AIR-753-10/71-TR-2
AD 737698

ETHNIC GROUP MEMBERSHIP AS A MODERATOR IN THE PREDICTION OF JOB PERFORMANCE: An Examination of Some Less Traditional Predictors

James L. Farr
Brian S. O'Leary
C. Michael Pfeiffer
Irwin L. Goldstein
C. J. Bartlett

TECHNICAL REPORT NUMBER II

Prepared under Contract to
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research
Department of the Navy

Principal Investigator: Albert S. Glickman

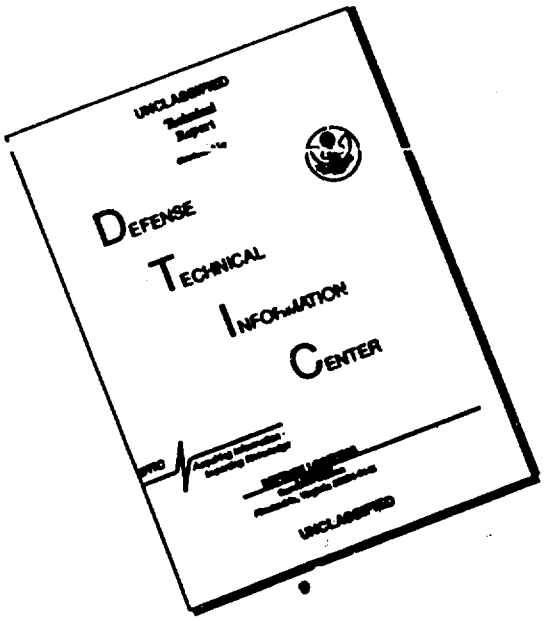
Contract No. N00014-68-C-0341

NR No. 151-277

American Institutes for Research
Washington Office
Institute for Research on Organizational Behavior

October 1971

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

TO: DISTRIBUTION LIST

SUBJECT: O'Leary, Brian S., Farr, James L., & Bartlett, C. J. Ethnic Group Membership as a Moderator of Job Performance. Technical Report Number 1, April, 1970. Washington: American Institutes for Research

ERRATUM

Please note and correct erratum on page 132, paragraph 2, of Technical Report Number 1, the first of the two reports on this research project.

The first two sentences should read:

> Perhaps the most important finding of this phase of the research project is the fact that what might be called Model V (as in "valid" -- no difference on prediction or criterion, equal subgroup validities) occurred so infrequently. Traditional personnel selections assume that Model V is operative (i.e., a single regression line can be used for all subgroups in a population).

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
American Institutes for Research
710 Chatham Center Office Building
Pittsburgh, Pennsylvania 15219

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
ETHNIC GROUP MEMBERSHIP AS A MODERATOR IN THE PREDICTION OF JOB PERFORMANCE: AN EXAMINATION OF SOME LESS TRADITIONAL PREDICTORS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report Number II

5. AUTHOR(S) (Last name, first name, initial)
Farr, James L.; O'Leary, Brian S.; Pfeiffer, C. Michael; Goldstein, Irwin L.; & Bartlett, C. J.
Principal Investigator: Albert S. Glickman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1971		

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-68-C-0341	AIR-753-9/71-TR-2
b. PROJECT NO. NR 151-277	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES
This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research Department of the Navy Washington, D. C.

13. ABSTRACT

Phase I of this project represented an attempt to obtain a picture of the effects of cultural factors on existing selection procedures using psychological tests. In examining 765 predictor-criterion relationships 219 cases were found of unfairness to a racial subgroup had these tests been used for selection.

Phase II represented an attempt to obtain a picture of the effects of cultural factors on less traditional measures. Eight studies were carried out including as predictors work sample tests, simulators, motor skill tests, culture-free tests, biographical information, and learning measures. Altogether, 400 additional predictor-criterion relationships were examined including 272 with these less traditional predictors.

The use of less traditional predictors did not result in an improvement in validity or in a reduction of unfairness. The most frequent occurrence was a lack of validity for either ethnic group, with the finding of validity combined with no unfairness being exceedingly rare. Although some of the less traditional showed promise it was concluded that situation-specific validation with all predictors appears essential in order to avoid possible inadvertent bias to minority groups through unfair selection measures.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Academic Performace Criteria						
Culture Free Tests						
Cultural Factors						
Discrimination						
Ethnic Group						
Job Performance Criteria						
Minority Groups						
Moderator Variables						
Selection Tests						
Test Bias						
Test Fairness						
Test Validity						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY**: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. **REPORT SECURITY CLASSIFICATION**: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. **GROUP**: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE**: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES**: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S)**: Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE**: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. **TOTAL NUMBER OF PAGES**: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. **NUMBER OF REFERENCES**: Enter the total number of references cited in the report.

8*a*. **CONTRACT OR GRANT NUMBER**: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. **PROJECT NUMBER**: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. **ORIGINATOR'S REPORT NUMBER(S)**: Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. **OTHER REPORT NUMBER(S)**: If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES**: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES**: Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY**: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. **ABSTRACT**: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS**: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Phase I of this project (O'Leary, Farr, and Bartlett, 1970) represented an attempt to obtain a picture of the effects of cultural factors on existing selection procedures using psychological tests. Seven independent studies were conducted in which the validity of commercial and industrially developed selection tests examined separately for white and Negro subgroups. Examining mean differences in predictor and criterion measures along with the separate validity coefficients 765 predictor-criterion relationships were then classified according to the differential prediction models discussed by Bartlett and O'Leary (1969). Of the 765 relationships examined 357 Bartlett-O'Leary models were described with 219 cases of unfairness to a racial subgroup had these tests been used for selection.

Phase II of this project represents an attempt to obtain a picture of the effects of cultural factors on less traditional measures which might be utilized in selection. Due to the extent to which traditional psychological tests might lead to inadvertent racial discrimination it is hoped that less traditional measures may show fewer instances of unfairness when used for personnel selection. The purpose of the second phase of this project is to examine a number of these less traditional measures and evaluate them with regard to their fairness in selection from racially mixed groups.

Study #1 was carried out on a racially mixed group of taxicab drivers. Driver training simulators and a variety of psychomotor tests were used as predictors. Criteria included observation of street errors, accident record, and traffic violation record.

Study #2 was carried out on a group of high school students with several biographical inventory keys utilized as predictors along with a traditional ability test. Grades in school and teacher ratings were used as criteria.

Study #3 examined the prediction of college grades using an opinion questionnaire (University Student Census) as a predictor as well as personality and interest inventory measures in order to determine if these measures could improve the standard academic predictions.

Studies #4 and #5 examined workers in a health insurance company. The SRA Pictorial Reasoning Test which has been alleged to be culture-free was compared to more traditional ability measures in the prediction of supervisor ratings. In these studies a measure of cultural deprivation suggested by Guion (1966) was employed. This measure utilizes the discrepancy between the scores on the culture-bound test and a culture-free test as either a moderator or a predictor.

Study #6 compared Fundamental Achievement Series, another measure regarded as less culturally biased, with traditional ability measures in the prediction of success in Marine Corps training. As in the two previous studies the cultural deprivation index suggested by Guion was employed.

Study #7 examined a work sample testing procedure along with psychomotor tests in the prediction of turnover and training criteria for sewing machine workers. The work sample test provided a career preview of the job, serving as a job expectancy training program as well as a selection instrument. Thus self-selection as well as administrative selection was investigated.

Study #8 examined experimental learning tasks as predictors of programmed learning performance using a pre-post design. Three different learning tasks were utilized: a paired associate task; a concept learning task; and a principle learning task. The criterion consisted of learning performance on a programmed instructional unit as measured by gain in proficiency on a multiple choice test.

All of these less traditional predictions were then examined in terms of the differential prediction models described by Bartlett and O'Leary, and frequency of selection bias was compared with more traditional predictors typically used in personnel selection.

The Bartlett-O'Leary Models

All of the relationships are classified according to the differential prediction models which have been described in detail by Bartlett and O'Leary (1969). Scatter diagram representations of these eleven models are presented in Figures 1 through 11. The models are summarized briefly as follows:

Model 1 - The predictor is valid for both subgroups, and significant mean differences occur on both the predictor and the criterion.

Model 2 - The predictor is valid for both subgroups, and significant mean differences occur on the predictor but not the criterion.

Model 3 - The predictor is valid for both subgroups, and significant mean differences occur on the criterion but not the predictor.

Model 4 - The predictor is valid for both subgroups, and significant mean differences occur on the predictor and the criterion, but in opposite directions. This model differs from the first one in that the intercepts of the two regression lines are different for this one.

Model 5 - The predictor is valid for one subgroup but not the other, and there are no significant mean differences on either the predictor or the criterion.

Model 6 - The predictor is valid for one subgroup but not the other, and there are significant mean differences on the criterion but not the predictor.

Model 7 - The predictor is valid for one subgroup but not the other, and there are significant mean differences on the predictor but not the criterion.

Model 8 - The predictor is valid for one subgroup but not the other, and there are significant mean differences on both the predictor and the criterion.

Model 9 - The predictor is valid for both subgroups, but in opposite directions. There are no mean differences on either the predictor or the criterion.

Model 10 - The predictor is valid for both subgroups, but in opposite directions. There are significant differences on the predictor but not the criterion.

Model 11 - There is no validity in either subgroup, but there are significant differences on both the predictor and the criterion.

SCATTER DIAGRAM REPRESENTATIONS

OF THE BARTLETT - O'LEARY MODELS

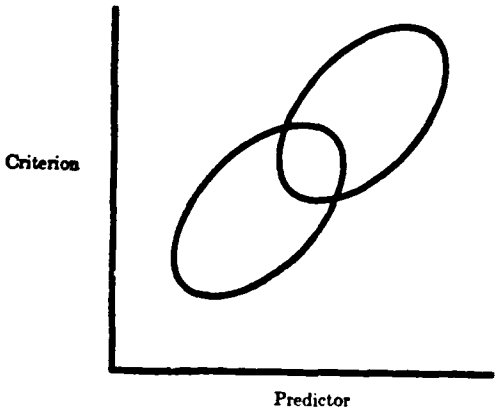


Figure 1

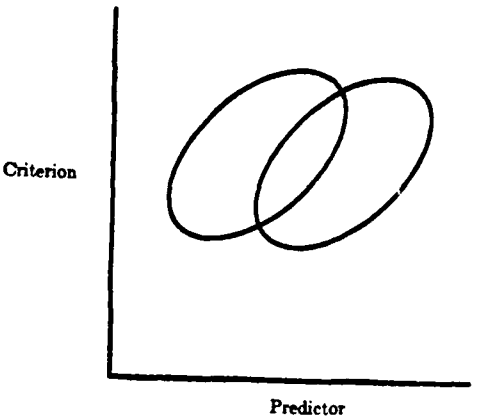


Figure 2

Figure 3

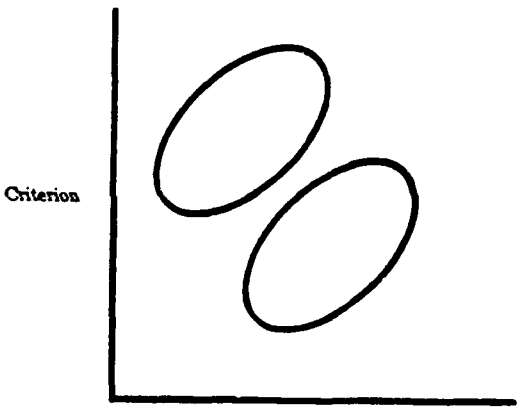


Figure 4

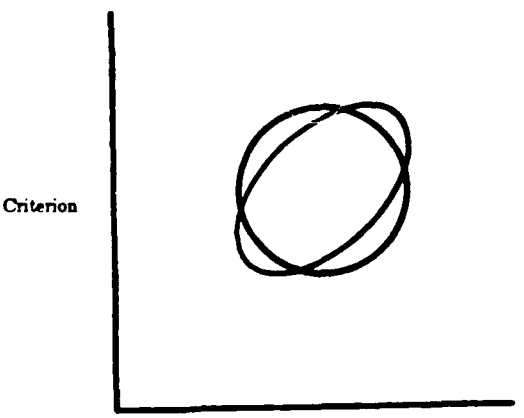


Figure 5

Figure 6

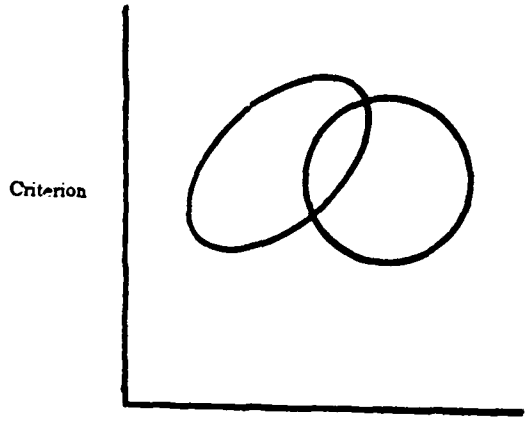


Figure 7

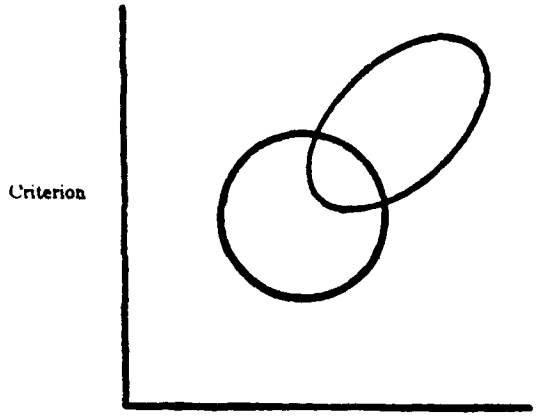


Figure 8

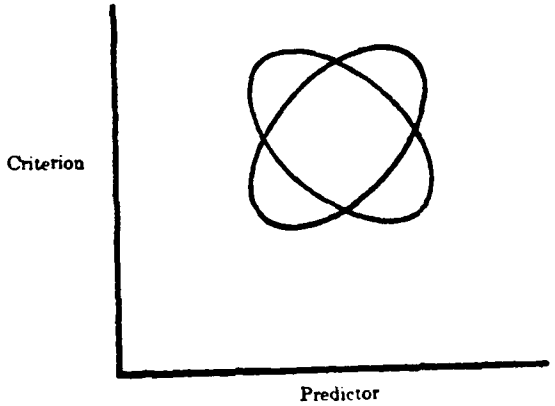


Figure 9

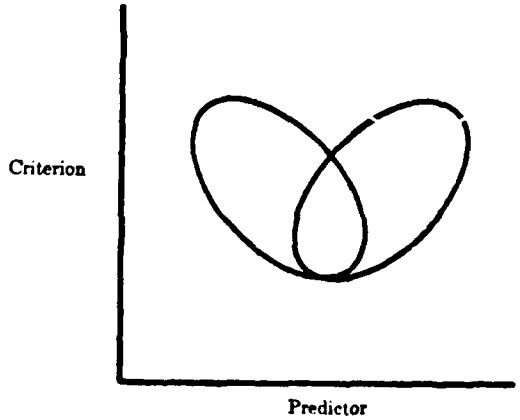


Figure 10

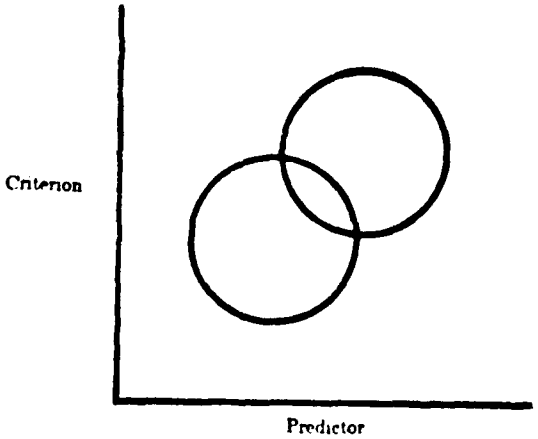


Figure 11

Study #1: Taxicab Drivers

The Relationship between Street Driving and Simulator and Perceptual-Motor Ability Test Performance

Sample

This investigation represents a reanalysis of data reported by Edwards, Hahn and Fleishman (1969) of the American Institutes for Research. The present study reports data analyses of the total group and the two racial subgroups. Complete details of the original study are found in Edwards, et al. The taxicab driver sample consisted of 303 workers of whom 152 were white and 151 were Negro. All of the drivers were employed in Washington, D.C. and generally drove in city traffic. As can be seen in Table 1, the white subgroup was older than the Negro subgroup and, consequently, also had more years of driving experience than the Negro subgroup. The Negro taxicab drivers had attained a significantly higher level of education than the white subgroup. The two subgroups had approximately equal annual driving mileage. All subjects were paid for their participation in the study.

Predictors

For the purposes of the present data analyses, a number of variables were investigated as possible predictors of driving behavior. Among these were two laboratory driving simulators and several psychomotor abilities measures including simple reaction time, jump reaction time, motor judgment and complex coordination test.

The two driving simulators employed were the Allstate Good Driver Trainer (AGDT) and the Aetna Drivotron. Both simulators consisted of a cab in which the driver sits, a motion picture screen and projector, and scoring consoles.

Table 1: Biographical Data - Taxicab Drivers

Variable	Group	$\bar{X}$	s	N	t
Age	Total	44.61	12.55	300	
	White	50.07	11.86	150	
	Negro	39.15	10.73	150	8.33**
Education (years)	Total	11.14	2.95	299	
	White	10.70	2.81	149	
	Negro	11.57	3.03	150	2.57*
Driving Experience (years)	Total	26.43	12.60	303	
	White	31.41	12.46	152	
	Negro	21.42	10.62	151	7.48**
Annual Driving Mileage (in 000's)	Total	38.64	15.32	272	
	White	37.34	13.23	145	
	Negro	40.13	17.33	127	1.50

* $p < .05$

** $p < .01$

The motion picture presented various driving situations and the driver reacted by steering, braking, signaling, etc., in response to the situations presented. The total number of incorrect responses was used as the only predictor measure with the AGDT. When the Drivotron was used, several measures including, steering errors, speed errors, brake errors, signal errors, and number of starbursts (number of accidents which would have resulted from errors committed) were combined to provide total error scores. Each simulator drive ranged from 20-25 minutes. Half of the sample began with the Drivotron, followed by the AGDT; the other half was given the AGDT followed by the Drivotron.

In addition to the driving simulator, the following measures were investigated as potential predictors:

a) Simple reaction time was measured by the time required for the S to depress a response button after the onset of a stimulus light. The S's hand rested on the response button at the beginning of each trial. A total score comprised of the sum of reaction times over 20 trials was used.

b) Jump reaction time was measured similarly with the additional requirement that the S place his hand 12 inches away from the response button until the stimulus light was activated.

c) The Motor Judgment Test measured the ability to make continuous anticipatory motor adjustments relative to changes in speed and direction of a continuously moving target. In the Motor Judgment Test, the S was confronted by two adjacent disks rotating at a constant speed. Each disk had black and white sections on its perimeter. Between these disks was a pointer whose speed of rotation the S could control. The S was not able to stop the rotation of this pointer completely nor could he exert control over the two rotating disks. The S was required to make as many revolutions of the pointer as possible without crossing the black areas on the rotating disks. To do this properly S had to

integrate his estimates of the speed of each disk, the pointer and his own control movements. Scores recorded were number of pointer revolutions and number of errors (crossing of black areas) during four 1-minute trials separated by 15-second rests.

d) The Complex Coordination Test is a measure of multilimb coordination which represented the ability of a S to make coordinated, simultaneous responses of several limbs in operating controls. Patterns of lights were presented whose positions were to be matched by appropriate adjustments of stick and rudder controls. A correct response was accomplished only when both the hands and feet had completed and maintained the appropriate adjustments, at which point a new pattern of lights to be matched were presented. The scores are the number of completed matchings during a four and a ten minute period.

Table 2 presents the predictor means, standard deviations and tests of significance of mean differences for the taxicab driver sample. The white subgroup committed significantly more errors on the AGDT than the Negro subgroup. On the Motor Judgment Test the white drivers had more errors but completed more total pointer revolutions than the Negro sample. There were no other significant predictor mean differences between the white and Negro subgroups.

Criteria

Three measures of driving performance were used as criterion measures. Street observations of actual driving behavior were made by two independent observers. The drivers were not informed that they were being observed. This was accomplished by having the observers ride in each cab from downtown Washington to Silver Spring, Maryland as if they were normal passengers. When the vehicle arrived at the AIR office in Silver Spring, the drivers were informed of the research project. However, they were not told that their driving behavior had been systematically observed. During the trip, each observer independently

Table 2: Predictor Means, Standard Deviations, N's and Tests of Significance of Mean Differences - Taxicab Drivers

Predictor	Group	$\bar{X}$	s	N	t
AGDT - Total Incorrect	Total	45.44	10.65	300	
	White	47.07	9.75	150	
	Negro	43.81	11.27	150	2.67**
Drivotron - Total	Total	42.95	5.80	303	
	White	42.78	5.87	152	
	Negro	43.11	5.74	151	0.49
Drivotron - Starbursts	Total	4.93	1.60	303	
	White	4.90	1.59	152	
	Negro	4.95	1.61	151	0.27
Drivotron - Steering	Total	9.93	1.96	303	
	White	9.93	1.89	152	
	Negro	9.94	2.04	151	0.04
Drivotron - Speed	Total	13.34	3.26	303	
	White	13.61	3.19	152	
	Negro	13.08	3.33	151	1.41
Drivotron - Brake	Total	10.41	2.77	303	
	White	10.36	2.63	152	
	Negro	10.46	2.92	151	0.31
Drivotron - Signal	Total	1.24	0.95	303	
	White	1.24	0.95	152	
	Negro	1.24	0.96	151	0.00

Table 2 (Contd)

Predictor	Group	$\bar{X}$	s	N	t
Simple Reaction Time	Total	232.75	53.99	299	
	White	236.02	42.56	151	
	Negro	229.41	63.55	148	1.06
Jump Reaction Time	Total	540.48	95.06	300	
	White	541.29	91.81	152	
	Negro	539.48	98.60	148	1.60
Motor Judgment Total Errors	Total	43.83	27.81	301	
	White	47.03	33.46	150	
	Negro	40.66	20.34	151	1.99*
Motor Judgment Total Revolutions	Total	53.05	19.49	301	
	White	56.98	23.14	150	
	Negro	49.14	14.03	151	3.55**
Complex Coord. 4 min.	Total	8.96	5.62	299	
	White	9.40	6.14	151	
	Negro	8.52	5.01	148	1.35
Complex Coord. 10 min.	Total	25.71	14.92	299	
	White	26.74	16.14	151	
	Negro	24.66	13.55	148	1.20

* $p < .05$

** $p < .01$

recorded errors made by the driver, using a partially pre-coded check list which had been developed on the basis of previous research (Edwards and Hahn, 1964). The form yielded a total error score which was used in the present analysis.

The total number of moving violations with which each driver had been charged and the total number of accidents in which each driver had been involved were also used as criterion measures. These data were obtained from the official records of the Department of Motor Vehicles.

As indicated in Table 3, the Negro drivers had been involved in more accidents and had been charged with more moving violations than the white subgroup. There was no significant difference between the two subgroups on the number of street errors recorded by the two observers.

Validity

Predictor-criterion correlations are presented in Table 4. Generally, low validity was found between the various predictors and criteria. The validity coefficients in Table 4 were corrected for age by partial correlation techniques. No predictor was valid for more than a single criterion. The observed street error criterion was slightly more predictable than recorded accidents on moving violations.

Models Illustrated

An examination of these data led to several cases which illustrated the Bartlett and O'Leary models (1969). They are:

a) Two examples of Model 5 which represents the situation in which there are no mean differences on either predictor or criterion measures and significant validity for only the white subgroup.

