would be obtained in every other school system in the country. Where true educational inequalities between majority and minority pupils exist, we should expect the present type of analyses to reveal these inequalities, and it would be surprising if they were not found in some school systems which provide markedly inferior educational facilities for minority pupils. It should be noted, on the other hand, that the present study was conducted in a school district which had taken pains to equalize educational facilities in schools that serve predominantly majority or predominantly minority populations. The success of this equalization is evinced in the results of the present analyses.

But we can take a bold step further and ask: Is equalization of educational facilities enough? Is the real meaning of equality of educational opportunity simply uniformity of facilities and instructional programs? Is it possible that true equality of opportunity could mean doing whatever is necessary to maximize the scholastic achievement of children, even if it might mean doing quite different things for different children in terms of their differing patterns of ability? Note that I did not say in terms of their ethnic or social class status, but in terms of their individual patterns of ability. The fact that different social class and ethnic groups show different modal patterns of ability, of course, means different proportions of various subpopulations will have different patterns of strengths and weakness in various mental abilities. Is such a fact to be deplored and swept out of sight, or should it be examined with a view to utilizing the differences in the design of instructional programs that might maximize each individual's benefits from schooling? A couple of years ago I wrote: "If we fail to take account either of innate or acquired differences in abilities and traits, the ideal of equality of educational opportunity can too easily be interpreted so literally as to be actually harmful, just as it would be harmful for a physician to give all his patients the same medicine. One child's opportunity can be another's defeat" (Jensen, 1968a, p. 3). At that time I suggested that we look for differential ability patterns that might interact with different instructional methods in such a way as to maximize school learning for all individuals and at the same time minimize individual and group differences in scholastic achievement and any other benefits derived from schooling.

In our laboratory research we have discovered two broad classes of abilities which show marked differences in their relation to social class and race (Jensen, 1968b, 1968d, 1970; Jensen & Rohwer, 1968, 1970). Briefly, what we have found is that children of low socioeconomic status, especially minority children, with low measured IQs (60 to 80) are generally superior to their middle-class counterparts in IQ on tests of associative learning ability: free recall, serial rote learning, paired-associates learning, and digit span memory. This finding has been interpreted theoretically in terms of a hierarchical model of mental abilities, going from associative learning to conceptual thinking, in which the development of lower levels in the hierarchy is necessary but not sufficient for the development of higher levels. Our hypothesis states that the continuum of tests going from associative to conceptual is the phenotypic expression of two functionally dependent but genotypically independent types of mental processes, which we call Level I and Level II. Level I processes are perhaps best measured by tests such as digit span and serial rote learning; Level II processes are represented in tests such as the Progressive Matrices. Level I and Level II abilities are distributed differently in upper and lower social classes and in different ethnic groups. Level I is distributed fairly evenly in all subpopulations. Level II, however, is distributed about a higher mean in upper than in lower social classes. The majority of children now called culturally disadvantaged show little or no deficiency in Level I ability but are about one standard deviation below the general population mean on tests of Level II ability. Children who are above average on Level I but below average on Level II ability usually appear to be bright and capable of normal learning and achievement in many life situations, although they have unusual difficulties in school work under the traditional methods of classroom instruction. Many of these children, who may be classed as retarded in school, suddenly become socially adequate persons when they leave the academic situation. But children who are below average on both Level I and Level II seem to be much more handicapped. Not only is their scholastic performance poor, but their social and vocational potential also seem to be much less than those of children with normal Level I functions. Yet both types of children look much alike in overall measures of IQ and scholastic achievement.

