

AN EXPLORATORY INVESTIGATION OF ETHNIC DIFFERENCES
WITHIN AN INDUSTRIAL SELECTION BATTERY

A Thesis
Presented to
the Faculty of the Department of Psychology
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Clay Leon Moore, Jr.
June 1966

363717

AN EXPLORATORY INVESTIGATION OF ETHNIC DIFFERENCES
WITHIN AN INDUSTRIAL SELECTION BATTERY

An Abstract of a Thesis
Presented to
the Faculty of the Department of Psychology
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Clay Leon Moore, Jr.
June 1966

An Abstract

This study was a comparison of the test performance of a group of ninety-one white applicants for employment in a petroleum refinery and a group of ninety-one Negro applicants, matched with respect to age and years of formal education.

The employment selection battery consisted of five tests: Test of Learning Ability; Test of Non-Verbal Reasoning; Test of Mechanical Comprehension, Form BB; Test of Chemical Comprehension; and the Advanced California Achievement Test Form W, math section only. Each of the tests had cut-off scores set at approximately the 70th percentile based on company norms. Each applicant to be considered for the interviewing step in the employment process had to exceed the cut-off score on each test.

Although the battery was not intended to be factorially pure, it can be said to provide a rough measure of two factors: a spatial reasoning factor (Non-Verbal Reasoning and Mechanical Comprehension) and a verbal factor (California Achievement and Chemical Comprehension). The Learning Ability test contains items measuring both factors.

The data indicate that matching on age and years of formal educational attainment does not eliminate the observed differences in test performance between ethnic groups since t tests were significant at the .001 level for all tests in the battery.

While a higher percentage of Negroes than whites failed each of the tests, the difference between these percentages for Negro and white was greatest in the case of the tests which appear to have a high spatial factor loading.

It was also observed that the difference between the mean scores for the Negroes and for the whites decreased as education increased, but that this decrease was not to the same degree for each of the tests. A statistic, PI, was computed for each test representing the percentage increase of the white mean score over the Negro mean score. On this basis the Mechanical Comprehension and Non-Verbal Reasoning tests show the largest increase, the California Achievement and Learning Ability tests are intermediate and the Chemical Comprehension test shows the smallest increase.

The major hypothesis of this study was rejected: there are significant differences between Negro and white test performance even when reported education is controlled. The second hypothesis was also rejected. It appears that tests heavily loaded with a spatial factor contribute more to the differences between Negro and white than do verbally oriented tests.

Table of Contents

Chapter		Page
1	Introduction	1
2	Review of the Literature	10
3	Procedures and Design	17
	Subjects	17
	Description of the Selection Battery	17
	Administration of the Tests	19
	Study Design	19
4	Results	21
	Comparison of Test Intercorrelations Across Ethnic Groups	21
	Comparison of Mean Test Scores Across Ethnic Groups	26
	Differential Effects of Selection Tests Across Ethnic Groups	30
5	Summary and Conclusions	35
	References	39

List of Tables

Table	Page
1 Comparison of White and Negro Differences on the Wechsler Bellevue Scale	14
2 Intercorrelations of Selection Test Battery for All Applicants	22
3 Intercorrelations of Selection Test Battery for All White Applicants	23
4 Intercorrelations of Selection Test Battery for Negro Applicants	24
5 Intercorrelations of Selection Test Battery for Matched Sample of Whites	25
6 Means and Standard Deviations of Matched Ethnic Groups for All Tests of Selection Battery . . .	27
7 Means and Standard Deviations on All Tests by Educational Category Across Ethnic Groups . . .	29
8 Comparison of Empirically Derived Order of Difficulty for Each Test Across Ethnic Groups .	31
9 Comparison of White Increase of Mean Score Over Mean Negro Test Score Across Educational Categories	33

Chapter 1

Introduction

Much has been written in recent years, both in popular literature and in scientific journals, concerning the various aspects of the role of the Negro in contemporary American society. This is largely a reflection of increasing pressures developing within our present social system. One of the significant areas for psychology is that dealing with psychological differences between white and Negro members of the population, particularly as measured by psychological tests.

Ingle presents one statement of the issue in two articles in Science (1964). He states:

- (1) The question as to whether the average differences among the races in test performance, school achievement, and behavior have a genetic as well as environmental basis is unresolved.
- (2) The issue is important and should be studied as a means to understanding the causes of social problems and correcting them.
- (3) It is time to propose, debate, and test by pilot studies means of preventing social problems, rather than to depend upon palliative methods.

The exchange of opinion and counterarguments continued in Science for several months. Various authors argued with Ingle's use of the word "race" and his meaning of intelligence. Most of the counterarguments were concerned with Ingle's social-political interpretations and applications of the data he presented on racial intelligence. At the

end of the battle one fact stood out clearly. There is a variety of opinion and seemingly very few definitive facts on the question of intelligence and racial differences.