Table 3: Criterion Means, Standard Deviations, N's, and Tests of Significance of Mean Differences - Taxicab Drivers

Criterion	Group	$\bar{X}$	s	N	t
Total Street Errors	Total	22.57	8.36	303	
	White	23.34	9.09	152	
	Negro	21.81	7.50	151	1.59
Accidents	Total	1.42	1.14	303	
	White	1.23	0.95	152	
	Negro	1.60	1.28	151	2.85**
Violations	Total	3.11	2.76	303	
	White	2.48	2.62	152	
	Negro	3.74	2.76	151	4.06**

** $p < .01$

Table 4: Predictor - Criterion Correlations - Total Group, White and Negro Subgroups

Taxicab Drivers(1)

Predictor	Group	Criterion: Street Errors r	Street Errors N	Accidents r	Accidents N	Violations r	Violations N
AGDT - Total Incorrect	Total	-04	300	-17**	300	00	300
	White	-07	150	-21**	150	01	150
	Negro	-04	150	-12	150	07	150
				(8)			
Drivotron - Total	Total	00	303	06	303	-03	303
	White	03	152	01	152	-01	152
	Negro	-03	151	09	151	-06	151
Drivotron - Starbursts	Total	-01	303	-08	303	04	303
	White	-07	152	-09	152	08	152
	Negro	08	151	-09	151	00	151
Drivotron - Steering	Total	01	303	-05	303	-02	303
	White	-04	152	07	152	00	152
	Negro	06	151	-16*	151	-04	151
				(7)			
Drivotron - Speed	Total	04	303	-03	303	-03	303
	White	-01	152	-04	152	02	152
	Negro	08	151	00	151	-03	151
Drivotron - Brake	Total	-04	303	-08	303	02	303
	White	-01	152	-05	152	03	152
	Negro	-06	151	-11	151	00	151

Table 4 (Contd)

Predictor	Group	Street Errors		Accidents		Violations	
		r	N	r	N	r	N
Drivotron - Signal	Total	-10	303	09	303	-02	303
	White	-13	152	09	152	-04	152
	Negro	-08	151	09	151	00	151
Simple Reaction Time	Total	04	299	00	299	06	299
	White	-07	151	06	151	11	151
	Negro	12	148	-02	148	06	148
Jump Reaction Time	Total	-07	300	-01	300	06	300
	White	-16*	152	02	152	01	152
	Negro	02	148	-03	148	11	148
		(5)					
Motor Judgment - Total Errors	Total	-02	301	-10	301	07	301
	White	-06	150	-08	150	16*	150
	Negro	02	151	-10	151	01	151
						(8)	
Motor Judgment - Total Revolutions	Total	03	301	-10	301	05	301
	White	01	150	-07	150	12	150
	Negro	02	151	-08	151	07	151
Complex Coord. - 4 min.	Total	16**	299	-12	299	-07	299
	White	15	151	-05	151	-02	151
	Negro	15	148	-13	148	-10	148
Complex Coord. - 10 min.	Total	17**	299	-13	299	-06	299
	White	18*	151	-08	151	-01	151
	Negro	13	148	-12	148	-08	148
		(5)					

* $p < .05$; ** $p < .01$

(1) The number in parentheses beneath the Negro group correlation indicates the model illustrated (after Bartlett and O'Leary, 1969).

b) A single example of Model 7 was observed in the taxicab driver data. Model 7 illustrates the situation in which there is a mean criterion difference and validity for only the Negro subgroup.

c) Two examples of Model 8 were illustrated by the predictor-criterion relationships. This model is representative of the situation in which there are significant subgroup mean differences in both predictor and criterion performance but significant validity for only one subgroup. It is interesting to note that only a single predictor variable exhibited significant validity with the Negro subgroup despite the fact that the two subgroups were of approximately equal size.

Multiple Regression Analyses

A stepwise regression procedure was used to select predictors from the total 13 available for the multiple regression analyses using the criterion developed by Lord (1950). The results of these analyses for the total group and the two racial subgroups are shown in Table 5. For the total group the multiple regression equation was developed on a random back sample of 200 cases and cross validated on a sample of 94 cases. For the racial subgroups the multiple regression equations were developed on back samples of 100 random cases each and cross validated on a sample of 47 cases each. None of the relationships found in the back samples held up in the cross validation groups. In general, different predictors tended to be chosen for the multiple regression equations in the two racial subgroups.

Table 6 presents the results of multiple regression analyses conducted using the total sample (N = 294) as the back sample on which the equations were developed and the racial subgroups as cross validation samples (N = 147 for both subgroups). As can be seen in Table 6, the multiple correlations held up in three of the six instances. The success of these regression equations

Table 5: Multiple Regression Analyses - Equations,

Back Sample R's and Cross-Validated R's

Group	Criterion	Equation
Total	Street Errors	$Y = 24.55 - .166\ X_2 + .216\ X_5 + .076\ X_{13}$ $R_{BACK} = .21$ (N=200); $R_{c-v} = .06$ (N=94)
	Accidents	$Y = 2.676 - .016\ X_1 - .072\ X_3 - .009\ X_{13}$ $R_{BACK} = .23$ (N=200); $R_{c-v} = .19$ (N=94)
	Violations	$Y = 1.894 + .006\ X_8 - .215\ X_{12} + .070\ X_{13}$ $R_{BACK} = .18$ (N=200); $R_{c-v} = .02$ (N=94)
White	Street Errors	$Y = 29.806 - .227\ X_2 - 1.235\ X_{12} + .535\ X_{13}$ $R_{BACK} = .28$ (N=100); $R_{c-v} = .05$ (N=47)
	Accidents	$Y = 1.723 - .024\ X_1 + .060\ X_4$ $R_{BACK} = .27$ (N=100; $R_{c-v} = .11$ (N=47)
	Violations	$Y = -.004 - .206\ X_4 + .127\ X_5 + .008\ X_8 + .017\ X_{10}$ $R_{BACK} = .33$ (N=100); $R_{c-v} = .16$ (N=47)
Negro	Street Errors	$Y = 11.438 + .340\ X_5 - .338\ X_6 + .028\ X_7 + .346\ X_{12}$ $R_{BACK} = .38$ (N=100); $R_{c-v} = .02$ (N=47)
	Accidents	$Y = 2.892 - .094\ X_4 - .046\ X_{12}$ $R_{BACK} = .26$ (N=100); $R_{c-v} = .01$ (N=47)
	Violations	$Y = 0.476 - .299\ X_7 + .004\ X_9 - .066\ X_{10} + .086\ X_{11}$ $R_{BACK} = .30$ (N=100); $R_{c-v} = .05$ (N=47)

Note: X_1 = AGDT, Total Incorrect; X_2 = Drivotron, Total;

X_3 = Drivotron, Starbursts; X_4 = Drivotron, Steering;

X_5 = Drivotron, Speed; X_6 = Drivotron. Brake;

X_7 = Drivotron, Signal; X_8 = Simple Reaction Time;

X_9 = Jump Reaction Time; X_{10} = Motor Judgment, Total Errors;

X_{11} = Motor Judgment, Total Revolutions; X_{12} = Complex Coordination, 4 min.; X_{13} = Complex Coordination, 10 min.

Table 6: Multiple Regression Analyses Based upon Total Sample (N=294) and Cross Validated on Racial Subgroups (N=147 for both White and Negro Subgroups)

Criterion	Total Sample Equation	R Total Sample	R White C-V	R Negro C-V
Street Errors	$Y = 21.297 - .637\ X_7 + .083\ X_{13}$	.18	.18*	-.13
Accidents	$Y = 2.739 - .017\ X_1 - .006\ X_{11} - .008\ X_{13}$	.23	.22*	.18*
Violations	$Y = 3.447 - .038\ X_{12}$	.08	-.01	.05

* $p < .05$

Note: X_i's refer to same variables as described in Table 5.

as compared to those developed in the subgroups separately is probably due to two factors, viz., the greater stability in the beta weights due to increased back sample size and the fact that the cross validation samples were included as part of the back sample.

Study #2: North Carolina Students:

The Prediction of Success for Ninth and Twelfth Grade Students

Sample

Academic data from a stratified random sample of 48 classes in the North Carolina public school system were obtained from the Institute for Behavioral Research in Creativity.[1] A large number of these classes were not racially integrated. Thus, rather than analyze the data for the total group which would include segregated classes, the authors have chosen to present data on two racially integrated classes - a ninth grade class and a twelfth grade class.[2] The ninth grade class consisted of 221 students, 166 white and 55 Negro; while the twelfth grade class was composed of 245 whites and 58 Negroes.

Predictors

Test scores on the California Test of Mental Maturity, a measure of general mental ability, were obtained for both the ninth and twelfth grade samples. Also included as predictors were four keys of the Alpha Biographical Inventory developed by the Institute for Behavioral Research in Creativity. This inventory consists of 300 items about childhood activities, experiences, sources of derived satisfaction and dissatisfaction, academic experiences, attitudes and interests, value preferences, and self-description evaluations. The four keys were Rank in Class, Grade Point Average, Leadership, and Creativity. The Rank in Class key was not available for the ninth grade sample.

---

(1) The authors would like to thank Robert Lacklen of the Richardson Foundation and Robert Ellison of the Institute for Behavioral Research in Creativity for providing these data.

(2) An analysis of the total sample has recently been reported by Ellison, James, Fox, and Taylor (1970).

The total sample was divided into approximately equal groups based on an odd-even split. The odd cases were used to develop the Rank in Class Key while the even cases were used as a cross-validation sample. The Grade Point Average Key, the Leadership Key, and the Creativity Key had been previously developed and thus the validity data presented are examples of validity generalization.

Table 7 presents the means, standard deviations, and tests of significance between means for the two racial groups in the ninth grade sample. White students scored higher than Negro students on the California Test of Mental Maturity. There was no difference between the mean performance of the two racial groups on any of the keys developed on the Alpha Biographical Inventory.

Mean scores for the two racial groups in the twelfth grade sample are presented in Table 8. As in the ninth grade sample, white students scored higher than Negro students on the California Test of Mental Maturity. Moreover, white students scored higher than Negro students on the Grade Point Average Key, the Leadership Key, and the Creativity Key of the Alpha Biographical Inventory. There was no difference between the mean scores of the two groups on the Rank in Class Key.

Criteria

The following different criterion measures were obtained: Classroom grades in English, math, science, social studies, and a total Grade Point Average. Also for the twelfth grade sample Rank in Class (position divided by class size) was computed.

Two teacher ratings were obtained on all students. The first, a Leadership rating, was an average of the student's ratings on the following traits: Participation in class discussion, Initiative, Dominance, Initiative of disruptive activities, Acceptance by others, Consideration, Interest in getting

Table 7: Predictors: Means, Standard Deviations and N's

Ninth Grade Students

Predictor	Group	$\bar{X}$	s	N	t
California Test of Mental Maturity	Total	103.60	15.47	202	
	White	106.32	15.28	157	
	Negro	94.11	12.15	45	4.91**
Alpha Biographical Inventory					
Average GPA Key	Total	102.51	21.86	221	
	White	103.44	22.22	166	
	Negro	99.71	20.68	55	1.09
Leadership Key	Total	102.03	22.76	221	
	White	102.69	23.46	166	
	Negro	100.05	20.56	55	.74
Creativity Key	Total	102.28	24.65	221	
	White	102.98	25.36	166	
	Negro	100.18	22.47	55	.73

** $p < .01$

Table 8: Predictors: Means, Standard Deviations and N's

Twelfth Grade Sample

Predictor	Group	$\bar{X}$	s	N	t
California Test of Mental Maturity	Total	105.11	14.15	241	
	White	107.71	13.22	205	
	Negro	90.33	9.46	36	9.41**
Alpha Biographical Inventory					
Rank in Class Key	Total	99.44	20.31	303	
	White	90.73	20.97	245	
	Negro	102.47	17.12	58	1.26
Average GPA Key	Total	102.19	18.75	303	
	White	103.55	19.22	245	
	Negro	96.48	15.51	58	2.60**
Leadership Key	Total	104.17	20.42	303	
	White	105.61	20.73	245	
	Negro	98.10	17.97	58	2.53*
Creativity Key	Total	104.30	22.06	303	
	White	105.85	22.41	245	
	Negro	97.74	19.36	58	2.53*

* $p < .05$
** $p < .01$

an education, Involvement in personal academic activities. The second, a creativity rating, was an average of the students ratings on the following traits: Generalization of new ideas, Seeks new solutions, Independence, Defends own ideas, Quality and quantity of questions, Originality, Enjoyment of complex problems, Participation in open-ended discussions, Interest in outside activites.

Mean criterion data for the ninth grade sample is presented in Table 9. White students obtained significantly higher grades than Negro students in English, math, and social studies as well as the overall Grade Point Average. There was no difference between the mean grades of the two racial groups in science. White students also received higher teacher ratings on Leadership and Creativity than Negro students.

Table 10 presents the mean criterion data for the twelfth grade sample. White students obtained higher grades and teacher ratings than Negro students on all the criterion measures.

Validity

Correlations between the criterion and predictor measures for the ninth grade sample are presented in Table 11. All predictors correlated significantly with the criterion measures for both racial groups. The relationship between the California Test of Mental Maturity and English grades was the only instance where a difference in the validity occurred between the racial groups. The Negro sample validity-coefficient was significantly greater than the white sample coefficient. However, both coefficients were significantly related to the criterion.

Correlations between the predictors and criterion for the twelfth grade sample are presented in Table 12. All of the predictors correlated significant with the Rank in Class criterion and teacher ratings of Creativity for

Table 9: Criteria: Means, Standard Deviations and N's

Ninth Grade Sample

Criterion	Group	$\bar{X}$	s	N	t
Grades					
English	Total	23.56	9.77	219	
	White	25.37	9.49	164	
	Negro	18.18	8.63	55	4.95**
Math	Total	21.82	10.64	219	
	White	23.05	10.47	164	
	Negro	18.18	10.38	55	2.98**
Science	Total	22.24	11.77	219	
	White	23.11	12.16	164	
	Negro	19.64	10.18	55	1.90
Social Studies	Total	23.06	9.87	219	
	White	24.27	9.66	164	
	Negro	19.45	9.70	55	3.18**
Average Grade Point	Total	22.89	9.08	219	
	White	24.17	8.91	164	
	Negro	19.07	8.60	55	3.69**
Teacher Ratings					
Leadership	Total	29.73	7.26	219	
	White	30.92	7.02	164	
	Negro	26.13	6.84	55	4.39**
Creativity	Total	25.18	9.59	219	
	White	26.62	9.44	164	
	Negro	20.82	8.76	55	4.00**

** $p < .01$

Table 10: Criteria: Means, Standard Deviations and N's

Twelfth Grade Sample

Criterion	Group	$\bar{X}$	s	N	t
Rank in Class	Total	44.35	28.41	299	
	White	41.26	27.87	244	
	Negro	58.07	26.89	55	4.05**
Grades					
English	Total	19.93	7.94	299	
	White	20.99	7.76	243	
	Negro	15.36	7.13	56	4.95**
Math	Total	20.61	11.26	244	
	White	21.33	11.37	210	
	Negro	16.18	9.54	34	2.49*
Science	Total	17.84	10.25	176	
	White	19.15	10.21	142	
	Negro	12.35	8.55	34	3.57**
Social Studies	Total	21.13	10.16	293	
	White	22.01	10.42	239	
	Negro	17.22	7.87	54	3.76*
Average Grade Point	Total	19.87	8.29	300	
	White	20.91	8.32	244	
	Negro	15.34	6.52	56	5.42**
Teacher Ratings					
Leadership	Total	29.59	6.87	303	
	White	30.40	6.83	245	
	Negro	26.16	6.00	58	4.33**
Creativity	Total	27.94	8.44	303	
	White	29.06	8.34	245	
	Negro	23.21	7.17	58	4.91**

* $p < .05$
** $p < .01$

Table 11: Predictor - Criterion Correlations

Total Group, Whites and Negroes

Ninth Grade Sample(1, 2)

Criterion		Predictor: Alpha Biographical Key					
		CTMM	N	G.P.A.	Leadership	Creativity	N
Grades							
English	Total	48**	201	55**	46**	47**	219
	White	38**	156	53**	43**	45**	164
	Negro	65**@ (1)	45	64** (3)	61** (3)	61** (3)	55
Math	Total	55**	201	55**	50**	50**	219
	White	52**	156	53**	43**	49**	164
	Negro	54** (1)	45	64** (3)	51** (3)	47** (3)	55
Science	Total	53**	201	68**	62**	64**	219
	White	52**	156	70**	63**	64**	164
	Negro	47** (2)	45	61**	58**	60**	55
Social Studies	Total	42**	201	61**	55**	56**	219
	White	34**	156	59**	52**	54**	164
	Negro	56** (1)	45	69** (3)	65** (3)	66** (3)	55
Average Grade Point	Total	58**	201	70**	62**	63**	219
	White	52**	156	70**	62**	63**	164
	Negro	62** (1)	45	73** (3)	69** (3)	69**	55
Teacher Ratings							
Leadership	Total	58**	202	72**	71**	71**	221
	White	54**	157	70**	71**	70**	166
	Negro	52** (1)	45	82** (3)	79** (3)	79** (3)	55
Creativity	Total	64**	202	64**	64**	64**	221
	White	62**	157	62**	64**	64**	166
	Negro	53** (1)	45	71** (3)	69** (3)	70** (3)	55

(1) Decimals are omitted.

(2) Number in parentheses below the correlation for the Negro sample indicates the model illustrated.

* p .05

** p .01

@ Indicates those models in which a significant difference exists between the validity coefficients for the two ethnic groups.

Table 12: Predictor - Criterion Correlations

Total Group, Whites and Negroes

Twelfth Grade Sample(1, 2)

Criterion		Predictor: Alpha Biographical Inventory				
		CTMM	N	Rank in Class	G.P.A.	N
Rank in Class	Total	- 54**	238	67**	- 69**	299
	White	- 53**	204	71**	- 72**	244
	Negro	- 55** (1)	34	50**@ (3)	- 47**@ (1)	55
Grades						
English	Total	50**	237	- 54**	57**	299
	White	50**	203	- 61**	62**	243
	Negro	15@ (8)	34	- 17@ (6)	18@ (8)	56
Math	Total	28	198	- 43**	45**	244
	White	26	178	- 46**	47**	210
	Negro	38	20	- 12@ (6)	09@ (8)	34
Science	Total	42**	138	- 59**	59**	176
	White	39**	115	- 63**	62**	142
	Negro	27 (8)	23	- 39* (3)	36* (1)	34
Social Studies	Total	51**	233	- 61**	62**	293
	White	51**	201	- 66**	66**	239
	Negro	47** (1)	32	- 22@ (6)	22@ (8)	54
Average Grade Point	Total	51**	238	- 63**	.65	300
	White	50**	204	- 69**	69**	244
	Negro	34** (1)	34	- 28*@ (3)	27*@ (1)	56

Criterion		Predictor: Alpha Biographical Inventory		
		Leadership	Creativity	N
Rank in Class	Total	- 58**	- 59**	299
	White	- 60**	- 61**	244
	Negro	- 43** (1)	- 42** (1)	55

Table 12 (Contd)

Criterion		Predictor: Alpha Biographical Inventory — Leadership	Creativity	N
Grades				
English	Total	50**	52**	299
	White	55**	57**	43
	Negro	15@ (8)	15@ (8)	56
Math	Total	34**	34**	244
	White	36**	36**	210
	Negro	07@ (8)	07@ (8)	34
Science	Total	54**	55**	176
	White	55**	56**	142
	Negro	36* (1)	34* (1)	34
Social Studies	Total	56**	56**	293
	White	60**	61**	239
	Negro	19@ (8)	17@ (8)	54
Average Grade Point	Total	56**	57**	300
	White	60**	61**	244
	Negro	24@ (8)	23@ (8)	56

Criterion		Predictor: Alpha Biographical Inventory — CTMM	N	Rank in Class	G.P.A.	N
Teacher Ratings						
Leadership	Total	33**	241	- 50**	50**	303
	White	29**	205	- 52**	51**	245
	Negro	29 (8)	36	- 39** (8)	36** (1)	58
Creativity	Total	42**	241	- 50**	50**	303
	White	36**	205	- 52**	51**	245
	Negro	44** (1)	36	- 37** (3)	34** (1)	58

Table 12 (Contd)

Criterion		Predictor: Alpha Biographical Inventory — Leadership	Creativity	N
Teacher Ratings				
Leadership	Total	55**	55**	303
	White	55**	55**	245
	Negro	44** (1)	43** (1)	58
Creativity	Total	55**	55**	303
	White	56**	56**	245
	Negro	42** (1)	41** (1)	58

(1) Decimals are omitted.
(2) Number in parentheses below the correlation for the Negro sample indicates the model illustrated.
* $p < .05$
** $p < .01$
@ Indicates those models in which a significant difference exists between the validity coefficients for the two ethnic groups.

both racial groups. In two cases, the relationship between Rank in Class and the Rank in Class key and the Grade Point Average key, the validity coefficients for the white sample were significantly higher than those for the Negro sample. English and math grades were predictable for the white sample but not for the Negro sample.

The California Test of Mental Maturity predicted social studies grades and Grade Point Average for both racial groups. However, social studies grades and Grade Point Average were predictable only for the white sample using the Leadership and Creativity keys of the biographical inventory. The Rank in Class key and the Grade Point Average key predicted Grade Point Average for both racial groups; however, the white sample validity coefficients were significantly greater than those for the Negro sample.

Science grades were predictable for the white sample but not for the Negro sample using the California Test of Mental Maturity as the predictor. When the biographical keys are used as predictors, the science grades are equally predictable for the two racial groups.

Teacher ratings of Leadership were predictable for both racial groups using the biographical keys but were only predictable for the white sample when the California Test of Mental Maturity is used as the predictor.

Models Illustrated - Ninth Grade

a) An analysis of the predictor-criterion relationships for the ninth grade sample indicates that eighteen cases of Model 3 were illustrated. (The number in parentheses below the correlation for the Negro sample in Table 11 indicates the model illustrated). In all illustrations of this model white students scored higher than Negro students on the criterion measures but there was no difference between the two groups on the predictors. The test was valid for both racial groups. In every illustration of this model, the use of a single regression line or expectancy table would result in over-prediction of criteria performance for the Negro sample.

b) Six illustrations of Model 1 were illustrated in the relationship between the California Test of Mental Maturity and the criterion measures. White students scored higher on both the predictor and criterion and the test was valid for both racial groups. In this model the difference in predictor performance is associated with a difference in criterion performance, thus bias in the test is not present for predicting this criterion.

c) Only one other model was represented in the ninth grade sample. Model 2 was illustrated in the relationship between the California Test of Mental Maturity and Science grades. Although there was a difference between the mean performance of the two racial groups on the predictor, there was no difference in the criterion. The test was valid for both racial groups. The use of a single regression line would result in over-prediction of criterion performance for white students and under-prediction for Negro students.

The regression tests for the analysis of covariance (Potthoff, 1969) are presented in Table 13. This analysis indicated that there was no difference in the slope of the regression line for the two racial groups in any of the predictor-criterion relationships. However, in all but six of the relationships there was a difference in the intercepts for the two racial groups.

Models Illustrated - Twelfth Grade

Two models, Model 1 and Model 8, occurred quite frequently in the twelfth grade sample.

a) Model 1 is illustrated in the relationship between the Grade Point Average, Leadership, and Creativity Keys of the Alpha and the following criteria: Rank in Class, Science grades, Grade Point Average, and ratings of Leadership and Creativity. In all illustrations of this model white students scored higher than Negro students on both the predictor and criterion and the test was valid for both racial groups. Since the difference in predictor

performance is paralleled by a corresponding difference in criterion performance the test is not biased against either subgroup.

b) Model 8 is illustrative of the situation where there is a difference in both predictor and criterion performance but the test is valid only for one subgroup. This model was illustrated in the relationship between the Grade Point Average, Leadership, and Creativity keys of the Alpha Biographical Inventory and English, math, and social studies grades, as well as Grade Point Average. In all illustrations of this model the test is valid only for the white group, and white students score higher on both the predictor and criteria. Valid predictions can be made since the white group scored higher on both the predictor and criterion. However, the test is clearly not appropriate for the Negro sample.

c) Two other models were also illustrated in the twelfth grade sample. The relationship between the Rank in Class key of the Alpha Biographical Inventory and the criteria, Rank in Class, science grades, Grade Point Average, as well as the Leadership and Creativity ratings illustrate Model 3. There was no difference in the mean scores of the two groups on the biographical key but white students obtained higher grades and ratings. Although the biographical key is valid for both racial groups, the use of a single regression line would result in over-prediction of the criterion for Negro students and under-prediction for white students.

d) The last model observed in this sample was Model 6 as illustrated in the relationship between the Rank in Class biographical key and math and social studies grades. There is no difference in the mean scores of the two groups on the biographical key but white students obtained higher grades. The key is valid only for the white sample. Although this relationship is not likely to provide any differential selection rates for the two groups, the test is not appropriate as a predictor for the Negro sample.

Table (8WC) presents the regression tests for the analysis of covariance (Potthoff, 1969) for the twelfth grade sample. Differences in the stage of the regression line for the two groups were observed in sixteen predictor - criterion relationships. Likewise, differences in intercepts for the two groups were also observed in sixteen relationships.

One must conclude from the results of the twelfth grade sample, that using biographical data as predictors does not necessarily reduce mean differences for different ethnic groups. In this sample only one biographical key yielded nonsignificant differences in mean scores for the two racial groups. In addition, the results from the ninth grade sample illustrate that elimination of racial differences in predictor performance does not always reduce test bias. In fact if there exists a racial difference in criterion performance reduction of differences in mean predictor performance actually produces test bias.