These findings are important because they help to localize the nature of the intellectual deficit of many children called culturally disadvantaged. We must ask whether we can discover or invent instructional methods that engage Level I more fully and thereby provide a means of improving the educational attainments of many of the children now called culturally disadvantaged? In our current instructional procedure are we utilizing so exclusively those mental abilities we identify as IQ (Level II) that children who are relatively low in IQ but have strength in other abilities are unduly disadvantaged in the traditional classroom? The whole complex process of classroom instruction as we know it has evolved in relation to a relatively small upper-class segment of Anglo-European stock. The modal pattern of development in learning abilities of this group has probably shaped to a considerable degree the particular educational procedures public education has long regarded as standard for everyone, regardless of differences in cultural background or inherited patterns of ability. But so far we have not successfully met the challenge presented by our ideal of a rewarding education for all segments of the population, with their diverse patterns of ability.

Looking, for example, at the factor scores shown in Table 6 we note that the minority groups are not significantly below the majority group on Factor III (Memory), which we would identify with Level I ability. Lest anyone try to argue that these "pure" factor scores do not correspond to any "impure" scores that could be obtained with actual tests, we can look at Figures 4 and 5, showing the grade-to-grade growth curves of a good Level II test (Raven's Progressive Matrices) and a good Level I test (a composite of the three digit memory tests).

Insert Figures 4 and 5 about here

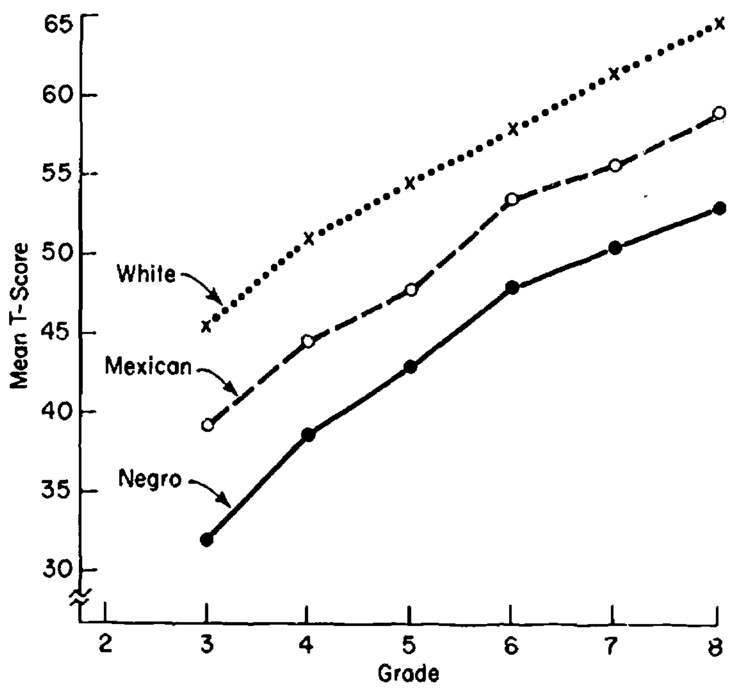

Fig. 4. Mean T scores ($\bar{X}$ = 50, SD = 10) on Raven's Progressive Matrices in Grades 3 to 8.