Disregarding Ingle's social-political interpretations and suggestions, his suggestion for more scientific investigation into the area of racial differences in intelligence is appropriate. The fact that there was such a response to his views and the diversity of ideas expressed attest to the need for more intensive research.

Of special concern is the use of psychological tests as tools for these suggested studies. Ingle states (1964, p. 376) that psychological tests are not culture-free and to evaluate the question of intelligence and racial differences "we need valid, culture-free measures of intelligence." This is not a new suggestion. The early 50's saw the intensive research effort of Kenneth Eells and others to seek the answers to this general question. Unfortunately the trend did not last for an extended period of time. With the advent of the major Negro push for equal rights in the early 60's and up to the present, criticisms of testing and similar evaluations as a form of discrimination have increased. There has been, however, no substantial increase in interest by psychologists to evaluate these criticisms until recently.

There have been studies to determine what the public thinks about psychological testing. Carter (1964) summarizes

evidence presented at a symposium on psychological tests and public responsibility. In this investigation 1,500 adults and 10,000 high school students were studied. Forty-one per cent of a representative group of Americans over eighteen years of age were opposed to using intelligence tests for admission to college. Thirty-seven per cent were against using tests in job selection and 50 per cent against using them in job promotion. The high school student group felt more strongly. Fifty-four per cent thought it unfair to use tests for college admission, 53 per cent were against using tests for job selection and 62 per cent against tests for job promotion. Another study on attitudes toward psychological testing was conducted by D. W. Fiske (1965). The 589 subjects interviewed were between 21 and 64 years of age and had completed at least the sixth grade. The sample was quite representative with respect to sex, race, and age as compared with the over-all population of this country. When asked "How good or poor do you think tests would be for finding out what a person's aptitudes, skills and abilities are?," 86 per cent stated "fairly good" to "very good." Asked the same question with respect to interests, 79 per cent responded "fairly good" to "very good." In the area of interpersonal relations and personality measurements the subjects' responses were distributed about equally in all four categories from "very good" to "very bad."

Another question raised concerned the subjects' feelings while taking tests for a job or admission to college or school. A large majority of the subjects expressed general interest in the tests and also expressed the usual anxiety associated with test performance. In no case did a large minority express antagonism toward the test. Thus, it appears the public has mixed opinions with respect to tests and their use.

One may now ask, What is the state of affairs with respect to industrial screening tests or, in industry's terminology, general ability tests? In The Negro Salaried Worker (Gourlay, 1965) the following question was raised: "Should testing procedures used to select white applicants be used to test and evaluate Negroes?" This question had never really been mentioned publicly. Management's increased attention to this problem is unquestionably, in part, a reaction to increased governmental interest as evidenced by the 1964 Civil Rights Act, particularly that section dealing with fair employment. The events in the "Motorola Case" have also created attention.

In the Motorola case a Negro applicant claimed the company was discriminating against him because of his race. The company claimed that the applicant was rejected because he failed their pre-employment test which is designed to predict an applicant's trainability. The applicant stated,

however, that he had passed the test and was thereby rejected because of his race.

In the formal proceedings with an examiner of the Illinois Fair Employment Practices Commission the company stated that its test is "completely race-free" and "administered fairly to all applicants." In the examiner's ruling the test was judged to be unfair to "culturally deprived and disadvantaged groups" and the company was ordered to discontinue the use of the test (Bus. Week, 1964, p. 122).

Later, before the Illinois Fair Employment Practices Commission, the examiner's ruling was modified in that the commission felt the examiner's reasons for discontinuing the test were beyond its jurisdiction. The IFEPC upheld his findings in that Motorola discriminated in the first step of the hiring process and ordered it to stop committing "the unfair employment practice complained of in this complaint" but the commission did not specify what that practice was. The commission went on to say that it did not "foreclose the possibility that tests of this nature are inherently discriminatory against persons alien to the predominant middle-class white culture in this society" (Bus. Week, 1964, p. 130). Thus the commission did not settle the issue but avoided it. French, in his comprehensive review of this case (1965), stated that the commission did find Motorola guilty of discrimination and that the applicant had been denied employment because of his race. He further states

that the commission later made a clarifying statement to the effect that someone at Motorola had marked the applicant as having failed the test. As French pointed out, the applicant did not claim the test discriminated nor did the commission state that the test was discriminatory. Yet, the examiner's ruling implied that the test is biased. The commission's report further emphasized the need to investigate the construction and use of employment screening devices.

Evidence reported by Gourlay suggested that few companies have made adjustments to evaluate the alleged bias of screening tests. This is illustrated by the following statement:

I believe that Negro applicants for positions in engineering, accounting, advertising, etc., should go through the same testing programs as whites. These are generally people with college educations, and therefore the test questions should be readily understandable to them. In the case of Negro applicants for semi-skilled or skilled factory or clerical work, I feel a more intensive interview might be used to supplement test results (p. 67).

It is apparent that this statement assumed educational training to be equal for Negroes and whites at the college level, and at the same time implied that education below the college level is not equal for the two groups. Thus one should supplement test results in the case of applicants for lower skill level jobs.