Table 13: Analysis of Covarance for Homogeneity of Regression

Ninth Grade

Criterion	Predictor: Alpha Biographical Inventory — C.T.M.M. F_1(1)	F_2(2)	F_3(3)	G.P.A. Key F_1	F_2	F_3
Grades						
English	6.16** (2,197)	3.43 (1,197)	8.78** (1,198)	14.03** (2,215)	.42 (1,215)	27.71** (1,216)
Math	.91 (2,197)	.67 1,197)	1.15 (1,198)	4.78** (2,215)	1.19 (1,215)	8.36** (1,216)
Science	.10 (2,197)	.03 1,197)	.18 (1,198)	2.24 (2,215)	1.78 (1,215)	2.69 (1,216)
Social Studies	2.83 (2,197)	3.83 (1,197)	1.82 (1,198)	5.96** (2,215)	1.41 (1,215)	10.48** (1,216)
Average Grade Point	1.93 (2,197)	1.83 (1,197)	2.02 (1,198)	8.87** (2,215)	.29 (1,215)	17.51** (1,216)
Teacher Rating						
Leadership	2.02 (2,198)	.34 (1,198)	3.71 (1,199)	14.88** (2,217)	1.90 (1,217)	27.75** (1,218)
Creativity	.58 (2,198)	.01 (1,198)	1.16 (1,199)	9.53** (2,217)	1.90 (1,217)	27.75** (1,218)

Criterion	Predictor: Alpha Biographical Inventory — Leadership Key F_1	F_2	F_3	Creativity Key F_1	F_2	F_3
Grades						
English	14.67** (2,215)	1.64 (1,215)	27.62** (1,216)	14.75** (2,215)	1.39 (1,215)	28.06** (1,216)

Table 13 (Contd)

Ninth Grade

Criterion	Predictor: Alpha Biographical Inventory — Leadership Key F_1	Leadership Key F_2	Leadership Key F_3	Creativity Key F_1	Creativity Key F_2	Creativity Key F_3
Math	5.31** (2,215)	1.30 (1,215)	9.30** (1,216)	5.22** (2,215)	1.02 (1,215)	9.43** (1,216)
Science	1.98 (2,215)	.37 (1,215)	3.59 (1,216)	2.05 (2,215)	.40 (1,215)	3.71 (1,216)
Social Studies	6.99** (2,215)	2.54 (1,215)	11.36** (1,216)	6.91** (2,215)	2.2 (1,215)	11.63** (1,216)
Grade Point Average	9.42** (2,215)	1.13 (1,215)	17.70** (1,216)	9.57** (2,215)	.90 (1,215)	18.24** (1,216)
Teacher Ratings						
Leadership	16.85** (2,217)	1.95 (1,217)	31.62** (1,218)	16.82** (2,217)	2.13 (1,217)	31.34** (2,217)
Creativity	10.99** (2,217)	.59 (1,217)	21.44** (1,218)	11.15** (2,217)	.61 (1,217)	21.72** (2,217)

* $p < .05$

** $p < .01$

(1) F_1 tests hypothesis that $\underline{F}(Y_{ij} \mid X_{ij}) = a + bX_{ij}$ for all i groups

(2) F_2 tests hypothesis that $\underline{F}(Y_{ij} \mid X_{ij}) = a_1 + bX_{ij}$ for all i groups

(3) F_3 tests hypothesis that $\underline{F}(Y_{ij} \mid X_{ij}) = a + b_1X_{ij}$ for all i groups
(valid test only if $\underline{F}$ is not significant)

Table 14: Analysis of Covariance for Homogeneity of Regression

Twelfth Grade

Criterion	Predictor: C.T.M.M. F_1(1)	C.T.M.M. F_2(2)	C.T.M.M. F_3(3)	Alpha Biographical Inventory: Rank in Class Key F_1	Rank in Class Key F_2	Rank in Class Key F_3
Rank in Class	5.71** (2,235)	10.60** (1,235)	.78 (1,236)	11.26** (2,295)	1.37 (1,295)	21.12** (1,296)
Grades						
English	8.11** (2,234)	16.21** (1,234)	.01 (1,235)	18.22** (2,295)	8.78** (1,295)	26.95 (,296)
Math	1.25 (2,195)	2.28 (1,195)	.21 (1,196)	3.70 (2,240)	2.97 (1,240)	4.40 (1,241)
Science	1.09 (2,135)	2.18 (1,135)	.01 (1,136)	9.63** (2,172)	2.72 (1,172)	16.37** (1,173)
Social Studies	6.57** (2,230)	10.30** (1,230)	2.73 (1,231)	11.40** (2,289)	11.19** (1,289)	11.21 (1,290)
Average Grade Point	6.46** (2,235)	12.70** (1,235)	.21 (1,236)	18.88** (2,296)	10.65** (1,296)	26.26 (1,297)
Teacher Ratings						
Leadership	1.81 (2,238)	2.97 (1,238)	.65 (1,239)	9.42** (2,299)	.50 (1,299)	18.36** (1,300)
Creativity	2.47 (2,238)	4.52 (1,238)	.41 (1,239)	12.56** (2,299)	.80 (1,299)	24.34** (1,300)

Table 14 (Contd)

Twelfth Grade

Criterion	Predictor: Alpha Biographical Inventory — G.P.A. Key F_1	F_2	F_3	Leadership Key F_1	F_2	F_3
Rank in Class	6.02** (2,295)	.72 (1,295)	11.34** (1,296)	6.25** (2,295)	1.02 (1,295)	11.48** (1,296)
Grades						
English	13.26** (2,295)	7.64** (1,295)	18.46 (1,296)	13.87** (2,295)	8.67** (1,295)	18.59 (1,296)
Math	2.84 (2,240)	2.34 (1,240)	3.43 (1,241)	2.88 (2,240)	2.36 (1,240)	3.38 (1,241)
Science	6.69** (2,172)	2.69 (1,172)	10.60** (1,173)	6.93** (2,173)	3.01 (1,172)	10.72** (1,173)
Social Studies	8.16** (2,289)	10.61** (1,289)	5.53 (1,290)	8.62** (2,289)	11.57** (1,289)	5.46 (1,290)
Average Grade Point	12.40** (2,296)	8.63** (1,296)	15.77 (1,297)	12.96** (2,296)	9.56** (1,296)	15.90 (1,297)
Teacher Ratings						
Leadership	6.45** (2,299)	.62 (1,299)	12.26** (1,300)	6.50** (2,299)	.74 (1,299)	12.27** (1,300)
Creativity	9.26** (2,299)	1.10 (1,299)	17.42** (1,300)	9.28** (2,299)	1.13 (1,299)	17.41** (1,300)

Criterion	Creativity Key F_1	F_2	F_3
Rank in Class	6.31** (2,295)	1.47 (1,295)	(1,296)
Grades			
English	13.61** (2,295)	8.42** (1,295)	18.34 (1,296)

Table 14 (Contd)

Twelfth Grade

Criterion	Predictor: Alpha Biographical Inventory, Creativity Key F_1	F_2	F_3
Math	3.05 (2,240)	3.16 (1,240)	2.41 (1,241)
Science	6.62** (2,173)	3.13 (1,172)	10.00** (1,173)
Social Studies	8.37** (2,289)	11.11** (1,289)	5.44 (1,290)
Average Grade Point	13.71** (2,296)	10.97** (1,296)	15.92 (1,297)
Teacher Ratings			
Leadership	6.41** (2,299)	.66 (1,299)	12.17** (1,300)
Creativity	8.97** (2,299)	.82 (1,299)	17.13** (1, 00)

* $p < .05$
** $p < .01$

(1) Degrees of freedom for all comparisons: $\underline{F}$ -(2,90); F_2 -(1,920); F_3 (1,91)
(2) F_1 tests hypothesis that $\underline{F}(Y_{ij}|X_{ij}) = a + bX_{ij}$ for all i groups.
(3) F_2 tests hypothesis that $\underline{F}(Y_{ij}|X_{ij}) = a_i + bX_{ij}$ for all i groups.
(4) F_3 tests hypothesis that $\underline{F}(Y_{ij}|X_{ij}) = a + b_iX_{ij}$ for all i groups.
(valid test only if F_2 is not significant)

Study #3: An Examination of the Variables Predicting Success in a University Setting

Sample

The sample consisted of students attending a large state university. Since predictor data was not available for all subjects, the sample sizes for the various predictors are different. The Negro group consisted of all the Negro students who entered the university in September, 1968, while the white group was a random sample of the remaining students who entered that semester.

Predictors

Predictors included:

a) standard academic predictors : Scholastic Aptitude Test - Verbal (SAT-V), Scholastic Aptitude Test - Mathematical (SAT-M), and high school grade point average;

b) non-intellectual variables, including the California Psychological Inventory (CPI), the Holland Vocational Preference Inventory (HVPI);

c) an opinion questionnaire known as the University Student Census (USC).

High school grade point average was a standardized score having a mean of 3.00 and a standard deviation of 1.00, used to equate different high school grading systems. The CPI consists of 18 scales, with an anxiety scale developed by Leventhal (1966), included as a nineteenth scale. The HVPI consists of 11 scales, and the USC includes 17 statements to which students were asked to mark their agreement or disagreement. The statements of the USC are as follows:

30. Most of my courses are stimulating and exciting.

30a. The University should actively recruit black students.

31. Most faculty advisors here act like they really care about students.

32. Most instructors here act like they really care about students.

33. Most administrators here act like they really care about students.

34. The University should use its influence to improve social conditions in the state.

35. University students have ample opportunity to participate in University policy making.

36. I am here for an education; let other people "get involved" on campus.

37. There should be a special college for new students undecided as to their major.

38. The University should suspend students who disrupt the normal operation of the University.

39. Most courses require intensive study and preparation outside the classroom.

40. Most organized student activities on campus are ridiculous.

41. Major University-wide events draw lots of support and enthusiasm.

42. There are many facilities and opportunities on campus for individual creative activities.

43. At the beginning of a course, there is no way to tell who will get what grades.

44. Channels for expressing student complaints are readily available.

45. I was a leader in high school.

Mean predictor scores are given in Table 15. On the three traditional academic predictors, the Negro subgroup scores were significantly lower than the scores of the white subgroup. For the non-academic predictors it becomes less meaningful to speak of one group scoring higher than another. However,

Table 15: Predictors - Means, Standard Deviations, N's,

and Tests of Significance of Mean Differences

Predictor	Group	$\bar{X}$	s	N	t(1)
SAT-V	Total	469.4	97.2	304	
	White	501.8	95.4	178	
	Negro	423.5	80.0	126	7.50**
SAT-M	Total	492.8	106.5	304	
	White	533.1	98.9	178	
	Negro	435.8	89.9	126	8.74**
High School Grade Point Average	Total	2.99	.99	304	
	White	3.17	1.00	178	
	Negro	2.72	.92	126	3.63**
California Psychological Inventory					
Dominance	Total	25.42	6.55	272	
	White	25.72	6.72	193	
	Negro	24.67	6.07	79	1.20
Capacity for Status	Total	17.47	4.17	272	
	White	17.89	4.17	193	
	Negro	16.44	4.03	79	2.62**
Sociability	Total	23.46	5.37	272	
	White	23.74	5.57	193	
	Negro	22.78	4.81	79	1.34
Social Presence	Total	33.21	6.41	272	
	White	33.91	6.63	193	
	Negro	31.52	5.52	79	3.04**
Self Acceptance	Total	21.17	4.22	272	
	White	21.41	4.28	193	
	Negro	20.57	4.02	79	1.49
Sense of Well-Being	Total	31.69	6.75	272	
	White	32.22	6.66	193	
	Negro	30.41	6.85	79	2.01*
Responsibility	Total	28.46	5.44	272	
	White	28.59	5.50	193	
	Negro	28.13	5.31	79	0.63

Table 15 (Contd)

Prediction	Group	$\bar{X}$	s	N	t(1)
Socialization	Total	36.52	6.89	272	
	White	36.53	7.05	193	
	Negro	36.51	6.50	79	0.02
Self-Control	Total	24.46	8.66	272	
	White	24.09	8.82	193	
	Negro	25.37	8.25	79	1.10
Tolerance	Total	19.03	5.23	272	
	White	19.65	5.19	193	
	Negro	17.51	5.05	79	2.20*
Good Impression	Total	14.42	6.12	272	
	White	14.22	6.27	193	
	Negro	14.91	5.74	79	0.96
Communality	Total	24.77	3.41	272	
	White	25.04	3.34	193	
	Negro	24.11	3.53	79	2.04*
Achievement via Conference	Total	24.33	5.28	272	
	White	24.13	5.45	193	
	Negro	24.82	4.84	79	0.97
Intellectual Efficiency	Total	34.92	6.47	272	
	White	35.47	6.19	193	
	Negro	33.57	7.04	79	2.20*
Psychological Mindedness	Total	9.81	2.92	272	
	White	9.93	3.01	193	
	Negro	9.49	2.69	79	1.12
Flexibility	Total	9.68	4.04	272	
	White	10.04	4.12	193	
	Negro	8.80	3.74	79	2.30*
Femininity	Total	20.36	5.19	272	
	White	20.34	5.22	193	
	Negro	20.42	5.17	79	0.11
Leventhal's Anxiety Scale	Total	5.41	2.41	272	
	White	5.40	2.42	193	
	Negro	5.42	2.41	79	0.06
Holland Vocation Performance Index					
Realistic	Total	1.90	2.67	230	
	White	1.87	2.54	164	
	Negro	2.27	2.96	66	0.96

Table 15 (Contd)

Predictor	Group	$\bar{X}$	s	N	t(1)
Intellectual	Total	4.21	4.18	230	
	White	4.06	4.10	164	
	Negro	4.59	4.36	66	0.87
Social	Total	5.10	3.92	230	
	White	4.82	3.89	164	
	Negro	5.80	3.93	66	1.72
Conventional	Total	2.04	2.93	230	
	White	1.59	2.51	164	
	Negro	3.17	3.55	66	3.28**
Enterprising	Total	2.85	2.93	230	
	White	2.79	2.87	164	
	Negro	3.00	3.08	66	0.49
Artistic	Total	4.19	3.98	230	
	White	3.90	3.82	164	
	Negro	4.92	4.31	66	1.76
Self Control	Total	10.07	3.59	230	
	White	9.93	3.50	164	
	Negro	10.41	3.84	66	0.91
Masculinity	Total	6.25	3.09	230	
	White	6.23	3.11	164	
	Negro	6.30	3.04	66	0.15
Status	Total	8.29	2.28	230	
	White	8.19	2.21	164	
	Negro	8.55	2.47	66	1.07
Infrequency	Total	6.01	2.95	230	
	White	5.78	2.89	164	
	Negro	6.59	3.03	66	1.89
Acquiescence	Total	9.31	5.03	230	
	White	8.95	4.52	164	
	Negro	10.21	6.05	66	1.52
University Student Census					
#30	Total	2.47	1.08	246	
	White	2.39	1.00	141	
	Negro	2.58	1.16	105	1.34
#30a	Total	1.64	1.49	250	
	White	2.34	1.32	144	
	Negro	0.68	1.13	106	10.39**

Table 15 (Contd)

Predictor	Group	$\bar{X}$	s	N	t(1)
#31	Total	2.15	1.18	250	
	White	2.03	1.11	144	
	Negro	2.29	1.25	106	1.70
#32	Total	2.27	1.13	248	
	White	2.00	0.93	143	
	Negro	2.63	1.27	105	4.29**
#33	Total	2.57	1.19	250	
	White	2.43	1.14	144	
	Negro	2.75	1.25	106	2.10*
#34	Total	1.70	1.37	250	
	White	1.86	1.25	144	
	Negro	1.49	1.49	106	2.07*
#35	Total	2.38	1.32	250	
	White	2.17	1.17	144	
	Negro	2.66	1.47	106	2.82**
#36	Total	2.76	1.20	250	
	White	2.80	1.24	144	
	Negro	2.69	1.15	106	0.71
#37	Total	1.98	1.31	250	
	White	1.99	1.34	144	
	Negro	1.95	1.28	106	0.24
#38	Total	2.27	1.34	149	
	White	1.92	1.32	144	
	Negro	2.73	1.24	105	4.89**
#39	Total	1.46	1.14	250	
	White	1.65	1.12	144	
	Negro	1.22	1.13	106	2.98**
#40	Total	2.69	1.11	250	
	White	2.68	1.05	144	
	Negro	2.70	1.98	106	0.14
#41	Total	2.05	1.26	250	
	White	1.94	1.11	144	
	Negro	2.20	1.44	106	1.54
#42	Total	1.67	1.21	250	
	White	1.53	1.10	144	
	Negro	1.84	1.34	106	1.94

Table 15 (Contd)

Predictor	Group	$\bar{X}$	s	N	t(1)
#43	Total	1.69	1.33	250	
	White	1.76	1.30	144	
	Negro	1.59	1.36	106	1.00
#44	Total	2.32	1.14	249	
	White	2.25	1.04	144	
	Negro	2.42	1.27	105	1.12
#45	Total	1.81	1.26	247	
	White	1.90	1.29	142	
	Negro	1.68	1.20	105	1.36

(1) t ratios are between the means of the white and Negro samples

* $p < .05$

** $p < .01$

there were significant differences between the racial groups on eight of the nineteen CPI scales, one of the eleven HVPI scales, and seven of the seventeen USC items.

In the eight instances where mean differences were found on the CPI scales, the mean of the white subgroup was of greater magnitude than the mean of the Negro subgroup. The scales included Capacity for Status, Social Presence, Sense of Well-Being, Tolerance, Communality, Achievement via Independence, Intellectual Efficiency, and Flexibility. The only HVPI scale for which a mean difference was found was the Conventional scale, on which the Negro subgroup has a greater mean score than the white subgroup. In four of the seven instances where significant differences were found on USC items, the white subgroup agreed more with the statement than the Negro subgroup. These statements were: "Most instructors here act like they really care about students"; "Most administrators here act like they really care about students"; "University students have ample opportunity to participate in University policy making"; and "The University should suspend students who disrupt the normal operation of the University." The Negro subgroup agreed more than the white subgroup with the following statements: "The University should actively recruit black students"; "The University should use its influence to improve social conditions in the state"; and "Most courses require intensive study and preparation outside the classroom."

Criterion

The criterion used was grade point average at the end of the freshman year. Since the sample sizes were substantially different for the various predictors, a separate test for criterion differences was made for each of the predictors. In every case, the Negro sample scored significantly lower than the white sample on the criterion measure. The criterion data for each predictor group is presented in Table 16.

Table 16: Criterion - Means, Standard deviations, N's, and Tests of Significance of Mean Differences for each Group of Predictors

Criterion	Group	$\bar{X}$	s	N	t(1)
Grade Point Average (Academic Predictors)	Total	2.00	.84	304	
	White	2.22	.88	178	
	Negro	1.69	.68	126	5.90**
Grade Point Average (CPI as a Predictor)	Total	2.06	.86	272	
	White	2.20	.88	193	
	Negro	1.70	.70	79	4.92**
Grade Point Average (HVPI as a Predictor)	Total	2.14	.81	230	
	White	2.27	.83	164	
	Negro	1.82	.67	66	4.26**
Grade Point Average (USC as a Predictor)	Total	2.21	.88	251	
	White	2.47	.63	144	
	Negro	1.88	.59	107	7.65**

(1) t ratios are between the means of the white and Negro samples.

** $p < .01$

Validity

Table 17 presents the validity coefficients for the total, white, and Negro samples. The standard academic predictors (SAT-V, SAT-M, and high school grades) are the best predictors of college success for both racial groups. The validity coefficients of the SAT's are of greater magnitude for the Negro subgroup, but high school grades do not predict as well for the Negro group as for the white group.

The magnitude of the non-intellectual validity coefficients is generally lower than the academic variables. Nevertheless, ten of the nineteen CPI scales were significantly related to university grade point average for the white group; and seven were significantly related for the Negro group. Similarly, five HVPI scales reached significance for the white sample, and three for the Negro sample. Of the seventeen USC items only three exhibited a significant relationship for the Negro group, and none for the white sample.

Models Illustrated

In twenty-two of the fifty predictor-criterion relationships examined, the predictor was not valid for either racial group or for the total group. The remaining cases illustrated the following Bartlett-O'Leary models (1969):

a) There were six cases of Model 1, in which the predictor was valid for both racial groups although there were significant mean differences on both the predictor and the criterion. Since this procedure would result in the selection of those most likely to succeed, there is justification for employing the predictor in a selection decision. All three standard academic predictors are found among the examples of Model 1. However, the significant validity coefficient for the Negro high school grade point average is much lower than that obtained for the white sample and the test for homogeneity of of regression indicated that the regression slopes were different for the two racial groups.

Table 17: Predictor - Criterion Correlations Total Group, Whites, Negroes[1, 2]

Criterion:	Predictor: Group	SAT-V	SAT-M	High School Grades			N
GPA	Total	54**	43**	60**			304
	White	46**	31**	64**			178
	Negro	50** (1)	41**	44**a			126
		CPI Dominance	CPI Capacity for Status	CPI Sociability	CPI Social Presence	CPI Self-Acceptance	
	Total	12*	19**	05	-01	04	272
	White	12	15	04	-10	01	193
	Negro	07 (11)	20 (11)	-01	09	06	79
		CPI Sense of Well Being	CPI Responsibility	CPI Socialization	CPI Self Control	CPI Tolerance	
	Total	20**	26**	26**	15*	28**	272
	White	15*	25**	23**	20**	27**	193
	Negro	26* (1)	26* (3)	38** (3)	10 (6)	15 (8)	79
		CPI Good Impression	CPI Communality	CPI Ach. via Conformance	CPI Ach. via Indep.	CPI Intell. Eff.	
	Total	10	21**	21**	35**	27**	272
	White	13	13	22**	32**	23**	193
	Negro	10	34** (8)	28* (3)	35** (1)	28* (1)	79

Table 17 (Contd)

Criterion:	Group	CPI Psyc. Mindedness	CPI Flexibility	CPI Femininity	CPI Anxiety Scale	N
		Predictor				
GPA	Total	12*	14*	17*	-10	272
	White	16*	09	17*	-11	193
	Negro	-08	16	19	-07	79
		(6)	(11)	(6)		

	Group	HVPI Realistic	HVPI Intellectual	HVPI Social	HVPI Conventional	HVPI Enterprising	N
	Total	-09	11	19** 8	-02	02	230
	White	-13	16*	23**	-03	02	164
	Negro	04	02	25*	20	08	66
			(6)	(3)			

	Group	HVPI Artistic	HVPI Self-Control	HVPI Masculinity	HVPI Status	HVPI Infrequency	N
	Total	18**	04	-25**	16*	-07	230
	White	25**	01	-25**	21**	-16a	164
	Negro	12	17	-26*	12	31*	66
		(6)		(3)	(6)	(6)	

	Group	HVPI Acquiescence	N
	Total	05	230
	White	10	164
	Negro	04	66

Table 17 (Contd)

Criterion:	Group	Predictor					
		USC-#30	N	USC-#30a	N	USC-#31	N
GPA	Total	-13*	246	25**	250	02	250
	White	-02	141	04	144	07	144
	Negro	-22* (6)	105	01	106	05	106
		USC-#32	N	USC-#33	N	USC-#34	N
	Total	-26**	248	-02	250	-08	250
	White	-10	143	05	144	-08	144
	Negro	-23* (8)	105	00	106	-23* (8)	106
		USC-#35	N	USC-#36	N	USC-#37	N
	Total	-01	250	10	250	04	250
	White	08	144	09	144	12	144
	Negro	07	106	09	106	-11	106
		USC-#38	N	USC-#39	N	USC-#40	N
	Total	-05	249	17**	250	-01	250
	White	11	144	08	144	03	144
	Negro	01	105	15 (11)	106	-07	106
		USC-#41	N	USC-#42	N	USC-#43	N
	Total	04	250	09	250	05	250
	White	08	144	04	144	07	144
	Negro	09	106	06	106	06	106

Table 17 (Contd)

Criterion:	Predictor Group	USC-#44	N	USC-#45	N
GPA	Total	04	249	-05	247
	White	07	144	-10	142
	Negro	10	105	-09	105

(1) Decimals are omitted.

(2) Number in parentheses below the correlation for the Negro sample indicates the model illustrated.

* $p < .05$

** $p < .01$

a indicates those models in which a significant difference exists between the validity coefficients for the two ethnic groups.

b) There are five examples of Model 3 in the data: three CPI scales and two HVPI scales. Although these scales are equally valid in the racial groups, the significantly lower criterion performance of the Negro sample would lead to a selection of more Negroes who would fail to meet criterion standards.

c) For the models describing differential validity, there were eight examples of Model 6 where there was no difference for the predictor scores but one subgroup obtained higher criterion scores, and four examples of Model 8 where there were significant subgroup mean differences in both predictor and criterion performance but significant validity for only one subgroup. In seven cases the predictor was valid for only the white sample, and in five cases for only the Negro subgroup.

d) Five examples of Model 11 were found. In these cases, although the predictor appears valid when looking at the total group, it has no validity in either racial group.

There are two cases which illustrate a more stringent requirement that the validity coefficients of the racial groups be significantly different for identification of differential validity. The first is the Infrequency scale of the HVPI, which was identified as a case of Model 6 by the first method of analysis. The second instance is high school grade point average which was identified as Model 1 by the first method of analysis. Even though high school grades are a valid predictor ($p < .01$) for both racial groups, the validity coefficient is sufficiently lower for the Negro group to make the difference in coefficients significant ($p < .05$).