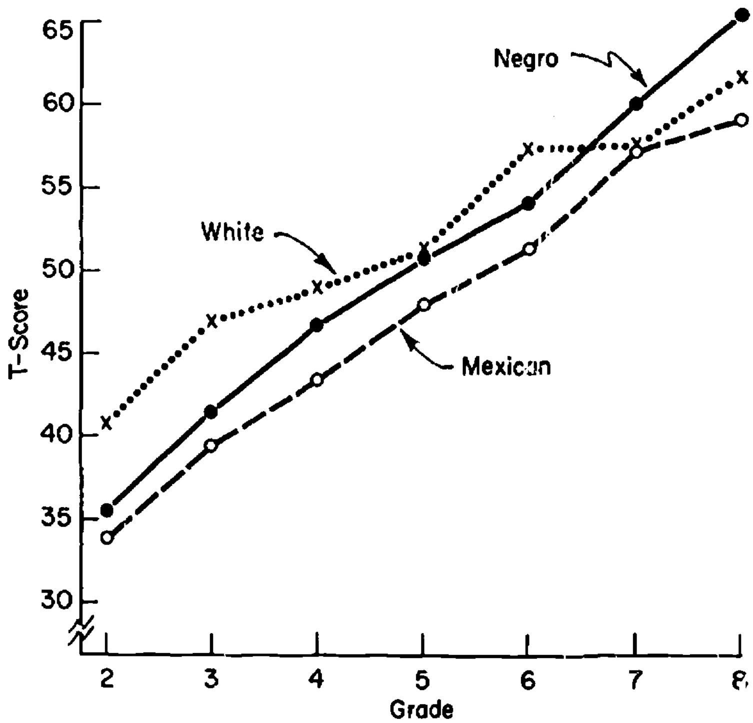

Fig. 5. Mean T scores ($\bar{X}$ = 50, SD = 10) on composite Memory score in Grades 2 to 8.

The results of both tests have been put on the same scale of T scores, with an overall mean of 50 and a standard deviation of 10 (based on the standard deviation of raw scores in the white group at grade 5). The differences between the growth curves shown in Figures 4 and 5 are striking. The approximately one standard deviation difference between the Negro and white groups on the Level II test (Matrices) can be seen to have rather drastic implications in terms of grade level comparisons. By drawing a horizontal line from the Negro or Mexican mean at any grade to the point where it crosses the curve for the white group and dropping a perpendicular to the baseline, we can read off the grade equivalent of the minority group mean. The average Negro 8th grader in this school system, for example, performs on the matrices at a level equivalent to white children at grade 4.5. Mexican children at grade 8 perform at grade 6.3. The grade 6 performance of Negroes and Mexicans is equivalent to the white's performance in grades 3.4 and 4.5, respectively.

On the other hand, note the small differences between the groups on the Level I test (Memory Span) in Figure 5. It is interesting to conjecture whether instruction in scholastic skills specifically aimed at Level I ability in children who are low in Level II would significantly reduce majority-minority differences in scholastic achievement. We do not know and can find out only through further research. If instruction is aimed only at Level II ability for all children,we should expect sizeable majority-minority differences in achievement. If instruction could somehow be aimed at Level I ability for all those children (regardless of ethnic identification)who are significantly stronger in Level I than in Level II, would their achievement be brought appreciably closer to that of the majority? Or is scholastic learning so intrinsically dependent on Level II ability that no form of instruction attempting to capitalize on Level I ability could possibly succeed beyond the most elementary aspects of any academic subject matter? Again, we do not know. But until these possibilities are explored, schools may be accused of cheating many children, especially large numbers of minority children, by providing uniform facilities but not sufficiently diversified instructional programs to minimize differences in achievement and also maximize the overall level of achievement.

Some scholastic subjects would seem to lend themselves more to Level I processes and instructional methods than other subjects. For instance, the learning of spelling and arithmetic computation would seem to be less dependent upon Level II ability than, say, reading comprehension, arithmetic concepts or arithmetic applications. If this is true, we should expect majority-minority differences to be smaller on the Level I types of subject matter than on the Level II types. Let us make the relevant comparisons in the data of the present study. Table 8 shows these comparisons in sigma units. They bear out our hypothesis; the pupils of

- - - - - - - - - - - - - -
Insert Table 8 about here
- - - - - - - - - - - - - -

both minority groups fall below the majority mean about one-fifth of a sigma more on Level II-like scholastic achievement than on Level I-like subjects. Clearly, school subjects which by their nature seem to permit greater utilization of Level I ability show smaller majority-minority differences than those subjects which involve more Level II ability. This raises the interesting question whether all scholastic subjects can be taught in ways that maximize their dependence on Level I and minimize their dependence on Level II. If this can be done for children who are low in Level II ability -- and we will never know without trying --

Table 8

Mean Sigmas (Based on White Group[1]) Below White Mean of Negro and Mexican Pupils in Grades 4-8 on Level I-Like and Level II-Like Tests of Scholastic Achievement