"We see no reason for adjusting our standards, but it

is possible that we may adjust them in response to pressure. It is our opinion that the bias is in the culture, not in the test" is another common viewpoint cited by Gurlay. One might wonder if it is not possible that test constructors, as products of their culture, could inadvertently build into their tests the biases of their culture. An evaluation of the history of the intelligence testing movement answers this question in the affirmative.

The position taken by industry so far has been defensive. Raymond S. Scruggs, Negro public relations manager for American Telephone and Telegraph Company, has taken a constructive position. He suggested that tests can have positive applications and that industry should investigate the potential of Negro applicants and employees. He went on to say, "I think researchers will find this a fertile field, . . . as one psychologist has said, testing can be used not only to eliminate people but also to identify weaknesses and suggest corrective measures which may assist in increasing skills" (Gurlay, 1965, p. 67).

Ginzberg (1956) provided some suggestions based on a very extensive research study. He found that in the area of educational attainment the Negro was at a considerable disadvantage. In summary, he stated:

The Negro population labors under a double educational handicap. Since the Negro goes to school for fewer years than do whites, he

receives far less preparation for life and work. What schooling he does receive is of inferior quality and therefore has less value than the schooling received by whites. A major weakness of Negro education is the poor preparation of Negro teachers. It has not been possible under segregation to break the cycle of poorly prepared Negro teachers teaching severely handicapped Negro students (p. 60).

This factor was also emphasized by Norgren (1959). He stated, "although strides have been made in raising the educational level of Negroes, there will still be difficulties in placing Negroes in companies which have specified educational requirements established as employment standards for jobs. Thus, a lack of proper education continues as a major and basic limitation in the employment of Negroes and this is so for all levels of educational achievement" (p. 5). He stated further (p. 23), the Negro employment problem is perpetuated by three interrelated factors: low income, negative family influence for advanced education, and low quality schooling.

Gourlay, with more recent data (1965, p. 21), shows that there has been a substantial increase in Negro underutilization, i.e., relatively fewer Negroes are in the labor force than ever before and still more critical is the increase of this non-participation at the younger age levels.

There is much that can and is being done. Society has begun to upgrade Negro education and to remove barriers from previously all white schools. This in the long run will

help. But what of the present? It could be that current pre-employment testing procedures have contributed to keeping Negroes out of the labor market. At the very least the evidence cited here suggests the need for investigations into the area of the importance of cultural differences in employment testing in industry.

Chapter 2

Review of the Literature

With the exception of the "Motorola Case," a review of the literature fails to indicate any specific data with respect to ethnic differences in test performance in the industrial setting. The data available are concerned primarily with ethnic differences in general intelligence test performance.

One of the earlier reviews was that of Klineberg (1944). He concluded that the evidence indicated that there is a difference in intellectual test performance between the Negro and white populations. His explanation for this difference was uncontrolled environmental factors. A review by Shuey (1958) reached the same conclusion, pointing out that the average Negro IQ was 85 whereas Klineberg reported the average Negro IQ to be 86. However, Shuey attributed this difference to hereditary differences in the intellectual capabilities of the two racial groups.

In still another review, Dreger and Miller (1960) take a stand somewhat between the two previous reviewers. They state:

The usefulness of Shuey's otherwise excellent work is limited by what appears to be a polemic attitude. Her book seems to be an attempt to prove a nonequalitarian hypothesis rather than being strictly a review of the literature. In this case, Shuey does the same rationalizing

from a hereditarian standpoint that Klineberg did in his earlier "review" from an environmental standpoint (p. 364).

Tyler (1963) very succinctly summarizes the data in such a fashion that several hypotheses emerge. The major notion generated from the environmentalist camp suggested that intelligence tests are unfair due to the manner in which they are constructed. This viewpoint fostered attempts to develop "culture-free" or "culture-fair" tests (these terms are used interchangeably) of intellectual ability.

A recent study by Kidd (1962) was concerned with the extent to which Cattell's 'Test of g: Culture-Free' is really culture-free. In her review of the literature she reported one study which indicated that Cattell's test was not culture-fair and three other studies which found the test to be culture-fair. Her own analysis supports the fairness of the test. She found 19 items of the matrix format to be free of cultural bias. Sperrazzo and Wilkins (1958) reported evidence on the Raven Matrices Test, a non-verbal intelligence test. Their study included subjects of an all white school, an all Negro school and one integrated school. They reported evidence showing that the Raven correlates with other measures of general ability. The Colored Progressive Matrices Test correlated .23 with the Otis Quick-Scoring Test, .30 with the California Test of

Mental Maturity, and .40 with the Kuhlmann-Anderson. An analysis of variance shows a significant difference with respect to race and socio-economic level and a race by sex by socio-economic level interaction term. The authors offered this explanation:

It is apparent from the significant race by socio-economic status interaction and the third-order interaction that a restriction on the interpretation of the race difference found is necessary. The measured differences in scores between races are related to the age, sex, and socio-economic status of the Ss. The results cannot be interpreted, therefore, as showing differences in intelligence between the races tested here. The difference found seems to depend upon variations of the non-race factors (p. 37).