Because of the large criterion differences between the Negro and white groups, the test for homogeneity of regression indicates a separate intercept in each case in which the test is appropriate. (See Table 18)

Table 18: ANACOVA for Homogeneity of Regression[1]

Predictor	Criterion GPA		
	F_1[2]	F_2[3]	F_3[4]
SAT-V	2.30	.00	4.61*
SAT-M	3.15*	.13	6.19*
High School	12.57**	8.84**	15.89**
GPA	(2,300)	(1,300)	(1,301)
CPI			
Dominance	9.48**	.18	18.83**
Capacity for Status	8.14**	.01	16.32**
Sociability	9.73**	.11	19.41**
Social Presence	11.37**	1.79	20.90**
Self-Acceptance	9.71**	.08	19.42**
Sense of Well-Being	8.57**	.13	17.05**
Responsibility	9.85**	.08	19.67**
Socialization	10.90**	.53	21.30**
Self-Control	11.45**	.77	22.15**
Tolerance	7.68**	1.26	14.10**
Good Impression	10.51**	.09	20.99**
Communality	9.03**	1.10	16.96**
Achievement Via Conf.	11.51**	.05	23.05**
Ach. Via Indep.	7.13**	.02	14.29**
Intellectual Effic.	8.04**	.08	16.05**
Psychol. Mindedness	10.92**	2.78	18.94**
Flexibility	8.81**	.11	17.57**
Femininity	10.30**	.02	20.65**
Leventhal's Anx. Scale	10.04**	.14	20.00**
Holland			
Realistic	8.07**	1.50	14.60**
Intellectual	8.72**	1.16	16.27**
Social	9.62**	.04	19.28**
Conventional	8.66**	1.79	15.48**
Enterprising	7.69**	.10	15.35**
Artistic	10.33**	1.69	18.91**
Self-Control	8.16**	.76	15.58**
Masculinity	7.94**	.06	15.88**
Status	9.14**	.99	17.28**
Infrequency	12.04**	9.14**	14.43**
Acquiescence	8.35**	.49	16.24**
	(2,226)	(1,226)	(1,227)

Table 18 (Contd)

Predictor	Criterion GPA		
	F_1	F_2	F_3
USC			
#30	28.40**	2.81	53.59**
#30a	17.82**	.05	35.73**
#31	28.32**	.18	56.64**
#32	22.25**	.31	44.31**
#33	27.98**	.36	55.75**
#34	30.60**	.75	60.52**
#35	28.64**	.18	57.29**
#36	27.17**	.02	54.55**
#37	29.61**	3.40	55.27**
#38	28.97**	.91	57.06**
#39	25.97**	.35	49.72**
#40	27.96**	.59	55.43**
#41	28.36**	1.67	54.91**
#42	26.78**	.06	53.70**
#43	28.03**	1.19	54.82**
#44	28.16**	.00	56.55**
#45	28.81**	.00	57.86**

(1) Degrees of freedom for each ratio are shown in parentheses below each column.

(2) F_1 tests hypothesis that $E(Y_{ij} \mid X_{ij}) = a + b\, X_{ij}$ for all i groups.

(3) F_2 tests hypothesis that $E(Y_{ij} \mid X_{ij}) = a_i + b\, X_{ij}$ for all i groups.

(4) F_3 tests hypothesis that $E(Y_{ij} \mid X_{ij}) = a_i + b_i\, X_{ij}$ for all i groups.

* $p < .05$

** $p < .01$

Multiple Regression Analysis

In addition to the zero-order correlational analysis, a multiple regression analysis was performed using SAT-V, SAT-M, and high school grade point average. The results of this analysis are found in Table 19. Both racial groups appear equally predictable using the multiple predictor. There is a difference of only .02 between the cross-validated coefficients of the groups. Although sample size did not permit cross-validation, multiple regression analyses were performed for sexually and racially divided groups. Table 19 shows the multiple R's for the groups to be very similar, ranging from .61 to .67.

Table 19: Cross-validated Multiple R's

Predictors: SAT-V, SAT-M, and High School Grade Point Average

	Multiple R	Cross-validated R	N
Total	.67	.62	152
White	.62	.61	89
Negro	.61	.59	63
White Male	.63		79
Negro Male	.61		64
White Female	.65		99
Negro Female	.67		62

Study #4: The Prediction of Work Success in a Health Insurance Company

Sample

The sample consisted of 209 workers for a large health insurance company. Of the 209, 158 were white and 51 were Negro. The employees worked as approvers, coders, keypunch operators, special assistants, and computer operators.

Predictors

The first predictor was the Thurstone Test of Mental Alertness (TMA) which provided a verbal, quantitative, and total score. The second predictor, the Pictorial Reasoning Test, was developed by Science Research Associates with a non-verbal format and has been referred to as a culture-free test. Both tests were administered to new employees but were not used for selection purposes.

Significant differences in performance were found ($p < .01$) between the white and Negro groups on the verbal, quantitative, and total scales of the TMA, but no significant difference was found on the Pictorial Reasoning Test. Predictor means are presented in Table 20.

Criteria

All seven criteria were ratings of workers by their supervisors along a five point scale. Workers were rated on their quantity of work, quality of work, accuracy, knowledge of their job, job aptitude, flexibility, and overall effectiveness. The mean rating for the white group was significantly ($p < .05$) higher than that for the Negro group on three criteria: quantity of work, job aptitude, and overall effectiveness. Criterion means for the total group and the two subgroups are shown in Table 21.

Table 20: Predictors - Means, Standard Deviations, N's, and Tests of Significance of Mean Differences

Total Insurance Workers

Predictor	Group	$\bar{X}$	s	N	t(1)
Thurstone Test of Mental Alertness					
Verbal	Total	31.42	10.92	209	
	White	33.99	9.78	158	
	Negro	23.47	10.38	51	6.55**
Quantitative	Total	22.67	8.82	209	
	White	24.02	8.49	158	
	Negro	18.49	8.50	51	4.02**
Total	Total	54.14	18.55	209	
	White	58.07	17.03	158	
	Negro	41.96	17.73	51	5.79**
Pictorial Reasoning Test	Total	46.64	8.04	209	
	White	47.02	8.09	158	
	Negro	45.47	7.77	51	1.20

(1) t ratios are between the means of the white and Negro samples.

** $p < .01$

Table 21: Criteria - Means, Standard Deviations, N's,

and Tests of Significance of Mean Differences

Total Insurance Workers

Criterion	Group	$\bar{X}$	s	N	t(1)
Quantity of Work	Total	3.46	.68	209	
	White	3.52	.67	158	
	Negro	3.25	.69	51	2.47*
Quality of Work	Total	3.53	.75	209	
	White	3.55	.77	158	
	Negro	3.43	.73	51	.98
Accuracy	Total	3.38	.78	209	
	White	3.41	.78	158	
	Negro	3.27	.75	51	1.12
Knowledge of Job	Total	3.58	.68	209	
	White	3.63	.63	158	
	Negro	3.49	.81	51	1.65
Job Aptitude	Total	3.70	.75	208	
	White	3.75	.73	157	
	Negro	3.49	.81	51	2.14*
Flexibility	Total	3.16	.89	208	
	White	3.22	.89	157	
	Negro	2.98	.86	51	1.68
Overall Effectiveness	Total	3.53	.74	208	
	White	3.59	.76	157	
	Negro	3.33	.68	51	2.17*

(1) t ratios are between the means of the white and Negro samples.

* $p < .05$

Validity

Predictor-criterion correlations are presented in Table 22. Only one relationship reached significance ($p < .05$) for the Negro sample, while eight correlations were significant for the white subgroup. The correlation trends for the two groups were different with 25 of 28 correlations for the Negro group being negative correlations while all the correlations for the white group were positive. Quantity of work was the most predictable criterion for the white group. All four test scales correlated significantly ($p < .05$) with the ratings on that measure. Two of the TMA scales were significantly related to quality of work and flexibility for the white group. The only significant relationship found for the Negro group was between the Pictorial Reasoning Test and ratings of overall effectiveness. Although the Pictorial Reasoning Test tends to minimize the difference in mean performance between the white and Negro groups, its validity is no higher than that of the TMA.

Models Illustrated

a) Three examples of Model 8 were illustrated in the prediction of quantity of work from TMA scales. There is a significant mean difference on both the predictor and the criterion, with a significant validity coefficient for only the white group. If the TMA were used to select for the criterion, quantity of work, these examples of Model 8 would not lead to unfair discrimination since the mean differences fall on both the predictor and the criterion. However, the low validity coefficients preclude any practical use of the predictors even though they are significant for the total group.

b) Three examples of Model 7 were found. The TMA-Quantitative scale and the TMA-Total scale are valid predictors of quality of work for the white group, and the TMA-Total scale is a valid predictor of flexibility for the white group. However, none of these correlations are significant for the Negro group. Moreover,

Table 22: Predictor-Criterion Correlations

Total Group, Whites, and Negroes

Total Insurance Workers(1, 2)

Criterion		Predictor: Thurstone Test of Mental Alertness: Verbal	Quantitative	Total	Pictorial Reasoning	N
Quantity of Work	Total	15*	14*	15*	11	209
	White	15*	17*	18*	18*	158
	Negro	-13 (8)	-13 (8)	-14 (8)	-19a (6)	51
Quality of Work	Total	08	10	09	03	209
	White	13	18*	16*	09	158
	Negro	-18	-19a (7)	-20a (7)	-20	51
Accuracy	Total	01	01	01	-01	209
	White	02	00	01	00	158
	Negro	-15	-06	-12	-09	51
Knowledge of Job	Total	02	04	03	04	209
	White	03	09	06	13a	158
	Negro	-17	-20	-20	-22	51
Job Aptitude	Total	05	06	06	-03	208
	White	06	09	07	01	158
	Negro	-19	-16	-19	-21	51
Flexibility	Total	16*	13	16*	04	208
	White	17*	14	16*	06	157
	Negro	00 (7)	01	01 (7)	-08	51
Overall Effectiveness	Total	09	08	09	-01	208
	White	11a	11	11a	05a	157
	Negro	-24	-18	-23	-28*	51

(1) Decimals are omitted.

(2) Number in parentheses below the correlation for the Negro sample (6) indicates the model illustrated (see Appendix A).

* $p < .05$

a indicates significant difference between white and Negro correlations.

since Negroes score significantly lower on the predictors while there were no differences on the criterion, the use of the TMA as a selection instrument would be unfair to the Negroes.

c) There were two examples of Model 6 where there were no significant differences on the predictor, but the white group scored significantly higher on the criterion. In one case, the predictor (Pictorial Reasoning Test) was valid for the white group alone; and in one case, for the Negro group alone.

Applying the more stringent requirement that a significant difference exist between the Negro and white correlations, four of the nine models described above remain. A significant difference between Negro and white correlations was found in two other instances, but in both cases the predictor was inappropriate since it had not correlated with the criterion in either subgroup.

The results of the analysis of covariance for homogeneity of regression, given in Table 23, indicated that bias exists in nine instances (as determined by a significant F_1 statistic). Four of these instances involved the prediction of quantity of work; three involved prediction of knowledge of job; and two instances were found in the prediction of overall effectiveness. The test indicated that separate regression slopes were appropriate in six instances: the prediction of quantity of work by TMA-Q, TMA-T, and the Pictorial Reasoning Test; the prediction of quality of work by the TMA-Q and the TMA-T; and the prediction of knowledge of job by the Pictorial Reasoning Test. In the prediction of job aptitude by the TMA-V, one F_3 statistic was significant indicating that separate intercepts are appropriate for the two racial groups.

Cultural Deprivation

Since the TMA is assumed to be culture bound and the Pictorial Reasoning Test purports to be a culture-free test, there is an opportunity to compare the predictability of the two tests in various population subgroups. If the criterion

Table 23: Analysis of Covariance for Homogeneity of Regression

Total Insurance Workers[1]

Predictor	Criterion											
	Quantity of Work			Quality of Work			Accuracy			Knowledge of Job		
	F_1[2]	F_2[3]	F_3[4]	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
Thurstone Test of Mental Alertness												
Verbal	3.32*	3.71	2.90	1.75	3.11	.38	1.11	.83	1.39	2.63	1.65	3.59
Quantitative	3.92*	3.86*	3.92*	2.52	4.65*	.39	.67	.09	1.25	3.44*	3.75	3.09
Total	3.69*	4.26*	3.06	2.29	4.26*	.31	.90	.44	1.36	3.08*	2.79	3.34
Pictorial Reasoning	5.49**	5.40*	5.45*	2.00	3.16	.83	.74	.28	1.20	4.33*	5.43*	3.17
	(2,205)	(1,205)	(1,206)	(2,205)	(1,205)	(1,206)	(2,205)	(1,205)	(1,206)	(2,205)	(1,205)	(1,206)

Table 23 (Contd)

Predictor	Criterion								
	Job Aptitude			Flexibility			Overall Effectiveness		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
Thurstone Test of Mental Alertness									
Verbal	2.70	1.46	3.93*	.62	.67	.58	3.02*	3.32	2.69
Quantitative	2.49	1.56	3.42	.78	.33	1.24	2.51	2.07	2.93
Total	2.69	1.67	3.70	.65	.53	.77	2.91	3.04	2.74
Pictorial Reasoning	2.81	1.91	3.68	1.35	.68	2.02	3.49*	3.47	3.48
	(2.205)	(1,205)	(2,206)	(2,205)	(1,205)	(1,206)	(2,205)	(1,205)	(1,206)

(1) Degrees of freedom for each ratio are shown in parentheses below each column.

(2) F_1 tests hypothesis that $E(Y_{ij} \mid X_{ij}) = a + bX_{ij}$ for all i groups.

(3) F_2 tests hypothesis that $E(Y_{ij} \mid X_{ij}) = a_i + bX_{ij}$ for all i groups.

(4) F_3 tests hypothesis that $E(Y_{ij} \mid X_{ij}) = a_i + b_iX_{ij}$ for all i groups.

* $p < .05$

** $p < .01$

measures can be assumed to be culture-free, then the Pictorial Reasoning Test should demonstrate superior validity, especially in culturally deprived groups.

Two measures were tested as indicators of cultural deprivation. First, analyses were made for the two racial subgroups. A second measure of cultural deprivation suggested by Guion (1966) compares in standard score units, the culture-bound test and the culture-free test. The degree of discrepancy between the two could prove to be a measure of cultural deprivation. This measure could be used either as a moderator, or as a predictor itself. Culturally advantaged and culturally deprived groups were thus identified by the difference between the TMA-T and the Pictorial Reasoning Test. Those with a positive difference score were considered culturally advantaged, while those with a negative difference score were considered culturally deprived. Positive and negative difference score groups were identified within both the white group and the Negro group.

Table 24 presents the validity coefficients for the TMA-T and the Pictorial Reasoning Test for each of the groups. It appears that the TMA was more highly correlated with criterion measures for the culturally advantaged groups. Three out of seven correlations were significant for the white group while none were significant for the Negro group. Five correlations reached significance in the total group with positive difference scores (TMA-T minus Pictorial Reasoning Test) while none were significant for the total group with negative difference scores. The TMA correlated significantly with six criteria in the white positive group while none were significant in the white negative, and with one in the Negro positive while none were significant in the Negro negative group.

The Pictorial Reasoning Test exhibited little validity for any of the population subgroups. Of 77 relationships only four reach significance. Three of those four are found in groups which would be identified as culturally advantaged. Thus, even if a culture-bound test is inappropriate for culturally deprived

Table 24 : Predictor - Criterion Correlations for Groups

Divided According to Cultural Deprivation(1, 2)

Total Insurance Workers

Criterion	Total Group		White Group		Negro Group		Total Gp. Positive Diff.		Total Gp. Negative Diff.	
	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning
Quantity of Work	15*	11	18*	18*	-14	-19	24*	19	05	09
Quality of Work	09	03	16*	09	-20	-20	21*	09	-06	05
Accuracy	01	-01	01	00	-12	-09	12	00	-10	03
Knowledge of Job	03	04	06	13	-20	-22	20*	13	-14	-01
Job Aptitude	06	-03	07	01	-19	-21	15	06	-13	-02
Flexibility	16*	04	16*	06	01	-08	20*	11	04	05
Overall Effectiveness	09	-01	11	05	-23	-28*	21*	10	-13	-04
	(N=208)		(N=157)		(N=51)		(N=99)		(N=109)	

Table 24 : (Contd)

Criterion	White Gp. Positive Diff.		White Gp. Negative Diff.		Negro Gp. Positive Diff.		Negro Gp. Negative Diff.	
	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning
Quantity of Work	33**	28**	06	-21	-30	-45	-17	-06
Quality of Work	29**	13	01	12	-30	-38	-29	-11
Accuracy	17	04	-17	-06	-17	-32	-16	05
Knowledge of Job	31**	21	-23	-05	-47	-55	-24	-11
Job Aptitude	23*	12	-19	-07	-38	-41	-24	-10
Flexibility	25*	15	-01	-05	-10	-21	-01	05
Overall Effectiveness	31**	17	-19	-07	-62*	-59*	-24	-15
	(N=86)		(N=71)		(N=13)		(N=38)	

(1) Decimals are omitted.

(2) Sample size is indicated in parentheses beneath each column.

* $p < .05$

** $p < .01$

individuals, the use of a culture-free test for those individuals did not improve predictability.

Moreover, the relationship between the TMA and the criterion ratings appears to be positive for culturally advantaged groups and negative for culturally deprived groups. Twenty-one out of 28 relationships between the TMA and criterion measures were positive for culturally advantaged groups, while 23 out of 28 were negative for culturally deprived groups. In fact, all of the cases in which the TMA was negatively related to criterion measures in a culturally advantaged group were in the Negro group with positive difference scores.

The same pattern was found for the Pictorial Reasoning Test. In the culturally advantaged groups 21 out of 28 correlations were positive, while 20 out of 28 were negative in the culturally deprived groups. Again, all seven instances of a negative correlation of the TMA with criterion ratings in a culturally advantaged group occurred in the Negro group with positive difference scores.

The pattern of relationships is similar, regardless of whether race or difference score is used to measure cultural deprivation. In fact, the positive difference score group was composed of only 13 Negroes and 86 whites. The negative difference score group consisted of 38 Negroes and 71 whites. Hence the majority of individuals of both races receive the same classification regardless of the cultural deprivation measure.

Guion (1966) further proposes that cultural deprivation itself, as measured by a difference score, may be related to job success. Table 25 presents the correlations between the difference scores and criteria. Since only two out of 63 correlations reach significance ($p < .05$), there was virtually no relationship.

Multiple Regression

The multiple regression coefficients for each of the criteria, using all three TMA scales as well as the Pictorial Reasoning Test and the difference score as predictors, are found in Table 26 along with the cross-validated coefficients.

Table 25: Difference Score - Criterion correlations(1, 2)

Total Insurance Workers

Criterion	Total Group	White Group	Negro Group	Total Gp. Positive Difference	Total Gp. Negative Difference	White Gp. Positive Difference	White Gp. Negative Difference	Negro Gp. Postivie Difference	Negro Gp. Negative Difference
Quantity of Work	05	-01	05	05	-05	02	49**	27	-11
Quality of Work	05	05	01	15	-15	16	-12	12	-16
Accuracy	03	03	-03	15	-16	14	-18	28	-21
Knowledge of Job	-02	-08	02	08	-18	08	-28*	11	-13
Job Aptitude	08	05	02	10	-15	10	-19	03	-13
Flexibility	10	08	08	10	-01	09	06	20	-06
Overall Effectiveness	09	0	05	12	-13	14	-09	-12	-09
	(N-208)	(N=157)	(N=51)	(N=99)	(N=109)	(N=86)	(N=71)	(N=13)	(N=38)

(1) Decimals are omitted.

(2) Sample size is indicated in parentheses beneath each column.

* $p < .05$

** $p < .01$

Table 26: Multiple Regression Coefficients Predictors:

TMA-V, TMA-Q, TMA-T, Pictorial Reasoning,

Difference Score (TMA-T) - Pictorial Reasoning)

Criterion	Multiple R	SE	Cross-validated R	N
Quantity of Work	.23	.73	-.42**	104
Quality of Work	.19	.71	.03	104
Accuracy	.44	.73	.34**	104
Knowledge of Job	.52	.62	-.07	104
Job Aptitude	.48	.75	-.09	104
Flexibility	.41	.86	.29**	104
Overall Effectiveness	.46	.73	.38**	104

** $p < .01$

The original sample used included one half of the total group (N=104). Cross-validation was performed in the other half of the sample. The shrinkage of the multiple R made multiple prediction impractical for three criteria: quality of work; knowledge of job, and job aptitude. Moreover, the instability of the weights made prediction of quantity of work infeasible. The cross-validated coefficients are significant ($p < .01$) in the prediction of accuracy (R=.34), flexibility (R=.29), and overall effectiveness (R=.38).

Study #5: Prediction of Success for Clerical Workers of a Health Insurance Company

Sample

The sample consisted of a more homogeneous subgroup of 126 clerical workers in the health insurance company used in the previous study. Ninety-five workers were white, and 31 were Negro. All employees included in the sample were employed as clerks or clerk-typists.

Predictors

The same predictors employed for the total group of health insurance workers in the previous study were analyzed for this sample. White clerical workers scored significantly higher than Negro workers on all three scales of the TMA ($p < .01$). However, there was no significant difference between the two ethnic groups on the Pictorial Reasoning Test. Predictor means are presented in Table 27.

Criteria

Criterion ratings along the same dimensions employed in the previous study were used as criteria. Although the mean job performance ratings for the white group were higher than for the Negro group on all seven criterion scales, the difference was significant on only one, job aptitude ($p < .05$). Mean criterion scores are presented in Table 28.

Validity

Neither the TMA nor the Pictorial Reasoning Test was a valid predictor for either racial group. Of the 84 predictor-criterion correlations presented in Table 29 only one reached significance ($p < .05$). A larger number of significant correlations could be expected on the basis of chance factors alone. However, a

Table 27: Predictors - Means, Standard Deviations, N's, and Tests of Significance of Mean Differences

Clerical Workers

Predictor	Group	$\bar{X}$	s	N	t(1)
Thurstone Test of Mental Alertness					
Verbal	Total	29.67	9.44	126	
	White	32.02	8.56	95	
	Negro	22.45	8.38	31	5.39**
Quantitative	Total	20.89	7.65	126	
	White	22.00	7.68	95	
	Negro	17.48	6.56	31	2.92**
Total	Total	50.56	15.31	126	
	White	54.02	14.34	95	
	Negro	39.94	13.34	31	4.79**
Pictorial Reasoning Test	Total	46.77	7.78	126	
	White	46.72	8.29	95	
	Negro	46.39	6.63	31	0.22

(1) t ratios are between the means of the white and Negro samples.

** $p < .01$

Table 28: Criteria - Means, Standard Deviations, N's, and Tests of Significance of Mean Differences

Clerical Workers

Criterion	Group	$\bar{X}$	s	N	t(1)
Quantity of Work	Total	3.45	.69	126	
	White	3.51	.66	95	
	Negro	3.26	.73	31	1.77
Quality of Work	Total	3.46	.78	126	
	White	3.51	.77	95	
	Negro	3.32	.79	31	1.18
Accuracy	Total	3.40	.78	126	
	White	3.47	.77	95	
	Negro	3.23	.80	31	1.48
Knowledge of Job	Total	3.56	.72	126	
	White	3.63	.65	95	
	Negro	3.39	.88	31	1.38
Job Aptitude	Total	3.69	.82	126	
	White	3.78	.77	95	
	Negro	3.42	.92	31	2.14*
Flexibility	Total	3.13	.92	126	
	White	3.18	.93	95	
	Negro	2.97	.88	31	1.15
Overall Effectiveness	Total	3.51	.77	126	
	White	3.57	.75	95	
	Negro	3.29	.78	31	1.78

(1) t ratios are between the means of the white and Negro samples.

* $p < .05$

Table 29: Predictor - Criterion Correlations -

Total Group, Whites and Negroes

Clerical Workers(1, 2)

Criterion		Predictor: Thurstone Test of Mental Alertness — Verbal	Quantitative	Total	Pictorial Reasoning	N
Quantity of Work	Total	05	05	06	07	126
	White	05	10	08	16	95
	Negro	-22	-28	-30	-22	31
Quality of Work	Total	04	08	07	07	126
	White	09	17a	14a	15a	95
	Negro	-27	-36* (7)	-35	-31	31
Accuracy	Total	-03	-08	-06	-08	126
	White	-05	-10	-08	-06	95
	Negro	-26	-19	-25	-23	31
Knowledge of Job	Total	-03	-01	-02	05	126
	White	-07	05	-01	15a	95
	Negro	-17	-34	-28	-27	31
Job Aptitude	Total	+07	06	07	-09	126
	White	06	10	09	-06	95
	Negro	-23	-23	-26	-23	31
Flexibility	Total	07	01	05	02	126
	White	07	03	06	05	95
	Negro	-11	-18	-16	-10	31
Overall Effectiveness	Total	07	02	05	-04	126
	White	09	05	08a	03	95
	Negro	-29	-30	-33	-26	31

(1) Decimals are omitted

(2) Number in parentheses below the correlation for the Negro sample indicates the model illustrated.

a indicates significant difference between white and Negro correlations.

* $p < .05$

consistent negative relationship between the scores on both tests and the criteria was found for the Negro sample. All 28 predictor-criterion correlations were negative for the Negro group, while 21 of 28 correlations were positive for the white group. Moreover, the culture-free examination was not more valid than the TMA for either racial group.

Models Illustrated

These data in this study provide an example of Model 7 where there is validity for only one subgroup. The relationship between the ratings of quality of work and the TMA - Quantitative scale presented in Table 30 indicate that the white subgroup scored higher on the test but the test was only valid for the Negro subgroup. Since an increase in the predictor score was not associated with an increase in the criterion rating, the test is not appropriately utilized in this instance. Actually, a better predictor score for the Negro subgroup is indicative of poorer job performance. This illustration meets the additional criterion of a significant difference between validity coefficients.