Tests	Negro (N=1,107)	Mexican (N=1,276)
Level I-Like Tests:		
Spelling	.62	.52
Arithmetic Computation	.56	.36
Level II-Like Tests:		
Paragraph Meaning	.90	.75
Arithmetic Concepts	.71	.60
Arithmetic Applications	.72	.55

it should reduce not only the scholastic achievement gap between majority and minority children but the achievement differences among all children of every group. If it succeeds, it would do so, not by pulling anyone down toward the common average, but by capitalizing on each child's particular strengths and minimizing the role of his particular weaknesses in learning any given kind of subject matter. This would seem to be an avenue worth exploring in our efforts to achieve not only equality of educational opportunity but greater equality of scholastic performance as well.

References

Bowles, S., & Levin, H. M. The determinants of scholastic achievement -- an appraisal of some recent evidence. Journal of Human Resources, Winter, 1968.

Burt, C. The gifted child. British Journal of Statistical Psychology, 1961, 14, 123-139.

Coleman, J. S., et al., Equality of educational opportunity. U. S. Office of Education, Washington, D. C., 1966.

Gough, H. G. A short social status inventory. Journal of Educational Psychology, 1949, 40, 52-56.

Guilford, J. P. Fundamental statistics in psychology and education. 3rd Edition. New York: McGraw-Hill, 1956.

Ilg, Frances L., & Ames, Louise Bates. School readiness: Behavior tests used at the Gesell Institute. New York: Harper & Row, 1964.

Jensen, A. R. Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. Proceedings of the National Academy of Sciences, 1967, 58, 149-157.

Jensen, A. R. Social class, race, and genetics: Implications for education. American Educational Research Journal, 1968, 5, 1-42. (a)

Jensen, A. R. Patterns of mental ability and socioeconomic status. Proceedings of the National Academy of Sciences, 1968, 60, 1330-1337. (b)

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123. (a)

Jensen, A. R. Intelligence, learning ability, and socioeconomic status. Journal of Special Education, 1969, 3, 23-35. (b)

Jensen, A. R. A theory of primary and secondary mental retardation. In Ellis, N. R. (Ed.), International review of research in mental retardation. Vol. IV. New York: Academic Press, 1970.

Jensen, A. R. Hierarchical theories of mental ability. In Butcher, H. J. (Ed.), Research on intelligence. London: Methuen, in press.

Jensen, A. R. & Rohwer, W. D., Jr. Mental retardation, mental age, and learning rate. Journal of Educational Psychology, 1968, 59, 402-403.

Jensen, A. R., & Rohwer, W. D., Jr. An experimental analysis of learning abilities in culturally disadvantaged children. Final Report, Office of Economic Opportunity. Contract No. OE02404, 1970.

Thorndike, R. L. Community variables as predictors of intelligence and academic achievement. Journal of Educational Psychology, 1951, 42, 321-338.

Wechsler, D. The measurement and appraisal of adult intelligence. 4th Edition. Baltimore: Williams & Wilkins, 1958.

Wilson, A. B. Educational consequences of segregation in a California community. Racial isolation in the public schools, Appendices, Vol. II, p. 185. U.S. Commission on Civil Rights, Washington, D.C., 1967.

Wiseman, S. Education and environment. Manchester University Press, 1964.

Footnotes

[1]Alameda, Contra Costa, Marin, Napa, San Francisco, San Joaquin, San Mateo, Santa Clara, Solano, Sonoma.

[2]A smaller rank order (e.g., 1) indicates: high reading scores, high median IQ, high proportion of minorities, high expenditure per child, high teacher salaries, high tax rate, high teacher/pupil ratio (i.e., smaller classes), and a larger number of administrators per 100 pupils.

[3]Lorge-Thorndike Verbal and Nonverbal IQ, Figure Copying, Raven's Matrices, Making X's, Listening-Attention, and three memory tests.