These studies suggest that tests can be constructed with cultural biases reduced.

Data collected on recruits during the two World Wars suggested the migration hypothesis. In some cases the Northern Negro recruits scored higher as a group than white recruits from some Southern states. Tyler (1963) pointed out two possible explanations for these data: (1) superior Negroes migrate to Northern cities, and (2) educational advantages in the North contribute to better test performance. Klineberg (1935) and more recently Lee (1951) have shown that the assumption of selective migration alone cannot explain the dramatic improvements in test performance.

Still another hypothesis is proposed by Anatasi (1963). She suggested that "Negroes tested suffered from an

unusually barren perceptual experience in early life" (p. 114). Two independent studies--Davidson et al. (1950) and DeStephens (1953)--supported this hypothesis. Both studies used the Wechsler Bellevue scale (Davidson using Form I and DeStephens using Form I and II) in their analysis of ethnic differences. Table 1 indicates that, of the verbal subtests, only the arithmetic scale yielded a significant difference between Negro and white samples. On the performance tests there were three scales in agreement: the Negro groups scored significantly lower on Picture Completion, Block Design, and Digit Symbol. The Performance IQ and the Full Scale IQ were significantly different in favor of the whites for both the Davidson and DeStephens samples.

According to Davidson et al., these performance scales measure various aspects of visual-motor coordination.

"Differing modes of perceptual ability are elicited, but the psychomotor speed with which these modes are expressed weigh heavily in the final quantitative evaluation of subjects."

They went on to suggest that the difference is:

possibly due to the fact that Negroes in our society have no incentive to do things rapidly. It is our hypothesis that this "why hurry" attitude, in the face of a relative lack of attainable socio-economic goals, conditions the responses of Negroes as a group so as to result in lower mean Performance subtest scores on the Wechsler scale (p. 491).

Woods and Toal (1957) furnished similar data using the Revised Beta Test. Using matched samples of Negro and white

Table 1
Comparison of White and Negro Differences on the
Wechsler Bellevue Scale

Subtest	Significant Mean Differences	
	Davidson	DeStephens
Information	.10	.05
Comprehension	.10	.05
Digit Span	.10	.05
Arithmetic	.05	<.05> .01
Similarities	.10	.05
Picture Arrangement	.05	.05
Picture Completion	.01	.01
Block Design	.01	.01
Object Assembly	.10	.01
Digit Symbol	.01	<.05> .01
Verbal IQ	.10	<.05> .01
Performance IQ	.01	.01
Full Scale IQ	.01	.01

IQ scores, they investigated the subtest differences. All subtests except the Maze test were significant at the .05 level. The Digit Symbol and the Identities were in favor of the Negro whereas the Error Recognition, Formboard, and the Picture Completion subtests were in favor of the whites. Woods and Toal suggested that Negroes do better on tests requiring perceptual speed and accuracy, which is contrary to the previous data of Davidson and DeStephens. They went on to suggest that subtests Error Recognition and Picture Completion are culturally loaded and thus favor whites as does the Formboard subtest.

Still another study by Higgins and Sivers (1958) compared Stanford-Binet and Colored Raven Progressive Matrices for children with low socio-economic status. A group of 789 children from schools serving the lowest socio-economic communities were selected. The group was composed of 789 subjects classified according to the following dimensions: 349 Negroes and 440 whites; 389 boys and 400 girls; 271 seven year olds, 273 eight year olds, and 245 nine year olds. Their procedure called for administration of the Stanford-Binet followed by the Colored Raven Progressive Matrices. The results indicated no differences by race on the Stanford-Binet but did show a significant difference by race on the Colored Raven Progressive Matrices. Thus, they accepted the hypothesis that there is no significant difference between

verbal and non-verbal IQ scores for low socio-economic status white children and failed to accept the same hypothesis for Negroes. Higgins and Sivers suggest "that intelligence test heavily loaded with non-verbal items may discriminate against Negro children" (p. 468).

From this review two general statements may be made. First, quality and quantity of education affect test performance; thus, tests which reflect knowledge gained basically in school could result in lower test scores for Negroes. Secondly, tests which require extensive use of non-verbal or perceptual skills may handicap the Negro.

Therefore, two general hypotheses are formulated: (1) there is no difference in ethnic group test performance for the tests in the selection battery, and (2) there is no difference in ethnic group test performance between verbally oriented test and tests heavily loaded with spatial items.

Chapter 3

Procedures and Design

Subjects

The subjects used in this study were new male applicants for petroleum refinery employment. The total applicant population consisted of 608 individuals, 517 whites and 91 Negroes, all having at least a high school education and ranging in age from 20 to 41. The mean age of the new applicants was 26 years.