Table 31 presents the results of the analysis of covariance for homogeneity of regression. The F_1 statistic, which simultaneously tests the equality of regression slopes and intercepts, was significant in 11 out of 28 instances. A significant F_2 statistic ($p < .05$) found in seven instances indicated that the regression slopes were not equal. These seven cases included the five instances in which a significant difference had been found between the white and Negro validity coefficients. The four significant F_3 statistics which were found indicated that a common intercept was inappropriate.

Cultural Deprivation

Following the procedures employed in the previous study, the sample was divided into culturally advantaged and culturally deprived groups on the basis of difference scores (the difference between the TMA-T and the Pictorial

Table 30: Difference Score - Criterion Correlations(1, 2)

Clerical Insurance Workers

Criterion	Total Group	White Group	Negro Group	Total Gp. Positive Difference	Total Gp. Negative Difference	White Gp. Positive Difference	White Gp. Negative Difference	Negro Gp. Negative Difference
Quantity of Work	-01	-09	08	00	03	00	00	02
Quality of Work	-02	-06	03	05	-08	08	-12	-09
Accuracy	02	-02	-04	10	-12	12	-17	-15
Knowledge of Job	-07	-17	05	01	-11	04	-29*	-03
Job Aptitude	13	11	03	12	-08	14	-11	-11
Flexibility	02	-01	-04	01	01	00	07	-08
Overall Effectiveness	07	04	-01	18	-11	20	-14	-14
	(N=126)	(N=95)	(N=31)	(N=45)	(N=81)	(N=40)	(N=55)	(N=26)

(1) Decimals are omitted

(2) Sample size is indicated in parentheses beneath each column.

* $p < .05$

Table 31: Analysis of Covariance for Homogeneity of Regression

Clerical Workers

Predictor	Criterion											
	Quantity of Work			Quality of Work			Accuracy			Knowledge of Job		
	F_1(2)	F_2(3)	F_3(4)	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
Test of Mental Alertness												
Verbal	1.41	1.14	1.67	1.82	2.41	1.21	3.15*	1.68	4.59*	1.96	.36	3.58
Quantitative	2.24	2.62	1.85	3.49*	6.00*	.94	2.38	.57	4.21*	3.86*	4.96*	2.68
Total	1.91	2.20	1.60	2.96	4.92*	.98	3.21*	1.45	4.94*	2.80	2.27	3.29
SRA	3.16*	3.97*	2.30	3.22*	5.17*	1.23	1.89	.86	2.93	3.85*	5.31*	2.31
	(2,122)	(1,122)	(1,123)	(2,122)	(1,122)	(1,123)	(2,122)	(1,122)	(1,123)	(2,122)	(1,122)	(1,123)

Table 31: (Contd)

Predictor	Criterion								
	Job Aptitude			Flexibility			Overall Effectiveness		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
Test of Mental Alertness									
Verbal	2.90	1.89	3.88	.77	.70	.84	3.00	3.70	2.25
Quantitative	3.18*	2.48	3.84	1.23	1.09	1.37	3.10*	3.28	2.87
Total	3.27*	2.77	3.72	1.06	1.07	1.05	3.50*	4.47*	2.47
SRA	2.84	1.29	4.38*	.92	.45	1.39	2.69	2.30	3.06
	(2,122)	(1,122)	(1,123)	(2,122)	(1,122)	(1,123)	(2,122)	(1,122)	(1,123)

(1) Degrees of freedom for each ratio are shown in parentheses below each column.

(2) F_1 tests hypothesis that $E(Y_{ij} | X_{ij}) = a + bX_{ij}$ for all i groups.

(3) F_2 tests hypothesis that $E(Y_{ij} | X_{ij}) = a_i + bX_{ij}$ for all i groups.

(4) F_3 tests hypothesis that $E(Y_{ij} | X_{ij}) = a_i + bX_{ij}$ for all i groups.

* $p < .05$

Reasoning Test in standard score units). Ss with a positive difference score (i.e. those who performed better on the culture-bound exam) were considered culturally advantaged, while Ss with a negative score (i.e. those who performed better on the culture-free exam) were considered culturally deprived. An analysis was performed for positive and negative difference-score groups within the white and Negro groups and for the total white and total Negro groups. Meaningful data could not be reported for the Negro group with a positive difference score because the sample size (N=5) was too small.

Table 32 presents the validity coefficients for the TMA-T and the Pictorial Reasoning Test for each of the subgroups. Neither test exhibits substantial validity: the TMA is significant ($p < .05$) in only three out of 56 cases, and the Pictorial Reasoning Test did not reach significance in a single instance.

However, for the clerical groups, the correlations of both the TMA and the Pictorial Reasoning Test tended to be positive for groups identified as culturally advantaged, and negative for groups identified as culturally deprived. When the TMA was used as a predictor, 19 out of 21 predictor-criterion correlations were positive for culturally advantaged groups, and 23 out of 28 were negative for culturally deprived groups. When the Pictorial Reasoning Test was used as a predictor, 17 out of 21 correlations were positive for culturally advantaged groups, and 23 out of 28 were negative for culturally deprived groups.

The difference score-criterion correlations are presented in Table 32. Again there was virtually no relationship between the predictor and the criterion. One of 56 correlations was significant ($p < .05$) which was less than would be expected by chance.

Thus it appears that the Pictorial Reasoning Test offers no advantage over the TMA when used with culturally deprived individuals regardless whether cultural deprivation is defined on the basis of race or a difference score between a culture-bound and a culture-free test. Furthermore, the utility of a difference score as a predictor appears questionable.

Table 32 : Predictor - Criterion Correlations for Groups Divided According to Cultural Deprivation[1, 2]

Clerical Insurance Workers

Criterion	Total Group		Total White Group		Total Negro Group		Total Gp. Positive Diff.		Total Gp. Negative Diff.	
	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning
Quantity of Work	06	07	08	16	-30	-22	17	14	04	02
Quality of Work	07	07	14	15	-35	-31	23	15	-02	05
Accuracy	-06	-08	-08	-06	-25	-23	01	-08	-15	-05
Knowledge of Job	-02	05	-01	15	-28	-27	20	16	-12	-04
Job Aptitude	07	-09	09	-06	-26	-23	22	09	-10	-04
Flexibility	05	02	06	05	-16	-10	09	07	02	01
Overall Effectiveness	05	-04	08	03	-33	-26	30*	10	-13	-05
	(N=126)		(N=95)		(N=31)		(N=45)		(N=81)	

(1) Decimals are omitted.

(2) Sample size is indicated in parentheses beneath each column.

* $p < .05$

Table 32: (Contd)

Criterion	White Gp. Positive Diff.		White Gp. Negative Diff.		Negro Gp. Negative Diff.	
	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning	TMA	Pictorial Reasoning
Quantity of Work	24	18	03	04	-20	-21
Quality of Work	32*	18	06	14	-38	-26
Accuracy	05	-06	-19	-10	-35	-16
Knowledge of Job	29	20	-20	-02	-29	-24
Job Aptitude	26	09	-12	-06	-32	-17
Flexibility	13	10	01	-03	-14	-04
Overall Effectiveness	38*	13	-17	-09	-36	-18
	(N=40)		(N=55)		(N=26)	

(1) Decimals are omitted.

(2) Sample size is indicated in parentheses beneath each column.

* $p < .05$

Study #6: Prediction of Success in Marine Corps Training

Sample

The sample consisted of 193 Marine Corps recruits undergoing training at Service Support Schools. Eighty-eight of these were Negro and 105 were white. The sample size was smaller for specific analyses due to incomplete predictor or criterion information. Ss were enrolled in training courses for food service personnel, supply school personnel, and motor transport personnel. Personal data for the Ss are summarized in Table 33. Community rank is a measure of the size of the community in which the recruit had spent most of his life, and was rated on an eight point continuum ranging from "in the open country or in a farming community" at the low end to "in a suburb of a very large city (over 500,000 people)" at the high end. The difference between Negro and white samples was significant, with Negroes tending to reside in larger communities. Father's and mother's education was likewise rated on an eight point continuum with the lowest point representing no education or incomplete grade school and the highest point representing completion of graduate or professional school. There were no differences between the racial samples on the amount of parental education. Fathers occupation was ranked on a seven point scale suggested by Hollingshead (1949) ranging from unskilled workers to executives, proprietors of large concerns, or major professionals. Salary was ranked on a five-point scale in $2000 increments from less than $2000 to $10,000 or more. Whites were significantly higher on both father's occupation and salary measures. There were no significant mean differences for either age, education, or the number of weeks the recruits had been attending Service Support Schools. Mean ages were slightly less than 20 years; mean education was slightly greater than completion of the eleventh grade; and the mean number of weeks in school was slightly greater than three.

Table 33: Personal Data

	Group	$\bar{X}$	s.d.	N(1)	t(2)
Community Rank	Total	3.44	2.17	187	
	White	2.89	2.05	102	
	Negro	4.09	2.14	85	3.89**
Father's Education	Total	3.31	1.36	147	
	White	3.37	1.38	84	
	Negro	3.24	1.35	63	.57
Mother's Education	Total	3.59	1.39	160	
	White	3.68	1.38	90	
	Negro	3.47	1.41	70	.94
Father's Occupation	Total	2.80	1.27	170	
	White	3.04	1.33	96	
	Negro	2.49	1.14	74	2.83**
Salary	Total	3.41	1.20	104	
	White	3.75	1.14	57	
	Negro	3.00	1.14	47	3.31**
Age (years)	Total	19.78	1.67	179	
	White	19.77	1.77	98	
	Negro	19.78	1.56	81	.03
Education (years)	Total	11.37	1.61	187	
	White	11.30	1.83	102	
	Negro	11.45	1.30	85	.65
Weeks in School	Total	3.38	1.89	186	
	White	3.40	1.92	101	
	Negro	3.36	1.86	85	.14

(1) Total N is less than 193 because of incomplete data from some subjects.

(2) t ratios are between the means of the white and Negro samples.

** $p < .01$

Predictors

Three tests were used as predictors: the Area Aptitude Test; the Armed Forces Qualification Test (AFQT); and the Fundamental Achievement Series which yielded verbal, numerical, and total scores (FAS-V, FAS-N, and FAS-T, respectively). The Fundamental Achievement Series is a test of elementary verbal and numerical skills, and was regarded as less culture bound than either the Area Aptitude Test or the AFQT. The means, standard deviations, and the test of significance of mean differences are presented in Table 34. The means for the white sample were higher than the Negro sample for all five predictors ($p < .01$).

Criteria

The criterion of final class standing was obtained by converting rank in class to a stanine score. Although class standing of the whites appears slightly higher than that of the Negroes, the difference was not significant. The criterion mean, standard deviation, and test of significance for mean difference are presented in Table 35.

Validity

Predictor-criterion correlations for demographic information, test scores and AFQT-FAS difference scores are presented in Table 36 for the combined groups as well as the Negro and white subgroups. All test scores show significant validity for the white subgroup and no significant validity for the Negro subgroup. The only variables showing significant criterion correlations for the Negro group were age and education.

Models Illustrated

a) The model which occurred most frequently was Model 7. It appeared for six of the fourteen predictors examined, including all of the predictor tests except the Area Aptitude Test, as well as community rank and father's occupation.

Table 34: Predictors - Means, Standard Deviations, N's, and Tests of Significance of Mean Differences

Predictor	Group	$\bar{X}$	s	N(1)	t(2)
Area Aptitude Test	Total	97.49	13.74	192	
	White	100.67	16.08	105	
	Negro	93.67	8.91	87	3.79**
Armed Forces Qualification Test	Total	39.85	19.98	189	
	White	47.70	21.09	102	
	Negro	30.64	13.83	87	6.63**
Fundamental Achievement Series					
Verbal	Total	79.20	8.30	187	
	White	80.67	7.78	102	
	Negro	77.44	8.60	85	2.68**
Numerical	Total	53.34	7.39	187	
	White	55.90	7.13	102	
	Negro	50.27	6.51	85	5.56**
Total	Total	132.58	13.65	187	
	White	136.57	13.41	102	
	Negro	127.29	12.40	85	4.59**

(1) Total N is less than 193 because of incomplete data for some subjects.

(2) t ratios are between the means of the white and Negro samples.

** $p < .01$

Table 35: Criterion - Mean, Standard Deviation,

and Test of Significance for Mean Difference

Criterion		$\bar{X}$	s.d.	N	t(1)
Class Standing (Standing Score)	Total	4.78	1.95	187	
	White	5.08	2.00	102	
	Negro	4.41	1.83	85	.25

(1) t ratio is between the mean of the white and Negro samples.

Table 36: Predictor - Criterion Correlation

Total Group, Whites, and Negroes(1, 2)

Predictor	Group	r	N
Area Aptitude	Total	19**	186
	White	19	102
	Negro	07 (11)	84
Armed Forces Qualification Test	Total	35**	183
	White	47**	99
	Negro	05a (7)	84
Fundamental Achievement Series			
Verbal	Total	23**	181
	White	30**	99
	Negro	09 (7)	82
Numerical	Total	23**	181
	White	23**	99
	Negro	12 (7)	82
Total	Total	26**	181
	White	30**	99
	Negro	11 (7)	82
Community Rank	Total	-19**	181
	White	-20*	99
	Negro	-09 (7)	82
Father's Education	Total	15	144
	White	11	83
	Negro	19	61
Mother's Education	Total	10	156
	White	21*	88
	Negro	-10 (5)	68
Father's Occupation	Total	20**	164
	White	23*	93
	Negro	08 (7)	71
Salary	Total	09	101
	White	04	56
	Negro	-02	45
Age	Total	39**	173
	White	43**	95
	Negro	33**	78
Education	Total	33**	181
	White	40**	99
	Negro	23*	82
Weeks in school	Total	03	180
	White	07	98
	Negro	-04	82
Difference Score (AFQT minus FAS-T standard score unit)	Total	12	177
	White	24*	97
	Negro	-11a (5)	80

(1) Decimals are omitted.

(2) Number in parentheses below the correlation for the Negro sample indicates the Model illustrated.

a Indicates significant difference between white and Negro correlations.

* $p < .05$

** $p < .01$

In these instances, significant mean differences were found for the predictor, but not for the criterion. In each instance, the predictor was valid for the white sample but not for the Negro sample. Since Negroes scored significantly lower on the predictors but not on the criterion (the sole exception being that whites came from smaller communities than the Negroes, and hence had a lower predictor mean in that instance), the employment of these predictors as selection instruments would be unfair to the Negroes. The most striking example of Model 7 is that of the AFQT, where the white correlation was significantly different from the Negro. The widespread use of the AFQT despite its lack of validity for the Negro sample raises serious concerns.

b) Two examples of Model 5 occurred where the level of the mother's education and the difference score were significantly ($p < .05$) related to class standing for the whites, but not for the Negroes. No mean differences between racial groups were observed for either the predictors or the criterion. Lack of validity for the Negro sample makes these variables inappropriate as predictors for the total group.

The analysis of covariance for homogeneity of regression, the results of which are presented in Table 37 shows a significant F_1 statistic in six instances: prediction by the FAS-V, mother's education, salary, age, education, and the difference score. The F_1 statistic simultaneously tested the hypothesis that both the regression shapes and the intercepts were equal for the two groups. The F_2 statistic indicated the appropriateness of separate regression slopes were significant for the AFQT and the difference score. The F_3 statistic indicated the necessity for separate intercepts for the two racial groups when using either mother's education, salary, weeks in school, age, or education as predictors.

Cultural Deprivation

Since Guion (1966) has suggested that the difference, in standard score units, between a culture-bound test and a culture-free test may be an indicator of cultural deprivation, such a measure may be used as a predictor itself, or it may

Table 37: Analysis of Covariance for Homogeneity of Regression(1)

Predictor	F_1(2)	F_2(3)	F_3(4)
Area Aptitude	1.92 (2,181)	3.05 (1,181)	.77 (1,182)
Fundamental Ach. Series Verbal	3.26* (2,177)	2.95 (1,177)	3.53 (1,178)
Numerical	1.09 (2,177)	.52 (1,177)	1.67 (1,178)
Total	1.72 (2,177)	1.60 (1,177)	1.85 (1,178)
AFQT	2.65 (2,179)	5.29* (1,179)	.00 (1,180)
Community Rank	1.84 (2,177)	.79 (1,177)	2.89 (1,178)
Father's Education	1.90 (2,140)	.22 (1,140)	3.60 (1,141)
Mother's Education	4.44* (2,152)	3.72 (1,152)	5.07* (1,153)
Father's Occupation	1.50 (2,160)	.70 (1,160)	2.29 (1,161)
Salary	3.38* (2,197)	.09 (1,197)	6.73* (1,198)
Weeks in School	2.82 (2,176)	.46 (1,176)	5.19* (1,177)
Age	3.16* (2,169)	.26 (1,169)	6.09* (1,170)
Education	3.68* (2,177)	.30 (1,177)	7.09** (1,178)
AFQT minus FAS-T	4.33* (2,172)	5.41* (1,172)	3.18 (1,173)

(1) Degrees of freedom for each ratio are shown in parentheses below the statistic.
(2) F_1 tests hypothesis that $E(Y_{ij} | X_{ij}) = a + bX_{ij}$ for all i groups.
(3) F_2 tests hypothesis that $E(Y_{ij} | X_{ij}) = a_i + bX_{ij}$ for all i groups.
(4) F_3 tests hypothesis that $E(Y_{ij} | X_{ij}) = a_i + b_iX_{ij}$ for all i groups.
* $p < .05$
** $p < .01$

be used as a moderator variable to divide culturally advantaged and culturally deprived groups: Both of these suggestions were employed in this study by using the difference between the AFQT and the FAS-T, in standard score units. Those with a positive difference score were considered culturally advantaged; those with a negative difference score were considered culturally deprived. Positive and negative difference score groups were also identified within the white and Negro groups.

Validity coefficients for each of the predictors as well as for the difference-score are presented in Table 38. The difference-score only correlated with the criterion ($p < .05$) for the white group. Since the magnitude of its correlation was far below that of age, education, and the AFQT, its usefulness as a predictor is questionable.

As a moderator variable, the difference score shows relationships similar to those found for the race variable. Thirteen of the fourteen predictors presented in Table 38 had validity coefficients of a larger absolute magnitude in the white than in the Negro samples. Likewise, 11 of 14 coefficients were larger in the total group with positive difference scores than in the total group with negative difference scores. Although the same trend was observed in the Negro groups (10 of 14 correlations being greater in the positive difference score group), validities for the white group with negative difference scores were not consistently lower than those for the white group with positive difference scores.

Age and education provided the most consistently high correlation with the criterion across cultural groups. Both were significantly correlated ($p < .05$) for all but one of the subgroups. The FAS and the AFQT had a tendency to be correlated with final class standing only for those subsamples where white subjects are included. Both the FAS-T and the AFQT were significantly ($p < .05$) related to the criterion in every subsample except the total Negro sample and the Negro samples divided according to difference score.

Table 38: Correlations with Final Class Standing for Culturally Advantaged and Culturally Disadvantaged Groups

Group	Predictor							
	Area Aptitude	FAS-V	FAS-N	FAS-T	AFQT	Comm. Rank	Father's Education	Mother's Education
Total	.19** (186)	.23** (181)	.23** (181)	.26** (181)	.35** (183)	-.19** (181)	.15 (144)	.10 (156)
White	.19 (102)	.30** (99)	.23** (99)	.30** (99)	.47** (99)	-20* (99)	.11 (83)	.21* (88)
Negro	.07 (84)	.09 (82)	.12 (82)	.11 (82)	.05 (84)	-.09 (82)	.19 (61)	-.10 (68)
Total $\bar{c}$ Positive Difference Score(1)	.31** (84)	.33** (84)	.32** (84)	.36** (84)	.39** (84)	-.24* (84)	.20 (62)	.05 (71)
Total $\bar{c}$ Negative Difference Score	.06 (93)	.20 (94)	.20 (94)	.25* (94)	.31** (94)	-.14 (93)	.07 (78)	.10 (81)
White $\bar{c}$ Positive Difference Score	.29* (53)	.31* (53)	.23 (53)	.31* (53)	.33* (53)	-.16 (53)	.22 (42)	.01 (46)
White $\bar{c}$ Negative Difference Score	.05 (44)	.42** (44)	.38* (44)	.46** (44)	.60** (44)	-.23 (44)	-.02 (39)	.39* (40)
Negro $\bar{c}$ Positive Difference Score	.09 (31)	.16 (31)	.17 (31)	.19 (31)	.24 (31)	-.19 (31)	.17 (20)	.05 (25)
Negro $\bar{c}$ Negative Difference Score	.08 (49)	-.02 (50)	.04 (50)	-.01 (50)	-.11 (50)	-.05 (49)	.16 (39)	-.17 (41)

Table 38: (Contd)

Group	Predictor					
	Father's Occupation	Salary	Age	Educ.	Week in School	Difference Score
Total	.20** (164)	.09 (101)	.39** (173)	.33** (181)	.03 (180)	+.12 (177)
White	.23* (93)	.04 (56)	.43** (95)	.40** (99)	.07 (88)	+.24* (97)
Negro	.08 (71)	-.02 (45)	.33** (78)	.23* (92)	-.04 (82)	-.11 (80)
Total $\bar{c}$ Positive Difference Score	.25* (75)	.20 (48)	.41** (79)	.35** (84)	.07 (83)	.06 (84)
Total $\bar{c}$ Negative Difference Score	.07 (85)	-.09 (51)	.37** (90)	.29** (99)	.07 (93)	.10 (93)
White $\bar{c}$ Positive Difference Score	.30* (49)	.28 (33)	.50** (50)	.38** (53)	.22 (52)	.07 (53)
White $\bar{c}$ Negative Difference Score	.06 (42)	-.24 (22)	.32* (43)	.44** (44)	.08 (44)	.23 (44)
Negro $\bar{c}$ Positive Difference Score	.08 (26)	-.23 (15)	.21 (29)	.39* (31)	-.25 (31)	.03 (31)
Negro $\bar{c}$ Negative Difference Score	.08 (43)	-.01 (29)	.41** (47)	.12 (50)	.06 (49)	-.10 (49)

* $p < .05$
** $p < .01$
(1) $\bar{c}$ = with

Thus, it appears that cultural deprivation, as measured by a difference score (in standard score units) between a culture-bound test and a culture-free test, does not have utility as a predictor. As a moderator, it offers little beyond the more obvious moderator of race. An exception to the above statement, intimated by the data, may be a tendency to achieve greater predictability in white groups moderated on the basis of the difference score.

The difference score was tested further as a moderator according to the method suggested by Saunders (1956). The total sample was divided into back and cross samples (N=85 and N=84 respectively). Cross-validated multiple R's using each of the four most promising predictors (FAS-T, AFQT, Age, Education) individually with the difference score and an interaction term were never higher than the zero-order correlation coefficient between the predictor test itself and the criterion. Thus, it appears that when this procedure is used the utility of a difference score moderator is also limited. The cross-validated multiple R combining the FAS-T, AFQT, Age, and Education was .45, well above the cross-validated statistics using the difference score as a moderator.

The degree of shrinkage when all predictors were combined precluded the utility of a multiple regression equation: All multiple R's and cross-validated statistics are presented in Table 39.

Table 39: Multiple Regression Analyses(1)

Predictors	R	Cross-validated R
All predictors	.72	.26
FAS-T, AFQT, Age, Education	.58	.45
FAS-T, Difference Score, Interaction term	.39	.26
AFQT, Difference Score, Interaction term	.39	.30
Age, Difference Score, Interaction term	.49	.34
Education, Difference Score, Interaction term	.43	.31

(1) Sample sizes were 39 and 38 for back and cross-sample respectively when using all predictors, 85 and 84 for back and cross-samples respectively for all other analyses.

Study #7: Work Sample and Psychomotor Tests as Predictors of the Performance of Sewing Machine Operators

Sample

The subjects in this investigation were 160 female sewing machine operators employed by a large garment manufacturing company. The sample included 67 white and 93 Negro employees. Table 40 presents biographical data on the sample. The white and Negro subgroups did not differ in either age or educational level.

Procedure

A major purpose of the present study was to evaluate a work sample test in terms of the information that the test provided concerning the nature of the job. It was hypothesized that a work sample test would be useful as an aid to applicant self-selection for the job by providing the applicants with a realistic picture of the work environment. Thus, the work sample test provided a preview of the job that could help an applicant determine if the work seemed suited to her interests and abilities.

In order to test the hypothesis that the work sample test would be an effective source of information about the job, applicants were randomly assigned to one of three experimental groups. Group A did not take any tests prior to employment. Group B were administered two perceptual tests (a pinboard and a formboard). The perceptual tests were given in order to ascertain if testing, per se, had any influence upon decisions to take a job. Group C were administered the two perceptual tests and a work sample test, called the Career Determining Exercises. The Career Determining Exercises took about two hours to complete and was composed of items which required the applicant to handle pieces of fabric, to thread the machine, and to actually operate the sewing machine. The scoring of the work sample test was based upon time required to complete the exercises and where appropriate quality of the performance, e.g., the actual sewing required for some parts of the test.

Table 40: Biographical Data - Sewing Machine Operators

	Group	$\bar{X}$	s	N	t
Age	Total	23.69	7.53	160	
	White	24.90	8.80	67	
	Negro	22.68	6.11	93	1.77
Education (in years)	Total	11.10	1.15	160	
	White	11.13	1.04	67	
	Negro	11.09	1.22	93	0.22

In addition to a total score, three subscores for the Career Determining Exercises were obtained. A Sewing Dexterity subscore was computed for the items which were concerned with the applicants' ability to handle pieces of fabric. An Equipment Aptitude subscore was obtained for those items concerned with the applicants ability to thread and maintain the machine. The Machine Control subscore was computed for the items which required the applicant to actually sew certain standardized patterns.

Forty subjects were assigned to Groups A and B and 80 subjects to Group C. All applicants for the job were offered employment, regardless of their scores on the predictors. Since the Ss were assigned to the groups on a random basis, differing proportions of white and Negro applicants were assigned to the three groups. Group A was composed of 16 white and 24 Negro workers and Group C contained 36 white and 44 Negro applicants.

In order to assess the effectiveness of the Career Determining Exercises as a self-selection aid, Groups A, B, and C were compared with respect to voluntary turnover within the first six weeks of employment. It was hypothesized that Group C, which had been administered the Career Determining Exercises, would have a higher percentage of applicants who would refuse employment and, consequently, have a lower voluntary turnover rate than Groups A or B.

The turnover criterion was developed by classifying all employees into one of six categories. The categories were: Remaining on the Job; Refused Employment; Terminated Due to Lack of Progress; Terminated Due to Absenteeism; Voluntary Quit; and Involuntary Quit. The Involuntary Quit category included those workers who had to quit their jobs for such reasons as moving from the area, sickness in the family, etc. Voluntary Turnover was defined as the sum of the Voluntary Quit and Termination Due to Absenteeism categories. The Termination Due to Absenteeism category was included in Voluntary Turnover because the workers in this category had actually withdrawn from the organization. In most of these cases the workers were formally terminated by the company only after it was apparent that the workers did not intend to return to the job.

The validities of the Career Determining Exercises and the two perceptual tests (the pinboard and formboard) were also examined. Criteria included turnover within six weeks of employment and progress in training. Two measures of training progress were used, single cycle time and production rate. Single cycle time measured the amount of time required for the sewing machine operator to complete one unit of the task that she was learning. Production rate referred to the number of units completed per hour. Since the operators were being trained for a number of different sewing operations, both single cycle time and production rate were measured on a dichotomous scale. That is, the operator was scored as being above or below the expected rate for her operation. Each specific operation had an expected single cycle time and production rate which had been previously determined by time and motion analysis.

Progress in training and the turnover criterion were both measured at three intervals: two, four, and six weeks after employment.

Predictor Comparisons

Table 41 presents the means, standard deviations, and tests of the significance of the difference between subgroup means for the predictor variables used with the sewing machine operator sample. The white subgroup scored significantly higher than the Negro subgroup on all predictors except the Pinboard Perceptual Test.

Criterion Comparisons

There were no significant differences between the white and Negro subgroups with respect to either of the training criteria measured at any of the time intervals. With respect to the turnover criteria, the only significant difference between the white and Negro subgroups was the percentage remaining on the job at all three time intervals for Group C.

Table 41: Predictors - Means, Standard Deviations, N's, and Tests of Significance of Mean Differences - Sewing Machine Operators

Predictor	Group	$\bar{X}$	s	N	t
Pinboard (number correct)	Total	77.93	8.47	120	
	White	78.76	8.60	51	
	Negro	77.55	8.23	69	0.77
Formboard (in seconds)	Total	69.33	29.75	120	
	White	57.98	16.15	51	
	Negro	79.69	35.27	69	4.48**
Career Determining Exercises - Total	Total	78.00	15.98	80	
	White	85.90	9.14	36	
	Negro	70.89	17.29	44	4.86**
C.D.E. - Sewing Dexterity	Total	38.97	9.79	80	
	White	42.53	5.25	36	
	Negro	35.70	11.37	44	3.55**
C.D.E. - Equipt. Aptitude	Total	8.23	3.71	80	
	White	9.54	2.48	36	
	Negro	7.01	4.17	44	3.36**
C.D.E. - Machine Control	Total	30.04	7.59	80	
	White	33.36	6.12	36	
	Negro	26.98	8.03	44	3.97**

** $p < .01$

Evaluation of Career Determining Exercises as an Aid To Self-selection

Tables 42, 43, and 44 present the percentage data for the various turnover categories measured at 2, 4, and 6 week intervals, respectively. Tables 45, 46, and 47 present the same data for the combined category of Voluntary Turnover and the category of Remain on Job. The only significant difference between the experimental groups occurred for the white subgroup. The white subgroup in Group C (which had been administered the Career Determining Exercises) had a significantly lower voluntary turnover rate at the six week interval than the white subgroup in Group B. The data generally revealed that the administration of the Career Determining Exercises appears to be effective as an aid to self-selection. There was only a single significant difference between Group C and either Group A or B. However, in all instances, the Group C voluntary turnover rate for the white group was lower than the white employees of Groups A and B. No such consistent pattern was found for the Negro subgroup.

The reasons for the differential effect of the Career Determining Exercises administration upon the voluntary turnover rates of whites and Negroes are not clear. A possible explanation is that the white and Negro subgroups viewed different aspects of the work situation as the important determiners of work attitudes and behaviors. It is hoped that future research concerned with the determinants of work motivation for these workers will help to clarify these data.

Validity

Correlations of the various predictors with the training criteria are presented in Table 48 and with the turnover criteria in Table 49. Generally, low validities were obtained. In three instances training criteria were predictable for the Negro subgroup and in seven cases the turnover criteria were predictable for the white subgroup. The perceptual tests were valid in more instances than the work sample test or subtests, but the number of significant validity coefficients was low for all tests.

Table 42: Two Week Turnover - Percentage Data for Voluntary Turnover and Category of Remain on Job

Group		Turnover Category: Voluntary Turnover	Turnover Category: Remain on Job
	Total	12.5%	60.0%
A	White	12.5%	75.0%
	Negro	12.5%	50.0%
	Total	10.0%	75.0%
B	White	13.3%	80.0%
	Negro	8.0%	72.0%
	Total	6.2%	63.8%
C	White	2.8%	77.8%[a]
	Negro	9.1%	52.3%[a]

[a] Proportions significantly different at the .05 level.

Table 43: Four Week Turnover - Percentage Data for Voluntary Turnover and Category of Remain on Job

Group		Turnover Category: Voluntary Turnover	Turnover Category: Remain on Job
	Total	22.5%	45.0%
A	White	25.0%	62.5%
	Negro	20.9%	33.3%
	Total	20.0%	62.5%
B	White	33.4%	53.3%
	Negro	12.0%	68.0%
	Total	13.7%	56.2%
C	White	8.4%	72.2%[a]
	Negro	18.2%	42.2%[a]

[a] Proportions significantly different at the .05 level.

Table 44: Six Week Turnover - Percentage Data for Voluntary Turnover and Category of Remain on Job

Group		Turnover Category	
		Voluntary Turnover	Remain on Job
A	Total	25.0%	37.5%
	White	31.2%	43.8%
	Negro	20.9%	33.3%
B	Total	25.0%	52.5%
	White	40.0%[a]	46.7%
	Negro	16.0%	56.0%
C	Total	16.3%	52.5%
	White	11.2%[a]	66.7%[b]
	Negro	20.5%	40.9%[b]

[a] Proportions significantly different at the .05 level.

[b] Proportions significantly different at the .05 level.

Table 45: Two Week Turnover - Percentage Data

	Group	Involuntary Quit	Voluntary Quit	Termination - Absenteeism	Termination - Lack of Pregress	Remain on Job	Refused Employment	Total N
	Total	2.5%	10.0%	2.5%	10.0%	60.0%	15.0%	40
A	White	0.0	12.5	0.0	6.2	75.0	6.2	16
	Negro	4.2	8.3	4.2	12.5	50.0	20.8	24
	Total	0.0%	2.5%	7.5%	2.5%	75.0%	12.5%	40
B	White	0.0	0.0	13.3	0.0	80.0	6.7	15
	Negro	0.0	4.0	4.0	4.0	72.0	16.0	25
	Total	3.8%	5.0	1.2%	5.0%	63.8%	21.2%	80
C	White	2.8	0.0	2.8	0.0	77.8[a]	16.7	36
	Negro	4.6	9.1	0.0	9.1	52.3[a]	25.0	44

[a] Proportions significantly different at the .05 level.

Table 46: Four Week Turnover - Percentage Data

	Group	Involuntary Quit	Voluntary Quit	Termination - Absenteeism	Termination -- Lack of Progress	Remain on Job	Refused Employment	Total N
	Total	2.5%	20.0%	2.5%	15.0%	45.0%	15.0%	40
A	White	0.0	25.0	0.0	6.2	62.5	6.2	16
	Negro	4.2	16.7	4.2	20.8	33.3	20.8	24
	Total	2.5%	5.0%	15.0%	2.5%	62.5%	12.5%	40
B	White	6.7	6.7	26.7	0.0	53.3	6.7	15
	Negro	0.0	4.0	8.0	4.0	68.0	16.0	25
	Total	3.8%	6.2%	7.5%	5.0%	56.2%	21.2%	80
C	White	2.8	2.8	5.6	0.0	72.2[a]	16.7	36
	Negro	4.6	9.1	9.1	9.1	43.2[a]	25.0	44

[a] Proportions significantly different at the .05 level.

Table 47: Six Week Turnover - Percentage Data

	Group	Involuntary Quit	Voluntary Quit	Termination - Absenteeism	Termination - Lack of Progress	Remain on Job	Refused Employment	Total N
	Total	7.5%	22.5%	2.5%	15.0%	37.5%	15.0%	40
A	White	12.5	31.2	0.0	6.2	43.8	6.2	16
	Negro	4.2	16.7	4.2	20.8	33.3	20.8	24
	Total	7.5%	7.5%	17.5%	2.5%	52.5%	12.5%	40
B	White	6.7	13.3	26.7	0.0	46.7	6.7	15
	Negro	8.0	4.0	12.0	4.0	56.0	16.0	25
	Total	3.8%	8.8%	7.5%	6.2%	52.5%	21.2%	80
C	White	2.8	5.6	5.6	2.8	66.7[a]	16.7	36
	Negro	4.6	11.4	9.1	9.1	40.9[a]	25.0	44

[a] Proportions significantly different at the .05 level.

Table 48: Correlations of Predictors with Criterion of Turnover - Sewing Machine Operators[a]

		Turnover Category								
		Voluntary Turnover			Termination - Progress			Remain on Job		
Predictor	Group	Two Week	Four Week	Six Week	Two Week	Four Week	Six Week	Two Week	Four Week	Six Week
Pinboard[b]	Total	24*	31**	22*	03	04	00	15	17	11
	White	52**	59**	43**	13	17	05	33*	38*	22
	Negro	09[a] (5)	12[a] (5)	14 (5)	-08	-05	04	01 (6)	02 (6)	06
Formboard[b]	Total	08	01	00	15	02	00	14	02	00
	White	-28	-30*	-35*	11	08	03	-03	-07	-14
	Negro	20	11[a] (7)	06[a])7)	23	-02	-03	24	03	02
C.D.E.[c] Total	Total	11	14	10	06	10	-08	-03	-14	-10
	White	18	18	18	18	13	11	23	18	16
	Negro	-09	-07	06	01	-10	-07	-06	-10	-02
C.D.E.[c] Sewing Dexterity	Total	05	02	07	06	-05	-02	07	05	02
	White	15	16	16	25	27	24	28	30	27
	Negro	10	04	20	-01	-10	-06	07	-05	-06
C.D.E.[c] Equipt. Aptitude	Total	03	07	04	04	-06	-03	05	-01	00
	White	21	22	23	18	26	27	24	31	31
	Negro	-32	-18	-12	11	10	11	-16	-02	00
C.D.E.[c] Machine Control	Total	27*	24	24	07	-06	-05	-12	-15	-15
	White	11	10	10	04	-11	10	08	-06	-07
	Negro	07	15	15	-03	-13	-12	03	-02	01

[a] Correlations are point biserial. Decimals are omitted.

[b] White N = 44; Negro N = 54.

[c] White N = 30; Negro N = 33.

* $p < .05$.

** $p < .01$

a Correlations of subgroups significantly different at .05 level.

Table 49: Correlations of Predictors with Criterion of Training Progress - Sewing Machine Operators[a]

Predictor	Group	Training Progress: Two Week, Single Cycle	Two Week, Prod.	Four Week, Single Cycle	Four Week, Prod.	Six Week, Single Cycle	Six Week, Prod.
Pinboard[b]	Total	-10	00	-08	04	13	17
	White	-18	-03	-16	-08	-11	18
	Negro	-02	04	00	-01	53[a] (5)	19
Formboard[b]	Total	-32**	-17	16	03	-03	-14
	White	-21	-04	06	15	00	-18
	Negro	-37* (7)	-14	02	04	-18	-01
C.D.E.[c] Total	Total	03	21	07	23	09	18
	White	30	28	-11	03	-02	19
	Negro	-23	08	29	30	09	11
C.D.E.[c] Sewing Dexterity	Total	17	12	22	20	02	30
	White	15	31	01	02	-34	31
	Negro	22	-03	48	30	44	34
C.D.E.[c] Equipt. Aptitude	Total	28	31*	12	15	-09	-02
	White	10	26	11	14	06	34
	Negro	43* (7)	30	20	-04	-33	-51
C.D.E.[c] Machine Control	Total	-03	13	06	17	-17	15
	White	-16	22	-21	-07	-22	27
	Negro	12	02	42	30	-08	-03

[a] Correlations are point biserial. Decimals are omitted.

[b] For two week data, White N = 31, Negro N = 38; for four week data, White N = 26, Negro N = 32; for six week data, White N = 20, Negro N = 21.

[c] For two week data, White N = 25, Negro N = 21; for four week data, White N = 22, Negro N = 16; for six week data, White N = 19, Negro N = 10.

* $p < .05$.

** $p < .01$.

a Correlations for white and Negro subgroups significantly different at the .05 level.

Models Illustrated

All of the Bartlett-O'Leary (1969) models found with the sewing machine operator sample were differential validity models. There were no cases in which a predictor was valid for both the white and Negro subgroups. Four instances each of Models 5 and 7 were found. Model 5 illustrates the situation in which no significant differences are found between subgroups on either the predictor or criterion measures and the predictor is valid for only one subgroup. The pinboard perceptual test was valid for predicting voluntary turnover for the white subgroup at all three time intervals, illustrating Model 5. The pinboard test was valid for the Negro subgroup in predicting the single cycle measure of training progress at the six week interval.

Model 7 illustrates a situation in which there is a significant difference between subgroups for only the predictor variable and the predictor is valid for only one subgroup. The relationships between the formboard perceptual test and various criteria constituted three of the examples of Model 7 found with the sewing machine operator sample. The formboard test was valid for the white subgroup when the criterion was voluntary turnover at four or six weeks. The formboard was valid for the Negro subgroup with a criterion of the single cycle measure of training progress at the two week interval. The Equipment Aptitude subscore of the Career Determining Exercises was valid for the Negro subgroup for the criterion of single cycle time measured at the two week interval.

Model 6 was found for two predictor-criterion relationships with the sewing machine operator sample. Model 6 illustrates the situation in which the subgroups differ significantly on only the criterion measure and the predictors valid for only one subgroup. The pinboard test was valid for the white subgroup in the prediction of the turnover category of Remain on Job at the two and four week intervals.

The imposition of the additional criterion for model identification of a significant difference between subgroup correlations reduced the number of models found to three cases of Model 5 and two of Model 7. The Pinboard and formboard

perceptual tests were valid for the white subgroup only in two instances each for the voluntary turnover category in which significant subgroup differences in correlations were found. The only instance in which significant subgroup correlation differences were found and the predictor was valid for the Negro subgroup occurred with the relationship of the Pinboard Test and training progress at six weeks as measured by single cycle time.

Study #8: Learning Measures as Predictors of Task Performance in Two Ethnic Subgroups

Introduction

During the continuing controversy about the utility of psychological tests, some researchers have seriously questioned whether psychological tests can be validly employed to predict minority group performance. In their review of the learning patterns of the disadvantaged Stodolsky and Lesser (1967) indicate that the relationship between tested intelligence and performance on laboratory learning tasks is high for upper socioeconomic groups but negligible for lower socioeconomic groups. Moreover, since upper and lower socioeconomic children demonstrate similar performance on tasks which do not require transfer from previous learning, the learning ability of children from lower socioeconomic backgrounds may not be adequately reflected in general intelligence tests. Several investigators (Kirkpatrick, Ewen, Barrett, and Katzell, 1968; Mitchell, Albright, and McMurray, 1968; Ruda and Albright, 1968) have noted that tests may not have the same degree of validity for minority group members as they do for white subgroups. Also, in the first phase of this contract (O'Leary, Farr, and Bartlett, 1970) the investigators frequently found that tests were valid only for the white sample (e.g., Model 5, Bartlett and O'Leary, 1969).

In view of these data, both educators and employees have suggested that there is a need for the development of new testing techniques or substitutes for tests (Sheppard and Striner, 1968). The present study attempted to combine the techniques of differential psychology and the learning laboratory to obtain valid predictors for minority group members.

Rationale

Jensen (1968) has inferred "basic learning abilities" from an individual's performance on learning tasks which are relatively free from mediational processes or specific transfer from previous learning. If an individual has good basic learning ability and is given the appropriate environmental input, Jensen states that he should be able to acquire the learning sets, mediational habits, verbal associative networks, and the reservoir of transferable skills that largely constitute educability (1968). The learning tasks utilized by Jensen to measure these "basic learning abilities" include selective trial-and-error learning, free recall, serial and paired-associate learning. The experimental conditions were designed to minimize the effects of prior learning. Ferguson's theory (1954, 1956) on learning and human ability postulated a similar point of view. According to Ferguson, "the abilities of man, including the reasoning, number, perceptual, and spatial abilities, and whatever is subsumed under intelligence, are attributes of behavior, which through learning have attained a crude stability or invariance in the adult..." (p.121). That is, abilities are defined in terms of performance on psychological tests and the asymptotic performance, measured by these tests, is considered to be a crude limit of learning. Ferguson conceptualizes abilities as overlearned acquisitions - their stability is the result of overlearning.

Ferguson suggests that what is learned and the age at which it is learned is prescribed by cultural factors. Thus, different cultural environments lead to the development of different patterns of ability. The results of Lesser, Fifer, and Clark (1965) support this point. Theycompared middle-and-lower class children of four ethnic groups: Chinese, Jews, Negroes, and Puerto Ricans. The abilities measured were Verbal, Reasoning, Numerical and Spatial.

While an overall social class difference in mean performance was observed, the most important finding was that the pattern of abilities was different for each ethnic group. This pattern was invariant across social classes.

Rapier (1962) has also supported the view that intelligence tests measure acquired past knowledge or require the use of past knowledge in new situations. The assumption is that an individual's past learning is a valid predictor of what he will be able to learn in the future. McGeoch and Irion (1952) state that the predictive value of past learning rests upon two assumptions: (1) that all of the individuals tested have had an equal opportunity in everyday living or in school to learn the materials the test uses, and (2) although all individuals have had an equal opportunity to learn the materials of the test, there has been some differential learning. As a result, the differences which appear in a test score are a reflection of differences in intellectual ability.

An implication of these views is that differential reinforcement of basic ability patterns could result in various minority groups being at different points on the learning curve. Thus, if the Negro's cultural background reinforced a pattern of abilities which differed from that of the white subgroup, he would not be at the same point on the learning curve as his "equally capable" white counterpart. The frequently observed differences in mean test performance for white and Negro individuals (Anastasi, 1966; Krug, 1966) may be a function of the differential pattern of reinforcement of abilities in the two cultural groups. Moreover, ability tests measure the amount an individual has learned under almost infinite variations in the conditions of learning. Current psychological tests may be poor predictors of performance for minority group members because these variations in the conditions of learning have not been taken into account.

It is possible that an Ss performance in an actual learning situation does not depend primarily, as it does with traditional ability tests, upon what he has learned before he comes to the learning situation. The individual is simply given something to learn and the amount he learns is used as a measure of his ability. Standardization of the learning situation makes the conditions of learning more nearly equal for both majority and minority group members than the learning required for aptitude test performance. Moreover, research studies of minority group performance on simple learning tasks suggest that mean differences in performance for white and Negro individuals in standardized learning situations would be minimal as compared to mean differences in ability test performance for the two ethnic groups.

Some researchers have employed measures of performance in learning situations as predictors. Frederiksen, Carstater, and Stait (1947) suggested that scores derived from miniature learning situations might be used to estimate an individual's performance in future learning situations, i.e., final course grades might be predictable from how well the student learned similar concepts and principles taught in an hour's period of time before the course began. Allison's (1954, 1956) studies at U.S. Naval Training Schools demonstrated that measures of learning were related to measures of success in schools teaching mechanical-motor skills. Also, Allison (1956) found better predictions of success from a combination of the learning measures and the Navy Basic Test Battery than from the Basic Test Battery alone.

Wardrop (1967) utilized a programmed instruction unit as a complex miniature learning situation to predict classroom success. In the first of two studies, Wardrop compared the predictive validity of traditional psychological tests with performance on an 85 minute programmed text on "How to Study" using training success in Naval Electronics and Mechanical Schools as a criterion.

In addition, two perceptual and psychomotor learning tasks were employed; the DuBois-Bunch Learning test and a Numbers Test (Hackett, 1964). Pre-tests and post-tests were used to obtain a residual gain measure of learning on both the predictor and criterion tasks. Results of this study indicated that the programmed learning residual gain measure was approximately equal to the traditional psychological test (Navy General Classification Test) in the prediction of classroom performance; r = .27 vs. r = .28 for the electronics school and r = .23 vs. r = .30 for the mechanical school. Although the programmed instruction learning measure and the abilities measure showed some overlap, the independent component of the programmed learning measure when combined with the ability measure resulted in a somewhat higher validity coefficient (Wardrop, 1967).

The subjects in Wardrop's second study were students in an introductory psychology class. Two programmed learning tasks were utilized; a study skills program and a binary numbers program. Tests of verbal ability and numerical ability were the traditional psychological tests employed. Neither the traditional tests nor the residual gain measures correlated significantly with final course grades. The only significant correlations with the criteria were the post-test scores on the binary numbers program and a combined post-test score (binary numbers post-test score plus study skills program post-test score). Wardrop concluded that the results of these studies, while certainly not definitive, indicated that "learning tests" are as valid as traditional measures of intelligence in the prediction of classroom performance.

In the development of learning measures as predictors, one important factor which has not been systematically explored is the analysis of the criterion task. The situation is analogous to the development of an ability test battery to predict job performance without first conducting some type of job analysis. One reason that learning tasks have not consistently predicted criterion performance may be that researchers have primarily employed perceptual or psycho-

motor skills as predictors (Wardrop, 1967). It is likely that the more the predictive situation is similar to the criterionsituation, the higher the resultant validity coefficients. In other words, the more similar the skills required in the predictor situation are to the skills required in the criterion, the more valid the predictor. If an individual's learning is highly variable across different types of materials and across different learning situations (Jenkins, 1967), then systematic analysis of the criterion is especially important in the development of learning tasks as predictors.

An analysis of the criterion task in the present study employed Gagne's (1965) eight categories of learning which have been defined in terms of the different sets of conditions necessary for learning. The learning task was selected to represent the types of learning utilized in the criterion task.

In summary, current psychological tests may be poor predictors of minority group performance because the variations in the conditions of learning have not been taken into account. The present study employed a measure of the subject's performance on a miniature learning task derived directly from the criterion task as a predictor of his criterion task performance. Moreover, the conditions of learning are more nearly equalized for the two ethnic groups than the previous learning required for performance on aptitude tests in an attempt to control the bias present in traditional psychological tests.

Subjects

Ninety-four freshmen and sophomore students (46 white and 48 Negro) at the University of Maryland participated in the five hour experiment. All subjects were obtained through ads in the campus newspaper and were paid ten dollars for their participation. There were two experimental sessions, a two hour and a three hour session, run on consecutive evenings.

The Criterion Task

The greatest problem encountered by researchers validating tests for Negro and white groups is obtaining racially mixed groups. Frequently, integrated classrooms contain only a few Negroes. Combining classrooms or sections to obtain an adequate sample of minority group members only complicates the problem since there is no way to assess the equality of the various instructor's grading and instructional systems. In an attempt to overcome these difficulties, the present study utilized a programmed instruction unit as a criterion task. Since the programmed instruction unit attempts to equate the instruction for all subjects, students from different classrooms can be combined to obtain an adequate sample of minority group members. Moreover, research results indicate that learning accomplished through programmed instruction is at least as effective as "conventional" methods of instruction (Stolurow, 1961).

Chapter four of J. T. Gibson's (1968) programmed instruction text Industrial Psychology was selected as the criterion task. This self-contained programmed instruction unit covers the following topics: averages, the normal curve, percentiles, interquartile and semi-interquartile ranges, and the standard deviation. Sample frames are presented in Table 50

The 128 frames of this unit were divided into three ($3\frac{1}{2}$" by $8\frac{1}{2}$") booklets of approximately equal length. Each frame was presented on a separate page. The subject responded by writing his answer on a separate answer sheet. The subject's response consisted of either: (1) constructing an answer, (2) filling in a missing word, or (3) selecting the correct choice from several alternatives. Immediately following each question frame was another frame containing the correct answer. No time limit was imposed on the subject.