A sample of white applicants was selected from the total applicant population by matching a white applicant with a Negro applicant on reported educational attainment and age. Matching on education was exactly one to one but in some cases the matched ages varied as much as 3 years. This variation in age is not significant.

Description of the Selection Battery

The battery was composed of six tests. Five of these were used in determining the applicant's advancement to the interviewing step in the selection procedure. The other test was in the battery as an experimental test. All tests in the battery were considered to measure abilities relevant to success on the job. All of the tests were commercially published tests except the background questionnaire which was being developed by the company.

Tests of General Ability

The Test of Learning Ability, Form S, hereafter called Learning Ability, is a 54 item test with a 15 minute time limit. It is spiral-omnibus in design with items arranged in increasing order of difficulty. The three types of items are: block counting, vocabulary, and arithmetic.

The other test in this category has a 10 minute time limit and consists of 24 items. It is entitled Test of Non-Verbal Reasoning, hereafter called Non-Verbal, and is composed of spatial reasoning items utilizing geometric symbols. Both the Learning Ability and the Non-Verbal tests are published by Richardson, Bellows, Henry & Co., Inc.

Mechanical Comprehension

The test used to measure this ability is the Bennett Test of Mechanical Comprehension, Form BB, hereafter called Mechanical Comprehension. This test has no formally established time limit and is published by Psychological Corporation.

Chemical Comprehension

This test consists of 50 items about chemical subjects designed to minimize specific chemistry knowledge related to school training. The Test of Chemical Comprehension, hereafter called Chemical Comprehension, is published by Richardson, Bellows, Henry & Co., Inc. and has a 30 minute time limit.

Mathematical Ability

The test used to measure this ability was the Advanced California Achievement Tests, Form W, mathematics section only and is published by the California Test Bureau. Hereafter this test will be called California Math. It consists of two major subtests: mathematical reasoning and mathematical fundamentals, and has a time limit of 72 minutes.

Background History

The Personnel Questionnaire, Form R-B, is a company developed instrument. It was not used in determining an applicant's eligibility for the interviewing step. This questionnaire is untimed and contained 67 multiple-choice life-history items related to job success.

Administration of the Tests

The test battery, which required approximately three and a half hours to complete, was administered to groups of applicants in a community building located at the refinery site. This testing process extended over several weeks. All applicants were informed at the application stage that they were expected to take the test battery. All standard testing procedures were followed in the administration of the battery.

Study Design

The matched white and Negro samples were divided into three sub-categories according to their educational level. Level A (N = 43) included those who graduated from high school but had no college training, level B (N = 32) was defined as those who had up to two years of college, and level C (N = 16) included those with more than two years of college. The total sample size was 182.

Chapter 4

Results

The results of this study will be presented in the following order: (a) comparison of intercorrelations of the tests for the two ethnic groups, (b) comparison of mean test scores across ethnic groups, and (c) differential effects of the tests in the selection battery across ethnic groups.

A. Comparison of Test Intercorrelations Across Ethnic Groups

The intercorrelations among the tests computed on the total sample ($N = 608$) are presented in Table 2. As can be seen from Table 2, these correlations are in the usual range and fairly uniform across the five tests. All correlations are significant at the .01 level of confidence.

The intercorrelations computed on the all white applicant sample ($N = 517$), presented in Table 3, are very similar to the total sample except that the correlations are slightly but uniformly lower as compared with the total sample. This was apparently due to the smaller standard deviations in the all white sample. Again all correlations are significant at the .01 level.

Tables 4 and 5 present the intercorrelations computed on the Negro sample ($N = 91$) and the white matched sample

Table 2
Intercorrelations of Selection Test Battery for
All Applicants
(N = 608)

	1	2	3	4	5
1. Learning Ability	1.0				
2. Mechanical Comprehension	.46	1.0			
3. Chemical Comprehension	.58	.57	1.0		
4. Non-Verbal	.59	.53	.52	1.0	
5. California Math	.74	.50	.63	.66	1.0
Mean	35.5	26.1	31.8	11.9	90.6
SD	8.3	11.3	6.9	5.3	26.4

Note.--All correlations significant at $p < .01$.

Table 3
 Intercorrelations of Selection Test Battery for
 All White Applicants
 (N = 517)

	1	2	3	4	5
1. Learning Ability	1.0				
2. Mechanical Comprehension	.36	1.0			
3. Chemical Comprehension	.50	.53	1.0		
4. Non-Verbal	.51	.46	.47	1.0	
5. California Math	.69	.42	.56	.62	1.0
Mean	37.2	27.8	32.8	12.8	95.1
SD	7.0	10.7	6.4	5.8	23.7

Note.--All correlations significant at $p \leq .01$.