Three measures of criterion performance were obtained. The first was the number of errors in each of the three instruction booklets. Also, a final

Table 50: Sample Frames - Criterion Task

AVERAGES

1 Psychologists often study behavior in order to generalize about features that can be predicted or are typical of groups of people. Sometimes a group of subjects is measured for some characteristic of behavior, such as time for running the 100-yard dash. The performance of the group as a whole can be described by an average score.

Averages are used to (*choose one*):

a. describe individual behavior.
b. make generalizations of what can be expected of groups of individuals.

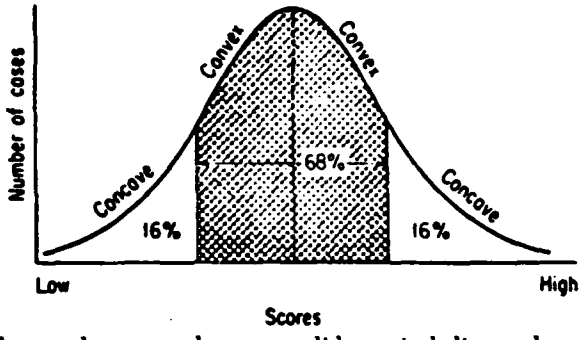
2 What does *average* mean? What is the *average* individual like? The term *average* has, in fact, a variety of meanings. It has been used to describe (*choose one*):

a. the typical individual.
b. the individual who appears most often in the group.
c. an individual whose measured characteristics cause him to fall descriptively in the middle of the group.
d. all of these.

3 Three measures are commonly used to determine an average: the arithmetic mean, the midscore or median, and the mode. Each measure means something *different* and is determined by different statistical methods. Thus it follows that two separate measures:

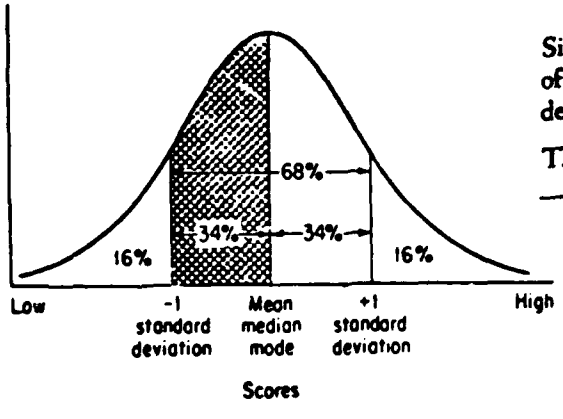
a. often give different estimates of average.
b. always give identical estimates of average.

60 Another property identifying the normal distribution is that a particular known percentage of the scores falls under each part of the curve. If a perpendicular line were drawn from the baseline to the two points where the curve changes slope from convex to concave, then 68% of the total area under the curve would be cut off as follows:



The shaded area between the two solid vertical lines, drawn from the points where the curve changes from convex to concave, represents _____ % of the cases in the distribution.

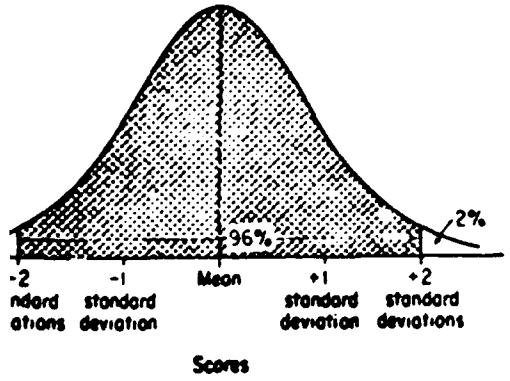
61 The distance along the baseline from the center of the distribution to one or the other of the solid lines perpendicular to the base is called a ***standard deviation.***



Since the mean, median, and mode cut the distribution in half, 34% (half of 68%) of the scores fall between the mean and a point one standard deviation above the mean (plus one standard deviation).

Thirty-four per cent of the scores fall between the mean and a point one _____ below the mean (minus one standard deviation).

2 If we double the distance along the baseline, represented by one standard deviation, we then include two standard deviations.



In this figure, ninety-six per cent of the scores fall between a point + 2 standard deviations and a point _____ from the mean

level of achievement was measured by a 30 item multiple-choice test (see Appendix A). All questions were designed to measure applications of concepts and principles to new situations and did not duplicate frames in the programmed instruction text. The multiple choice test was also administered prior to the programmed instruction task to obtain the third measure of criterion performance - a measure of gain in proficiency. Manning and DuBois (1958) have suggested that where: (1) trainees differ in their initial level of performance, that is, their prior education and experience has led to diversity in pre-training proficiency, and (2) the training curriculum does not ordinarily result in mastery of the job, but rather develops skills fundamental to efficient learning on-the-job, final grades may not adequately reflect the performance of the trainees. In this situation, improvement or gain-in-proficiency is considered to be the most significant dimension of student performance.

The most widely used measure of gain-in-proficiency is simply the difference between final and initial performance. This measure of change has been labeled "crude gain". However, many researchers (Manning and DuBois, 1962; Carver and DuBois, 1967; Wardrop, 1967; Cronbach and Furby, 1970) have indicated that "crude gain" scores are noted for their unreliability. In addition, crude gain scores and initial performance scores are usually negatively correlated, resulting in an over-prediction of learning for subjects with low initial scores and an under-prediction of learning for subjects with high initial scores. In an attempt to compensate for the deficiencies of the "crude gain" measure, Manning and DuBois (1962) have recommended the use of the "residual gain" measure. If Z_2 is the final score and Z_1 the initial measure (both expressed in Z score units), Z_2 can be divided into two uncorrelated parts: $\hat{Z}_2$ which correlates perfectly with Z_1, and $Z_{2.1}$ which is uncorrelated with Z_1.

Thus,

$$\text{residual gain} = Z_2 - \hat{Z}_2 = Z_2 - r_{12}Z_1 = Z_{2.1}$$

Residual gain is the deviation of final performance scores from the regression line of final or initial performance scores.

Manning and DuBois (1962) have suggested that the use of residual gain scores as a criterion for the validation of selection tests would serve to facilitate selection procedures oriented toward criteria of trainability or educability, rather than achievement at a particular point in learning. Since there may be a difference in the level of performance (both pre and post) for majority and minority group members, this type of criteria measure seems relevant for the validation of tests for minority group members. These members may be equally "trainable", as compared to their white counterparts, when measured by amount of gain in proficiency rather than final level of achievement.

Predictors

The experimental learning tasks utilized as predictors were also selected from an analysis of the criterion task using Gagne's (1965) eight categories of learning. This analysis revealed that the most frequent types of learning represented in the criterion task were concept and principle learning.

Three different types of learning tasks were used: a paired-associate task, a concept learning task, and a principle learning task. All the learning tasks were presented in a programmed instruction format similar to the criterion task. Since previous researchers indicated that paired-associate learning predicted academic success, the paired-associate task was included even though the analysis of criterion task indicated that this type of learning was not frequently represented in the criterion task. The paired-associate task was a modification of the task used by Duncanson (1966) in his study of learning and measured abilities. The task consisted of ten paired-associate items.

The ten response terms were selected from Noble's (1961) list of high M value nouns. The stimulus terms were nonsense syllables having a 47% association value. The actual pairs were: YAT-JEWEL, TIS-DINNER, ZUG-MONEY, REM-VILLAGE, SOZ-INSECT, QUN-GARMENT, XOW-HEAVEN, NOL-WAGON, RUH-OFFICE, BEK-KITCHEN. Each stimulus term was typed on a separate frame. The subject then responded by writing his answer on a separate answer sheet. Immediately following each stimulus frame was a frame which contained both the stimulus and response terms. The 10 pairs of words were randomized within each of the eight trials.

The concept task was a modification of a concept learning task developed by Allison (1960). The task consisted of 16 sets of four words with each set being assigned to one of the code letters A, B, C, or D. Four sets of words belong to each letter, with the code letter representing an underlying concept. The following concepts were employed: concept A was that one of the four words was a number; concept B was that one of the four words was a sport; concept C was that two of the four words were homonyms; concept D was that all four words were units of measurement.

Again a programmed instruction format was utilized in presenting the task. Each set of words was typed on a separate frame. The subject responded by writing his answer on a separate answer sheet. Immediately following each set of words, a frame was presented which contained the set and the correct classification letter. No time limit was imposed on the subject. The order of presentation of the sets of words was randomized as well as the spatial order of the words within each set. Each of the 16 sets of words was presented 8 times. Table 51 presents four of the sets of words used.

The principle-learning task was adopted from Schoer's programmed text, An Introduction to Statistics and Measurement. (1966) Forty-two frames were selected on the topic of nominal and ordinal measurement. Sample frames are presented

Table 51: Sample Frames - Concept Task

SLOP SEVEN LIFT MASK A	TRACK COOK TRAIL HOCKEY B
MIGHT NIGHT PURPLE MITE C	INCH MILE YARD FOOT D

in Table 52. The programmed instruction format of this task was the same as the one used in the criterion task. A 20 item test in which the student was required to identify whether a measure was nominal or interval was developed to measure final level of achievement as well as gain in proficiency (see Appendix B).

The traditional psychological tests used as predictors were the Wonderlic Personnel Test, and French's (1963) Wide Range Vocabulary (V-3) and Addition Test (N-1). Also included as a predictor was a Digit Span Test developed by the Navy Personnel Research Activity, San Diego, which yielded three scores; total number of digits correct, length of span where first error occurred, and longest correct span. Both the instructions and the problems of the Digit Span Test were administered by tape recorder. The test involved immediate memory - writing digits after a single hearing. The spans of digits ranged from 4 to 10 digits. Jensen (1968) has indicated that this type of test was more effective than conventional intelligence tests in predicting school grades for individuals from low socioeconomic backgrounds.

Predictor Performance

Table 53 presents the means, standard deviations, and tests of significance of mean differences for the two racial groups on the predictor variables. On the experimental learning measures, there were no differences in the mean performance of the two ethnic groups on the initial trial or pre-test measures. Although there were no differences on the final trial scores of the paired-associate task, white students obtained higher scores on the final level of achievement measures on both the concept and principle learning tasks.

The mean residualized gain scores for the two racial groups on the experimental learning measures are also presented in Table 53. There was no significant difference in the amount of gain on the paired-associate task for the two

Table 52: Sample Frame - Principle Learning Task

31 Equivalence within a category is involved in both _____ and _____ measurement.	34 The numerals rank-order the _____. They indicate
32 The concept of one category being "higher than," "greater than," or "more than" another category is involved in _____ but not _____measurement.	35 Assigning the grades A, B, C, D, and F to schoolwork is essentially ordinal measurement with the most preferred grade being ____ and the least preferred being ____.
33 The numerals used to indicate the categories in ordinal measurement must, then, not only name the categories but also reflect the _____ in which these categories lie on the scale.	36 If the numerals 1 through 5 were used instead of letters and the numeral 1 were assigned the A, the numeral 5 would have to be assigned the ____.

Table 53: Predictors Means, Standard Deviations, N's and Tests of Significance of Mean Differences - Learning Measure Study

Predictor	Group	$\bar{X}$	s	N	(1)
Paired - Associate Task					
First Trial	Total	.17	.77	94	
	White	.02	.15	46	
	Negro	.31	1.06	48	1.86
Final Trial	Total	7.38	2.83	94	
	White	6.83	2.89	46	
	Negro	7.92	2.69	48	1.87
Concept Task					
First Trial	Total	8.56	3.06	94	
	White	9.09	2.94	46	
	Negro	8.06	3.12	48	1.63
Final Trial	Total	13.73	3.09	94	
	White	14.70	1.55	46	
	Negro	12.81	3.85	48	3.11**
Principle Learning Task					
Pre-test	Total	20.70	5.11	94	
	White	21.70	5.05	46	
	Negro	19.75	5.04	48	1.85
Post-test	Total	28.35	6.94	94	
	White	30.98	5.93	46	
	Negro	25.83	6.94	48	3.82**

(1) t ratios are between the means of the white and Negro samples

* $p < .05$
** $p < .01$

Table 53 (Contd)

Predictor	Group	$\bar{X}$	s	N	t
Residual Gain Scores					
Paired-associate task	Total	.00	1.00	94	
	White	-.19	1.02	46	
	Negro	.18	.95	48	1.80
Concept task	Total	.00	.98	94	
	White	.25	.50	46	
	Negro	-.24	1.18	48	2.67*
Principle task	Total	.00	.96	94	
	White	.32	.86	46	
	Negro	-.31	.95	48	3.33**
Residual Gain Scores (using subgroup correlations)					
Paired-associate test	Total	-.03	1.00	94	
	White	-.25	1.00	46	
	Negro	.18	.95	48	2.11*
Concept task	Total	.00	.94	94	
	White	.25	.50	46	
	Negro	-.25	1.18	48	2.67*
Principle task	Total	.02	1.00	94	
	White	.35	.85	46	
	Negro	-.31	.95	48	3.51**
Wonderlic	Total	24.90	5.76	94	
	White	28.72	4.04	46	
	Negro	21.25	4.71	48	8.15**
Vocabulary Test	Total	23.44	8.45	94	
	White	27.87	7.09	46	
	Negro	19.19	7.54	48	5.68**
Addition Test	Total	36.78	11.20	94	
	White	37.48	10.64	46	
	Negro	36.10	11.76	48	.59
Digit Span Test					
Total Digits Correct	Total	122.06	13.24	94	
	White	124.52	9.80	46	
	Negro	119.71	15.61	48	1.78
Length of Span of First Error	Total	6.86	1.27	94	
	White	6.65	1.27	46	
	Negro	7.06	1.24	48	1.57
Length of Longest Correct Span	Total	8.34	1.02	94	
	White	8.52	.69	46	
	Negro	8.17	1.24	48	1.68

ethnic groups. These data agree with the lack of significance found between the mean performance of the two groups on both the initial or final trial measures. For both the concept and principle learning tasks, white students obtained higher gain scores than Negro students. Since there were no differences in the pre-test measures, these data are also reflected by the higher scores of the white subgroup on the final achievement measures.

Since the results of the residual gain analysis were dependent on the correlation between the pre- and post-test measures, the use of subgroup correlations in the formats for computing residual gain scores might produce different results. Table 53 presents the results of the residual gain analysis using subgroup correlations. The results of this analysis were similar to the results obtained using total group correlations between the pre- and post-test measures. The only exception occurred with the paired-associate task. The residualized gain scores computed using subgroup correlations were significantly different for the two ethnic groups with the Negro group showing more gain in proficiency.

White students scored significantly higher than Negro students on both the Wonderlic Personnel Test and the Wide-Range Vocabulary Test. There was no difference between the mean performance of the two groups on the Addition Test or any of the scores on the Digit Span Test.

Criterion Performance

Mean criterion scores for the two ethnic groups are presented in Table 54. Negro students had more errors on all three booklets of the programmed instruction criterion.

White students obtained higher scores on the multiple-choice criterion test administered both before and after the students were exposed to the programmed instruction. Thus, in terms of final level of achievement white students scored higher than Negro students.

Table 54: Criteria - Means, Standard Deviations, N's, and Tests of Significance of Mean Differences - Learning Measure Study

Criterion	Group	$\bar{X}$	s	N	$t^{(1)}$
Error Scores					
Booklet I	Total	3.36	3.10	94	
	White	2.02	1.86	46	
	Negro	4.65	3.50	48	4.53**
Booklet II	Total	11.57	8.36	94	
	White	9.84	6.02	46	
	Negro	13.23	9.90	48	2.12*
Booklet III	Total	13.24	8.52	94	
	White	9.83	5.20	46	
	Negro	16.52	9.76	48	4.12**
Pre-test	Total	11.52	3.78	94	
	White	13.63	2.98	46	
	Negro	9.50	3.34	48	6.25**
Post-test	Total	17.10	5.45	94	
	White	20.07	5.13	46	
	Negro	14.25	3.98	48	6.09**
Residual Gain - Criterion Task	Total	.00	.81	94	
	White	.21	.90	46	
	Negro	-.21	.65	48	2.57*
Residual Gain - Criterion Task (using subgroup correlations)	Total	.04	.83	94	
	White	.35	.89	46	
	Negro	-.25	.64	48	3.72**

(1) t ratios are between the means of the white and Negro samples

* $p < .05$
** $p < .01$

Residual gain scores were computed for the two groups using the pre- and post-measure. Inspection of Table 54 indicates that white students not only had higher pre- and post-test scores but they also exhibited more gain in proficiency as measured by the residual gain scores. A similar finding was obtained using the residual gain scores developed using subgroup correlations.

Validity

Correlations between the predictor and criterion measures are presented in Table 55. Performance on the paired-associate task did not predict performance on any of the criterion measures for either racial group separately, and correlated significantly with only two measures for the total group post-test and residual gain scores using subgroup correlations.

Initial level of performance on the concept task correlated with error scores on all three booklets for both the Negro group and the total group but did not correlate significantly for the white group. Neither final level of achievement nor residualized gain was predictable using initial level of performance as the predictor.

Final level of performance on the concept task did not correlate with final level of achievement on the criterion test for either ethnic group separately, but did correlate with final level of achievement for the total group. Although final level of achievement on the concept task predicted error scores on Booklet I for both whites and Negroes, it predicted errors on Booklets II and III only for the total group. Residualized gain scores were not predictable using final level of achievement on the concept task.

Initial level of performance on the principle learning task did not predict any of the criterion measures for the Negro group and predicted only two criteria for the total group i.e., final level of achievement and residualized gain scores using subgroup correlations. Final level of achievement on the

Table 55 : Predictor - Criterion Correlations

Learning Measure Study(1, 2)

Criterion	Group	Predictor: Paired Associate Task		Predictor: Concept Task	
		First Trial	Final Trial	First Trial	Final Trial
Error Scores					
Booklet I	Total	03	-11	-32	-44**
	White	-08	-21	-24	-29*
	Negro	-06	-25	-31* (6)	-37** (1)
Booklet II	Total	04	-13	-27**	-30**
	White	03	-12	-14	-21
	Negro	-01	-22	-31* (6)	-26
Booklet III	Total	03	-08	-26**	-36**
	White	01	-24	-01	-23
	Negro	-06	-16	-32* (6)	-28 (11)
Post Test	Total	-29**	-08	09	27**
	White	-06	05	01	15
	Negro	-26	-01	-01	15
Residual Gain - Criterian Task	Total	-20	-15	02	16
	White	-05	-11	-05	16
	Negro	-26	-11	-01	07
Residual Gain - Criterian Task (using subgroup correlations)	Total	-22*	-13	04	20
	White	-05	-05	-03	16
	Negro	-27	-09	-01	09

(1) Decimals are omitted
(2) Number in parentheses below the correlation for the Negro sample indicates the model illustrated.

* $p < .05$
** $p < .01$

Table 55 (contd)

Criterion		Principle Task		Residual Gain Scores		
Error Scores		Pretest	Posttest	Paired-Associate	Concept	Principle
Booklet I	Total	-04	-20	-12	-35**	-19
	White	-05	-15	-21	-13	-13
	Negro	10	-01	-24	-31* (8)	-03
Booklet II	Total	-17	-19	-13	-22*	-15
	White	-19	-07	-12	-12	-01
	Negro	-11	-15	-22	-19	-13
Booklet III	Total	-08	-19	-09	-29**	-17
	White	-06	-09	-24	-23	-07
	Negro	01	-04	-16	-20	-04
Post Test	Total	35**	51**	-07	25**	43**
	White	38**	32**	05	15	20
	Negro	22 (6)	49** (1)	01	16 (11)	45** (8)
Residual Gain - Criterion Task	Total	18	32**	-15	16	28**
	White	22	25	-11	19	18
	Negro	03	26 (11)	-09	07	27 (11)
Residual Gain - Criterion Task (using subgroup conditions)	Total	24**	39**	-12	19	33**
	White	30*	29*	-05	18	20
	Negro	06 (6)	31* (1)	-07	09	31* (8)

Criterion		Residual Gain Scores (using subgroup correlations)		
		Paired-associate	Concept	Principle
Booklet I	Total	-10	-35**	-20
	White	-22	-13	-14
	Negro	-24	-31* (8)	-04
Booklet II	Total	-12	-22*	-16
	White	-12	-12	-04
	Negro	-22	-19	-12
Booklet III	Total	-07	-29**	-18
	White	-24	-23	-08
	Negro	-16	-21 (11)	-04

Table 55 (contd)

		Paired-associate	Concept	Principle
Post Test - Criterian task	Total	-09	25**	46**
	White	05	15	27
	Negro	01	16	45**
			(11)	(8)
Residual Gain - Criterion Task	Total	-15	16	30**
	White	-12	19	22
	Negro	-09	08	27
				(11)
Residual Gain - Criterian task (using subgroup correlations)	Total	-13	19	36**
	White	-05	18	25
	Negro	-07	09	31*
				(8)

Criterion

		Wonderlic	Vocabulary	Addition
Booklet I	Total	-56**	-32**	-23*
	White	-35*	-08	-34*
	Negro	-44**	-16	-18
		(1)	(11)	(6)
Booklet II	Total	-38**	-12	-29**
	White	-44**	-11	-31*
	Negro	-29*	03	-27
		(1)		(6)
Booklet III	Total	-56**	-27**	-27**
	White	-66**	-31*	-37*
	Negro	-40**	02	-23
		(1)	(8)	(6)
Post Test	Total	63**	54**	31**
	White	43**	37**	33**
	Negro	48**	36*	35*
		(1)	(1)	(3)
Residual Gain - Criterion Task	Total	36**	36**	24**
	White	37**	36**	21
	Negro	14	17	27
		(8)	(8)	(11)
Residual Gain - Criterion Task (using subgroup correlations)	Total	45**	42**	28**
	White	41**	38**	27
	Negro	20	21	30*
		(8)	(8)	(6)

Table 55: (contd)

Criterion		Digit Span Test		
Error Scores		Total Digits Correct	Span of First Error	Longest Correct Span
Booklet I	Total	-43**	-12	-41**
	White	-14	-07	-16
	Negro	-48**	-31*	-44**
		(6)	(6)	(6)
Booklet II	Total	-27**	-11	-31**
	White	-13	-13	-14
	Negro	-26	-17	-33*
				(6)
Booklet III	Total	-34**	-10	-33**
	White	-16	-23	-04
	Negro	-34*	-16	-36*
		(6)		(6)
Post Test	Total	16	-05	19
	White	12	07	16
	Negro	04	01	11
Residual Gain - Criterion Task	Total	12	01	14
	White	21	12	22
	Negro	-02	-03	03
Residual Gain - Criterion Task (using subgroup correlations)	Total	13	-02	16
	White	16	10	20
	Negro	-01	-03	05

principle learning task predicted post-test scores and residualized gain scores using subgroup correlations for both racial groups but did not predict error scores for either racial group. Residualized gain scores using total group correlations were predictable only for the total sample.

Residualized gain scores on the paired-associate task did not predict any of the criteria for either racial group, while these scores on the concept task predicted error rates on all three booklets, as well as post-test scores for the total group, However, none of the criteria were predictable for each racial group separately with the exception that error scores on Booklet I were predictable for the Negro sample. Similar findings were obtained with residualized gain scores based on subgroup correlations.

Residualized gain scores on the principle learning task did not predict error scores on the criterion for any group. Post-test scores and residualized gain scores on the criterion task were predictable for the total group. However, inspection of subgroup correlations indicates that only the post-test scores and the residualized gain scores using subgroup correlations were predictable only for the Negro sample.

The Wonderlic Personnel Test predicted error scores on all these booklets and final level of achievement for both racial groups as well as the total group while residual gain scores were predictable only for the total group and the white subgroup. The Wide-Range Vocabulary Test predicted final level of achievement for both racial groups but predicted residual gain only for the white and total sample. Error scores were not predictable using the vocabulary test except for Booklet III for the white subgroup.

The addition test predicted error scores on all three booklets for the white sample and total group but did not predict error scores for the Negro sample. Although final level of achievement was predictable for both racial

groups, residual gain scores on the criterion test were predictable only for the total group. Residualized gain scores using subgroup correlations were predictable only for the total group and the Negro sample.

Neither the post-test scores nor the residualized gain scores were predictable using the Digit Span Test. However, the total digits correct score and the longest correct span score predicted error scores on all three booklets for the total group and predicted errors on Booklets I and III for the Negro sample. The length of span of first error score correlated with only one criterion-error scores on Booklet I for the Negro sample.

To summarize, perhaps the most consistent finding was that performance on the paired-associate task did not predict any of the criterion measures. The concept task exhibited some validity in predicting errors for the total group and in a few instances for the Negro sample. On the other hand, the principle learning task showed its greatest validity in the prediction of final level of achievement. In general, the validity of the residual gain measures was lower than that of the final level of performance measures but the general pattern of validity for the two types of measures was similar.

The Wonderlic predicted more criteria than any other predictor. In fact, the Wonderlic predicted all criterion measures for the white sample and all but the gain measures for the Negro sample. The Vocabulary and Addition Tests predicted final test performance for both racial groups but neither predicted errors for the Negro sample. The Digit Span test did not predict post-test performance or gain in proficiency for either racial group, but two of the scores exhibited validity in predicting errors scores for the Negro sample.