Table 4
 Intercorrelations of Selection Test Battery for
 Negro Applicants
 (N = 91)

	1	2	3	4	5
1. Learning Ability	1.0				
2. Mechanical Comprehension	.37	1.0			
3. Chemical Comprehension	.53	.39	1.0		
4. Non-Verbal	.48	.35	.34	1.0	
5. California Math	.72	.35	.61	.51	1.0
Mean	26.0	16.2	26.3	7.0	64.8
SD	8.7	9.0	6.8	4.5	26.5

Note.--All correlations significant at $p < .01$.

Table 5
 Intercorrelations of Selection Test Battery for
 Matched Sample of Whites
 (N = 91)

	1	2	3	4	5
1. Learning Ability	1.0				
2. Mechanical Comprehension	.17*	1.0			
3. Chemical Comprehension	.26	.27	1.0		
4. Non-Verbal	.39	.18*	.27	1.0	
5. California Math	.57	.14*	.38	.44	1.0
Mean	36.2	27.0	31.6	11.3	91.5
SD	6.0	8.9	5.6	4.6	20.8

Note.--*denotes r not significant at $p < .01$, all other correlations are significant at $p < .01$.

respectively. It is apparent that the correlations computed on the all Negro sample are very similar to those computed on the all white sample, although the correlations computed on the white matched sample tend to be lower. This result was possibly due to a restriction in range for three of the tests in the matched white sample resulting from matching on education and age. There was no evidence of any consistent differences in selection battery intercorrelations between the all white sample and the Negro sample.

B. Comparison of Mean Test Scores Across Ethnic Groups

As discussed in Chapter 3, 91 white applicants were matched on the basis of educational level and age with 91 Negro applicants. Therefore, a t test appropriate for related samples was computed on the mean differences for each of the five selection tests. These data are presented in Table 6. As can be seen from Table 6, highly significant differences in favor of the whites were obtained on all tests. Therefore, one must conclude that matching on years of education and age does not eliminate the advantage accruing to white applicants on the aptitude tests employed in this study.

Since the matching was accomplished by stratification on three levels of education: (1) high school graduates only, (2) two years of college or less, and (3) more than

Table 6
Means and Standard Deviations of Matched Ethnic Groups
for All Tests of Selection Battery
(N = 91 pairs)

		$\bar{X}$	SD	t_D
Learning Ability	W	36.2	6.0	9.893
	N	26.0	8.7	
Mechanical Comprehension	W	27.0	8.8	8.557
	N	16.2	9.0	
Chemical Comprehension	W	31.6	5.6	6.166
	N	26.3	6.8	
Non-Verbal	W	11.3	4.6	6.053
	N	7.0	4.5	
California Math	W	91.5	20.8	9.827
	N	64.8	26.5	

Note.--All t_D significant at $p < .001$.

two years of college, it was possible to compare Negroes and whites within the education stratifications. These data are presented in Table 7. It was apparent on the Learning Ability test as well as the Mechanical Comprehension test that there was a slight increase in mean score as a function of education for the white sample. However, the Negro sample shows a substantial increase between the "high school" and "less than two years of college" level and basically no difference between "less than two years of college" and "more than two years of college" level. Thus, on these two tests, the largest difference between the Negro and the white samples occurs at the high school level.

A similar trend occurs on the California Math test. The largest difference was at the high school level and there was some convergence at the higher educational levels.

The greatest amount of convergence was found on the Non-Verbal test. However, in this case the mean scores for the white sample actually decreased as a function of education. This is atypical and possibly due to sampling error. Nevertheless, assuming essentially equal scores for the whites the same convergence pattern seems to hold for Non-Verbal Reasoning.

Chemical Comprehension did not show this same pattern. There was no consistent convergence over educational levels.

Table 7
Means and Standard Deviations on All Tests by
Educational Category Across Ethnic Groups
(N = 91 pairs)

		W		N	
		$\bar{X}$	SD	$\bar{X}$	SD
Learning Ability	(N = 43) A	35.3	6.54	23.2	9.30
	(N = 32) B	36.6	4.70	28.6	7.41
	(N = 16) C	37.9	6.22	28.2	7.20
Mechanical Comprehension	A	26.5	8.63	14.3	7.88
	B	27.5	8.21	18.6	9.58
	C	27.2	10.47	17.1	10.38
Chemical Comprehension	A	30.6	4.76	24.5	6.04
	B	30.9	4.52	27.2	6.97
	C	35.8	7.57	29.6	7.04
Non-Verbal	A	12.2	4.70	5.8	4.03
	B	10.7	4.26	7.9	5.17
	C	9.9	4.35	8.1	3.39
California Math	A	86.2	19.34	55.8	23.84
	B	95.3	17.81	70.1	22.82
	C	98.1	26.08	77.7	31.03

C. Differential Effects of Selection Tests
Across Ethnic Groups

A measure of test difficulty is presented in Table 9. These measures were derived empirically by counting the number of Negro and white applicants who failed to meet the cut-off standard of each test in the battery. The cut-off scores were set at approximately the 70th percentile based on existing company norm tables for each test used in the selection battery. Assuming no differences between ethnic groups the expectancy of failures should be in the neighborhood of 70 per cent for both Negroes and whites. As can be seen from Table 8 the white sample failure percentage was considerably below the expected value whereas the Negro sample was considerably higher.

Further differences occur when individual tests are compared across ethnic groups. Whereas the Learning Ability test was the most frequently failed test for the Negroes, it was in fourth position of difficulty for the whites. Whereas the Chemical Comprehension test was the least frequently failed test for the Negro sample, it was the second most frequently failed test for the white sample. These differences, however, are not statistically significant.

A difference score (D) was computed for each test indicating the extent to which Negroes failed more frequently

Table 8
Comparison of Empirically Derived Order of Difficulty
for Each Test Across Ethnic Groups

	W (N = 91)		N		D%
	F _F	% _F	F _F	% _F	
Learning Ability	49	53.85	83	91.21	37.36
Mechanical Comprehension	45	49.45	78	85.71	36.26
Non-Verbal	53	58.24	81	89.01	30.77
California Math	65	71.43	81	89.01	17.58
Chemical Comprehension	56	61.54	72	79.12	17.58

Note.--D = %_F Negro - %_F white.

than did whites. As can be seen from Table 8 the D's indicate a differential failing effect among the selection battery tests for the Negro sample. The D's form two categories. The California Math and the Chemical Comprehension test show the least percentage difference for the white sample with D's less than 20 per cent. The Learning Ability, Mechanical Comprehension, and the Non-Verbal tests form another category all having D's greater than 30 per cent.

The major distinction between the two categories is the spatial factor. In the greater than 30 per cent D category two of the three tests are purely non-verbal measures requiring spatial reasoning while the third test (Learning Ability) is partially composed of spatial items. The less than 20 per cent D category is formed by the California Math which is simply arithmetic and the Chemical Comprehension test which measures general chemistry knowledge.

Table 9 presents another measure of test difficulty. The percentage increase (PI) was computed by taking the difference between the mean scores of the white matched sample and the Negro sample divided by the mean score of the Negro sample. The two college categories were combined to make the N similar to that of the high school category. As is apparent from Table 9, the spatial reasoning tests or non-verbal measures show the greatest difference between white and Negro samples. The California Math and the

Table 9
 Comparison of White Increase of Mean Score Over Mean
 Negro Test Score Across Educational Categories
 (Percentage Increase*)

	High School (Only) (N = 43)	High School (More Than) (N = 48)	Total (N = 91)
Learning Ability	52.2%	29.8%	39.2%
Mechanical Comprehension	85.1	52.5	62.9
Non-Verbal	110.0	31.6	61.2
California Math	54.5	32.5	41.2
Chemical Comprehension	24.8	16.4	20.2

Note.--*PI = $\frac{(XW - XN)}{XN} \times 100$

Learning Ability test are intermediate with the Chemical Comprehension test having the smallest difference between Negro and white samples.

Chapter 5

Summary and Conclusions

The purpose of this study was to investigate possible ethnic differences in an industrial selection test battery. The major hypothesis that there would be no difference in test performance between ethnic groups was rejected. All racial comparisons were significant at the .001 level of confidence. A comparison of ethnic groups at various educational levels indicated that the mean score of Negroes increased on all tests in the battery as a function of additional education whereas there was little or no gain for the white matched sample. The greatest increase in mean score of the Negro sample occurred between the "high school" and "less than two years of college" levels.

For three of the tests--Learning Ability, Mechanical Comprehension, and Chemical Comprehension--the matched white sample did not show a substantial improvement in mean score as a function of education. On one test, Non-Verbal, the mean score actually decreased, but this result was possibly due to sampling error as explained in Chapter 4. Only on the California Math did the matched white sample show definite increases in mean score as a function of increased educational attainment.

Since this study was an empirical research effort originating after the data had been collected, appropriate

statistical and design measures were lacking for testing the second hypothesis. This second hypothesis asked, Is there a difference in ethnic group test performance between tests predominantly loaded with a spatial reasoning factor and tests oriented with a high verbal factor?

The data indicated that the most difficult test for whites was not the most difficult for Negroes and the least difficult test for Negroes was not the least difficult for whites. In other words, there was a differential effect in the tests themselves, as brought out by the ethnic classification.

The major finding was the two distinguishable categories formed by the D's (the difference between the percentage of Negroes and percentage of whites who failed to meet the minimum score), as presented in Table 8. The category representing the largest D's was described as one which measures a spatial reasoning factor. The other D category was described as one measuring a verbal factor.

The literature review suggested two possible explanations for this racial difference in test performance. One was inferior educational training and the other was possible inferior perceptual development. Considering the type of sub-scales that make up the Learning Ability test, it may be noted that the block counting scale would offer a measure of spatial reasoning while the vocabulary and arithmetic scales

are of the verbal nature. Thus, the Learning Ability test has both biases incorporated into the total score since vocabulary and arithmetic are highly dependent on education. On the other hand, the Chemical Comprehension test has multiple-choice items relating to chemical concepts which were specifically designed to minimize any educational bias. Thus, the Chemical Comprehension test has no spatial reasoning bias and little if any educational bias.

Further indication of a differential effect was noted in Table 9. It is apparent from this table that the PI score decreases as a function of more education for all tests in the battery. The differential pattern is readily noticeable from the "total" column in Table 9. The Mechanical Comprehension and the Non-Verbal tests showed the greatest percentage increase in favor of whites. The Learning Ability and the California Math tests are intermediate with percentage increases in the area of 40 per cent while the Chemical Comprehension test showed the least percentage increase (20 per cent). In addition, data from Table 8 indicated that the Learning Ability test was in the largest D category (30 per cent), but in Table 9 this test was in the intermediate PI category. This fact suggested that the spatial factor is more important in contributing to ethnic differences than the educational bias. Thus, it is suggested that spatial reasoning tests (spatial factor) produce

the greatest difference between Negro and white samples as employed in this study.

In summary, this study indicated that ethnic differences occur in test performance and that it is on spatial reasoning kinds of tests that we observe the greatest differences between ethnic groups. This finding suggested that non-verbal measures may be more culturally biased than traditional verbal tests. This contradicts the underlying hypothesis of the "culture-fair" testing movement--that non-verbal measures do not discriminate across socio-economic levels or ethnic groups. If Anastasi's suggestion that barren cultural backgrounds retard the development of perceptual skills is correct, then one could predict a difference in spatial reasoning test performance between ethnic groups.

On the other hand, a recent study by Kidd (1965) suggests that certain characteristics of spatial items do not discriminate between socio-economic levels for Anglo-Americans and Mexican-Americans. Her study utilized the 'Test of g: Culture-Free' by Cattell. The possibility exist that certain types of spatial items may prove to be culture-fair, i.e., common to all cultures.

The data in this study, being suggestive rather than conclusive, indicate the need to intensify our efforts in exploring the perceptual bias hypothesis.

References

- Anastasi, Anne. Differential psychology. (3rd ed.)
New York: Macmillan, 1958.
- Carter, L. F. Psychological testing and public responsibility. Science, December, 1964, 146, 1695-1697.
- Davidson, K. S., Gibby, R. G., McNeil, E. B., Segal, S. J., & Silverman, H. A preliminary study of Negro and white differences in Form I of the Wechsler-Bellevue Scale. J. consult. Psychol., 1950, 14, 489-492.
- DeStephens, W. D. Are criminals morons? J. soc. Psychol., 1953, 38, 187-199.
- Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and whites in the United States. Psychol. Bull., 1960, 57, 361-402.
- Fiske, D. W. The subject looks at psychological tests. Proceedings of the 1964 Invitational Conference on Testing Problems. New Jersey: Educational Testing Service, 1965.
- French, R. L. The Motorola Case. Industr. Psychologist, 1965, 2, No. 3, 29-50.
- Ginzberg, E. The Negro potential. New York: Columbia University Press, 1964.
- Gourlay, J. G. The Negro salaried worker. New York: American Management Association, 1965.

- Higgins, C., & Sivers, Cathryne. A comparison of Stanford-Binet and Colored Raven Progressive Matrices I.Q.'s for children with low socio-economic status. J. consult. Psychol., 1958, 22, 465-458.
- Ingle, D. J. Racial differences and the future. Science, October, 1964, 146, 375-379.
- Ingle, D. J. Ingle replies. Science, December, 1964, 146, 1528-1530.
- The Issue of Race and Tests. Business Week, March 28, 1964, 122.
- Kidd, A. H. The culture-fair aspects of Cattell's Test of g: Culture-Free. J. genet. Psychol., 1962, 101, 343-362.
- Kidd, A. H., & Rivoire, J. L. The culture-fair aspects of the development of spatial perception. J. genet. Psychol., 1965, 106, 101-111.
- Klineberg, O. Negro intelligence and selective migration. New York: Columbia University Press, 1935.
- Klineberg, O. (Ed.) Characteristics of the American Negro. New York: Harper, 1944.
- Labor Briefs. Business Week, November 28, 1964, 130.
- Lee, E. S. Negro intelligence and selective migration: a Philadelphia test of Klineberg's hypothesis. Amer. sociol. Rev., 1951, 61, 227-233.
- Norgren, P. H. Employing the Negro in American industry. New York: Industrial Relations Counselor, Inc., 1959.

Shuey, A. M. The testing of Negro intelligence. Lynchburg, Va.: J. P. Bell, 1958.

Sperrazzo, G., & Wilkins, W. L. Further normative data on the Progressive Matrices. J. consult. Psychol., 1958, 22, 35-37.

Tyler, L. E. Test and measurements. Foundations Modern Psychology Series. New Jersey: Prentice-Hall Inc., 1963.

Woods, W. A., & Toal, R. Subtest disparity of Negro and white groups matched for I.Q.'s on the Revised Beta Test. J. consult. Psychol., 1957, 21, 136-138.