Models Illustrated

Five separate models were illustrated in this study. The specific model illustrated in each predictor-criterion relationship is enclosed in parenthesis below the correlation for the Negro sample in Table 55. The most frequently

illustrated model was Model 6, represented in fifteen cases. The relationship between all scores of the Digit Span Test and error scores on Booklet I are clear illustrations of Model 6. White students made less errors than Negro students but there was no difference in the mean performance of the two groups on the Digit Span Test. Moreover, the test was valid only for the Negro sample. Although this situation is not likely to lead to differential selection rates, it is not an optimal selection strategy since the best individuals in the white sample are not being selected.

Model 8, most clearly illustrated in the relationship between the residual gain scores on the principle learning task and post-test scores on the criterion test, was illustrated in eleven predictor-criterion relationships. Model 8 is illustrative of the situation where a difference in predictor performance is paralleled by a difference in criterion performance, but the test is valid only for one subgroup. In over half of the illustrations of this model, the test was valid only for the Negro sample. It is ironic that, even in those situations where the test was valid solely for the Negro sample, the use of total group validation procedures would result in test bias. This occurs because the Negro group scores lower on the predictor and would be less likely to be selected.

Model 11, as illustrated in the relationship between final level of achievement on the concept task and post-test scores on the criterion test, occurred eleven times. The test was not valid for either ethnic group separately but showed validity for the total group. The validity in this situation results from the fact that the two groups differ in both predictor and criterion performance. Failure to check the validity in the two groups separately would result in inadvertant racial discrimination.

Model 1 was illustrated in eight of the predictor-criterion relationships. The relationship between the Wonderlic and the criterion post-test clearly illustrates Model 1. White students scored higher on both the Wonderlic and

the criterion post-test. The test was valid for both racial groups. For this type of relationship the difference in predictor performance does not represent test bias, since the predictor reflects a corresponding difference in criteria performance.

The final model illustrated in this sample was Model 3. The mean performance of the two ethnic groups was approximately equal on the Addition test but white students obtained higher scores on the criterion post-test. The test was valid for both racial groups. If a total group validation procedure was utilized, this relationship would result in over-prediction of criterion performance for the Negro sample.

Discussion

Results of the present study tend to support the earlier findings of Jensen (1968) and Selmer and Iscoe (1963) indicating that there is no difference between whites and Negroes in the learning of paired-associate tasks. However, in the present study, performance on the paired-associate task was not related to performance on the programmed instruction criterion for either racial group.

Although measures of acquisition in paired-associate tasks frequently show substantial correlations with measures of academic performance, there have been a number of studies which have failed to show such a relationship (Travers, 1967). The failure of paired-associate tasks to consistently predict classroom learning may result from the factorial complexity of the classroom situation. In some learning situations paired-associate learning may be an important skill (i.e., foreign language vocabulary learning) while in other situations this type of learning may not be important (i.e., solving mathematical equations). The more similar the skills required in the predictor task are to those required in the criterion task, the more valid the predictor should be. The analysis of

the criterion task in the present study indicated that paired-associate learning was not frequently represented in the criterion. Thus, the failure of the paired-associate task to predict criterion performance in the present study may result from the fact that paired-associate learning conditions were not present in the criterion task.

Viewed in terms of mean performance levels, the use of the concept learning task as a predictor did not eliminate differences in mean predictor performance for the two racial groups. In fact, although there was no difference in the initial level of performance for the two groups, white students obtained a higher level of final achievement. Also, the residual gain score analysis revealed that white students showed more gain in proficiency than the Negro students.

The results of the residual gains analysis are in agreement with the findings of Stabler and Perry (1967). They compared the performance of white and Negro college students on a section of Hollard and Skinner's programmed text, The Analysis of Behavior . Despite the fact that the Negro and white students were matched on the basis of sex, age, IQ, and pretest scores, white students obtained higher scores on the post-test.

The concept learning task did exhibit some degree of validity for the Negro sample in predicting error scores but did not predict either final level of achievement or gain in proficiency for either racial group.

Since the analysis of the criterion task revealed mainly principle learning conditions, it was predicted that the principle learning task would be the most valid predictor. This prediction was not supported by the data since the Wonderlic test predicted more criteria than any other predictor. However, final level of achievement on the principle learning task predicted final level of achievement on the criterion test and residual gain scores using subgroup correlations for both racial groups. Moreover, residual gain scores on the

princple learning task predicted final level of achievement and residual gain scores using subgroup correlations for the Negro sample. In terms of mean performance level, the use of the principle learning task did not eliminate racial differences as white students obtained higher post-test scores as well as larger residual gain scores.

As a result of this study, three general conclusions concerning learning measures emerge. First, there is little difference in the learning ability of white and Negro students on relatively simple learning tasks such as the paired-associate task. However, performance on this relatively simple task may not predict performance in complex learning situations such as the programmed instruction criterion utilized in the present study. Second, with more complex learning tasks such as the concept and principle learning tasks, white students obtain higher levels of achievement and also exhibit a larger gain in proficiency than their Negro counterparts. Thus, it is unlikely that the use of such measures as predictors will eliminate mean differences in predictor performance for white and Negro subjects. It should be noted however, that it is not always desirable to reduce mean differences in predictor performance if there are corresponding differences in criterion performance. Third, the results of the present study lend partial support to earlier studies which indicate that it is feasible to use miniature learning situations to predict academic performance. In fact, inspection of Table 55, indicates that, at least with some criteria, the learning measures are more predictive of criterion performance for Negro students than for white students. This result is somewhat encouraging since O'Leary, Farr, and Bartlett (1970) found a large number of situations in which traditional tests were valid for whites but not for Negroes. Perhaps miniature learning situations can be used as predictors for Negroes in such situations.

The traditional psychological tests predicted final level of achievement for both racial groups. Errors scores were predictable for both racial groups using the Wonderlic as the predictor but were predictable only for the white sample using the Addition Test. The Wonderlic and the Vocabulary Test predicted gain scores for white students while the Addition Test predicted gain scores for the Negro group. The Wonderlic Test predicted more criteria than any of the other predictors.

Although there was no difference in the mean performance of the two racial groups on the Digit Span Test, the test did not predict either final level of achievement or gain in proficiency for either racial group. These results tend to be somewhat at variance with results reported by Jensen (1968) indicating that performance on digit span tests was predictable of academic success. Error scores were predictable for the Negro sample using the subtest scores of the Digit Span Test.

In conclusion, the results of the present study indicate that miniature learning situations show some promise as predictors of academic performance for minority group members. Use of the research paradigm utilized in this exploratory study in an actual training program seems warranted.

The regression tests for the analysis of covariance (Potthoff, 1969) are presented in Table 56. This analysis indicated that there was no difference in the slope of the regression line for the two racial groups in any of the predictor-criterion relationships. In approximately half of the predictor-criterion relationships there was a difference in the intercepts for the two racial groups.

Table 56: Analysis of Covariance for Homogenity of Regression

Learning Measure Study[1]

Criterion	Predictor: Paired-Associate Task					
	First Trial			Final Trial		
	F_1[2]	F_2[3]	F_3[4]	F_1	F_2	F_3
Error Scores						
Booklet 1	10.10**	.09	20.32**	12.54**	.81	24.32**
Booklet 2	1.89	.02	3.78	3.05	.82	5.29*
Booklet 3	8.40**	.01	16.98**	9.72**	.06	19.57**
Post Test	16.26**	.06	32.80**	17.82**	.09	35.90**
Residual Gain-Criterion Task	2.62	.03	5.28*	2.79	.03	5.61*
Residual Gain (Using sub-group Correlations)	5.79**	.04	11.66**	6.24**	.02	12.60**

Criterion	Concept Learning Task					
	First Trial			Final Trial		
	F_1	F_2	F_3	F_1	F_2	F_3
Error Scores						
Booklet 1	9.07**	1.01	17.12**	5.89**	.01	11.91**
Booklet 2	2.13	1.65	2.58	.71	.03	1.42
Booklet 3	9.18**	3.58	14.37**	5.17**	.01	10.45**
Post Test	17.70**	.01	35.7 **	15.00**	.58	29.56**
Residual Gain-Criterion Task	3.43	.09	6.85*	2.98	.96	5.0^
Residual Gain (Using sub-group correlations)	6.82**	.03	13.77**	5.90**	.96	18.84**

Table 56: (Contd)

Criterion	Predictor: Principle Learning Task — Pretest F_1	Pretest F_2	Pretest F_3	Posttest F_1	Posttest F_2	Posttest F_3
Error Scores						
Booklet 1	10.32**	.58	20.16**	8.05**	.23	16.01**
Booklet 2	1.44	.00	2.92	1.12	.29	1.98
Booklet 3	7.99**	.07	16.07**	6.53*	.01	13.19**
Post Test	16.79**	1.39	32.07**	10.41**	.01	21.06**
Residual Gain-Criterion Task	3.33	1.24	5.41*	1.38	.30	2.48
Residual Gain (using sub-group correlations)	6.78**	2.03	11.39**	3.39	.39	6.43*

Criterion	Predictor: Wonderlic F_1	Wonderlic F_2	Wonderlic F_3	Vocabulary F_1	Vocabulary F_2	Vocabulary F_3
Error Scores						
Booklet 1	1.37	1.81	.94	5.43**	.44	10.47**
Booklet 2	.19	.01	.38	1.42	.29	2.57
Booklet 3	.03	.00	.05	5.46**	1.27	9.63**
Post Test	2.34	.50	4.19	6.91**	.35	13.57**
Residual Gain-CriterionTask	1.56	2.96	.16	1.50	2.13	.87
Residual Gain (using sub-group correlations)	2.10	3.17	1.01	2.69	1.99	3.36

Table 56: (Contd)

Criterion	Predictor: Addition F_1	Addition F_2	Addition F_3	Digit Span Total Digit Corrects F_1	Digit Span Total Digit Corrects F_2	Digit Span Total Digit Corrects F_3
Error Scores						
Booklet 1	9.83**	.01	19.87	9.97**	3.14	16.40**
Booklet 2	1.83	.14	3.56	1.46	.45	2.49
Booklet 3	8.25**	.01	16.69**	7.13**	.75	13.54**
Post Test	19.20**	.22	38.52**	17.18**	.39	34.20**
Residual Gain-Criterian Task	3.19	.04	6.41*	4.08*	2.15	5.93*
Residual Gain (using sub-group correlations)	6.92**	.21	13.75**	7.10**	1.46	12.68**

Criterion	Predictor: Digit Span: Length of Span of First Error F_1	F_2	F_3	Digit Span: Length of Largest Correct Span F_1	F_2	F_3
Error Scores						
Booklet 1	13.40**	2.67	23.70**	9.15**	1.48	16.72**
Booklet 2	2.55	.27	4.86	1.46	.57	2.37
Booklet 3	9.60**	.06	19.34**	7.89**	1.85	13.80**
Post Test	18.16**	.11	36.56**	17.06**	.53	33.76**
Residual Gain-Criterion Task	3.84	.63	7.08*	3.97*	2.07	5.80*
Residual Gain (using sub-group correlations)	7.25**	.44	14.15**	7.05**	1.58	12.43**

Table 56: (Contd)

Criterion	Predictor: Residual Gain Scores								
	Paired-associate Task			Concept Task			Principle Task		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
Error Scores									
Booklet 1	12.45**	.79	24.16**	7.16**	.20	14.25**	8.12**	.06	16.34**
Booklet 2	3.03	.82	5.24	1.11	.00	2.34	1.48	.42	2.56
Booklet 3	9.66**	.06	19.46**	6.13**	.09	12.29**	6.84**	.00	13.83**
Post Test	17.91**	.06	36.14**	15.56**	.50	30.79**	12.51**	.48	24.69**
Residual Gain-Criterion Task	2.85	.05	5.71*	3.33	1.49	5.14	1.13	.00	3.30
Residual Gain (using subgroup correlations)	6.30**	.00	12.75**	6.24**	1.23	11.25**	3.99*	.00	8.07**

Criterion	Predictor: Residual Gain Scores - (using subgroup correlations								
	Paired-associate Task			Concept Task			Principle Task		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
Error Scores									
Booklet 1	12.62**	.68	24.64**	7.13**	.22	14.17**	7.94**	.07	15.97**
Booklet 2	3.10	.78	5.44**	1.09	.01	2.20	1.33	.26	2.41
Booklet 3	9.77**	.04	19.71**	6.08**	.08	12.21**	6.71	.01	13.56**
Post Test	17.72**	.04	37.77**	15.54**	.50	30.74**	11.66**	.07	23.49**
Residual Gain-Criterion Task	2.74	.07	5.46*	3.32	1.49	5.12	1.50	.09	2.94
Residual Gain (using subgroup correlations)	6.15**	.00	12.43**	6.23**	1.23	11.21**	3.71	.10	7.40**

* $p < .05$ ** $p < .01$

(1) Degrees of freedom for all comparisons: F-(2,90); F_2-(1,90); F_3-(1,91).

(2) F_1 tests hypothesis that $\underline{F}(Y_{ij}|X_{ij}) = a + bX_{ij}$ for all i groups.

(3) F_2 tests hypothesis that $\underline{F}(Y_{ij}|X_{ij}) = a_i + bX_{ij}$ for all i groups.

(4) F_3 tests hypothesis that $\underline{F}(Y_{ij}|X_{ij}) = a + b_iX_{ij}$ for all i groups.
(valid test only if F_2 is not significant)

Summary and Conclusions

The task of integrating the massive amount of data examined is a difficult one. The data were collected in such a wide variety of settings that the uncontrolled and often unknown situational variables have probably had far greater effects on the results than the measures included for study. The sample sizes have varied widely from study to study and from subgroup to subgroup. In order to account for the varying sample size emphasis in interpretation has been placed on statistical significance rather than absolute magnitude of validity coefficients. This decision was made on the basis of an assumption that no validity can be considered of great enough magnitude to be useful in actual personnel selection if its difference from zero can be reasonably attributed to chance.

Interpretation of potential bias in a selection instrument was made according to the Bartlett-O'Leary differential prediction models (1969) and the definition of test bias of Cleary (1966, p. 1): "A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if too high or too low a criterion score is consistently predicted for members of the subgroup when the common regression line is used." The presence of any of the Bartlett-O'Leary models would then indicate bias in the prediction of performance by using the measure for selection, the exception being Model 1. Model 1, where the measure is valid for both groups and differences in performance on both the predictor and the criterion are in the same direction, may result in identical regression lines for both subgroups, with accurately predicted performance differences resulting for both subgroups. The selection situation described under Model 1 cannot be considered resulting in unfair bias since differential selection from the subgroups would be considered a valid prediction of success.

Further assumed in all the interpretations of bias is that the criterion is a valid measure of performance. Implicit in this assumption is that the criterion measure itself is unbiased. This assumption must be subjected to serious question. Caution is therefore urged before a conclusion is drawn that any selection measure is fair because it validly predicts inferiority for a given subgroup. The criterion problem is one that has always plagued all areas of personnel research. The current research is (unfortunately) not immune from this plague.

The General Validity Picture

The sad state of the validity of psychological tests for the prediction of job performance was well summarized by Ghiselli (1955) more than fifteen years ago. Reported validities of tests were found to rarely exceed .50 and more likely to be in the .30 - .40 range. Unreported validities or in situations where no validation has been carried out may not present even this optimistic a picture.

Combining both phases of the present study, more than 1000 validity comparisons have been examined with particular emphasis on breaking these down into ethnic subgroups. By classifying the relationships found according to the Bartlett-O'Leary differential prediction models, it was hoped to demonstrate the frequency for which a measure might be inappropriate for combined group prediction. A Bartlett-O'Leary model was illustrated by more than 40% of the situations examined. These cases indicated the inappropriatness of using a regression line developed on the combined group. One might erroneously conclude on the basis of this summary that prediction on the basis of combined group analysis would thus be appropriate on the more than 50% where a Bartlett-O'Leary model was not found.

Bartlett and O'Leary did not discuss two situations, however, because they should be obvious, but nevertheless attention is drawn to them here. The first might be called Model 0 (as in Zero), where the measure has no validity in the combined group or either subgroup. Model 0 was found in almost half of all the validity comparisons made. Although the use of the test for selection will not

necessarily result in unfair discrimination in this situation a table of random numbers will serve the same purpose more economically, and will further assure no unfair discrimination. The other model is one that is equally as obvious, but unfortunately does not occur often. This might be called Model V (as in valid), where the measure is equally valid for both groups and where there are no differences on either the predictor or criterion. Model V, where it is appropriate to use the predictor for selection on both subgroups combined, was found in fewer than 10% of the situations analyzed. Most of the validities determined appropriate were found for one job (merchandise handler) where multiple measures for both predictors and criterion led to a large number of comparisons. If this particular job situation had not been included, Model V would have occurred in fewer than 2% of the comparisons made.

The large number of situations found where the predictors were completely inappropriate and the very few found where the predictors were appropriate strongly support the need for situation specific test validation for all situations.

The Search for the Unbiased Predictor

The goal of this phase of the project has been stated to be to examine a wider variety of predictors in hopes of determining a way of eliminating the test bias found prevelant in the more traditional measures studied in Phase I (O'Leary, Farr, and Bartlett, 1970). Table 57 reports the frequency of occurrence of the Bartlett-O'Leary models from Phase I. On the basis of these 765 comparisons with these traditional predictors it was concluded that test bias is clearly present in a large number of cases where heterogeneous groups are combined in making predictions. Unless other kinds of measures can be developed which do not lead to such bias, differential test validation must be carried out on identifiable subgroups of the population. The traditional model which assumes homogeneous groups is clearly inappropriate.

Table 57

Frequency of Bartlett - O'Leary Models Illustrated*

Model #	Frequency of Occurrence
1	16
2	28
3	18
4	0
5	163
6	60
7	39
8	28
9	0
10	4
11	1
Total	357
Total Between subjects comparisons	765

* Phase I (O'Leary, Farr, and Bartlett, 1970).

This second phase of the study has investigated less traditional predictors in the hopes of finding procedures which can be used to reduce test bias. Table 58 presents a frequency tabulation of Bartlett-O'Leary Model occurrence with the less traditional predictors. Additional information was also obtained on traditional selection measures and frequency of occurrence of the Bartlett-O'Leary Models is included in Table 58.

As can be seen the use of less traditional predictors does not reduce the frequency of occurrence of test bias. Test bias appears to occur at approximately the same frequency regardless of the nature of the predictor.

Studies #1 and #7 employed work sample tests in the form of a simulation of the job along with tests of motor skills. The result was a disappointing lack of validity for most measures employed.

Studies #4, #5, and #6 took a new look at the formerly disappointing approach of the culture-free test by examining difference scores between culture-loaded tests and the culture-free tests as an index of cultural deprivation. This technique which had been suggested by Guion (1966) proved equally disappointing.

Study #2 and #3 examined biographical information as a predictor of academic performance for high school and college students, respectively. Although biographical information shows some promise as a predictor for the high school students, the problem of differential performance on the criterion would lead to unfair bias in the measure without differential prediction. With differential prediction the results would be similar to those produced by the traditional ability measures.

Study #8 provided the widest disparity from traditional predictor-criterion performance studies, by studying learning measures as predictors as well as criteria in a laboratory setting. The use of traditional ability measures on subgroups from culturally deprived backgrounds has been criticized for the differential learning of the concepts required of these tests which may have taken place as a result of the cultural deprivation. Thus by examining learning measures

Table 58: Frequency of Barclett-O'Leary Models from Phase II

Model	Less Traditional	More Traditional
1	15	19
2	0	1
3	26	6
4	0	0
5	8	0
6	9	15
7	6	9
8	22	10
9	0	0
10	0	0
11	13	10
Total Models	99	70
Total Comparisons	272	128

in the laboratory rather than measures of previously learned material (i.e., traditional tests), it was hoped to reduce the bias for minority groups. The measures from the simple learning tasks showed some promise of validity in predicting learning of complex material, and in some cases eliminated differential predictor performance between the white and Negro groups. The learning measures, at least with some criteria, showed superior validity for the Negro group. Considering the finding that traditional tests usually show higher validities for the white group, this is encouraging. However the superior performance on the criterion by the white group, combined with the finding that the traditional intelligence measure predicted more criteria than any other predictor, leads to the conclusion that learning measures may provide only a slight glimmer of hope as unbiased predictors rather than providing a final solution to the problem. An attempt to extend this study to validate miniature learning measures as predictors of training criteria would appear warranted.

Ponce de Leon searched for the Fountain of Youth, Lancelot for the Holy Grail. The "Unbiased Predictor" has not been found either.

References

Allison, R. B., Jr. Learning measures as predictors of success in torpedoman's mates school. Office of Naval Research Technical Report. Princeton, N.J.: Educational Testing Service, 1954.

Allison, R. B., Jr. Learning measures as predictors of success in pipefitter and metalsmith schools. Office of Naval Research Technical Report. Princeton, N. J.: Educational Testing Service, 1956.

Allison, R. B., Jr. Learning parameters and human abilities. Technical Report. Educational Testing Service and Princeton University, 1960.

Anastasi, A. Some implications of cultural factors for test construction. In A. Anastasi (Ed.), Testing problems in perspective. Washington, D.C.: American Council on Education, 1966.

Bartlett, C. J. and O'Leary, B. S. A differential prediction model to moderate the effects of heterogeneous groups in personnel selection and classification, Personnel Psychology, 1969, 22, 1-17.

Carver, R. P., and DuBois, P. H. The relationship between learning and intelligence. Journal of Educational Measurement, 1967, 4, 133-136.

Cleary, T. Anne. Test bias: validity of the Scholastic Aptitude Test for Negro and white students in integrated colleges. College Entrance Examination Board, Research and Development Reports, RDR-65-6, No. 18, Educational Testing Service, June 1966.

Cronbach, L. J. and Furby, L. How should we measure change - or should we? Psychological Bulletin, 1970, 74, 68-80.

Duncanson, J. P. Learning and measured abilities. Journal of Educational Psychology, 1966, 57, 220-229.

Edwards, D. C. and Hahn, C. P. The use of filmed driver behavior in the study of accidents: The first year's findings. Technical Report. Washington: American Institutes for Research, 1964.

Edwards, D. S., Hahn, C. P., and Fleishman, E. A. Evaluation of laboratory methods for the study of driver behavior: The relation between simulator and street performance. Final Report, R69-7. Washington: American Institutes for Research, 1969.

Ellison, R. L., Fox, L. R., Fox, D. G., and Taylor, C. W. The identification of talent among Negro and white students from biographical data. Salt Lake City, Utah. Institute for Behavioral Research in Creativity, 1970.

Ferguson, G. A. On learning and human ability. Canadian Journal of Psychology, 1954, 8, 95-112.

Ferguson, G. A. On transfer and the abilities of man. Canadian Journal of Psychology, 1956, 10, 121-131.

Frederiksen, N., Carstater, E. D., and Stuit, D. B. Problems for further study. In D. B. Stuit (ed.), Personnel research and test development in the Bureau of Naval Personnel. Princeton, N. J.: Princeton University Press, 1947.

French, J. W., Ekstrom, R. B., and Price, L. A. Kit of reference tests for cognitive factors. (rev. ed.) Princeton, N. J.: Educational Testing Service, 1963.

Gagne, R. M. The conditions of learning. New York: Holt, Rinehart and Winston, Inc., 1965.

Ghiselli, E. E. The measurement of occupational aptitude. Berkeley and Los Angeles: University of California Press, 1955.

Gibson, J. T. Educational psychology. New York: Appleton, Century, Crofts, 1968.

Guion, R. M. Employment tests and discriminatory hiring. Industrial Relations, 1966, 5, 20-37.

Hackett, E. V. The use of learning tests in the analysis of training. In P. H. DuBois and K. M. Wientge (eds.), Strategies of research on learning in educational settings. Office of Naval Research and Office of Education Technical Report. St. Louis: Washington University, 1964.

Hollingshead, A. B. Elmtown's youth. New York: Wiley, 1949.

Jenkins, J. J. Individual differences in verbal learning. In R. M. Gagne's Learning and individual difference. Columbus, Ohio: Charles E. Merrill Books, Inc., 1967.

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

Kirkpatrick, J. J., Ewen, R. B., Barrett, R. S., and Katzell, R. A. Testing and fair employment. New York: New York University Press, 1968.

Krug, R. E. Some suggested approaches for test development and measurement. Personnel Psychology, 1966, 19, 24-34.

Lesser, G. S., Fifer, G. and Clark, D. H. Mental abilities of children from different social-class and cultural groups. Monographs of the society for research in child development, 30,(4), 1965.

Leventhal, A. M. An anxiety scale for the CPI. Journal of Clinical Psychology, 1966, 22, 459-461.

Lord, F. M. Efficiency of prediction when a regression equation from one sample is used in a new sample. Research Bulletin 50-40. Princeton, N.J.: Educational Testing Service.

Manning, W. H., and DuBois, P. H. Gain in proficiency as a criterion in test validation. Journal of Applied Psychology, 1958, 42, 191-194.

Manning, W. H., and DuBois, P. H. Correlational methods in research on human learning. Perceptual and Motor Skills, 1962, 15, 287-321.

McGeoch, J. A. and Irion, A. K. The psychology of human learning. New York: Longmans, Green, 1952.

Mitchell, M. D., Albright, L. F., and McMurray, F. D. Biracial validation of selection procedures in a large sourthern plant. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 575-576.

Noble, C. E. Measurements of association value (a), rated associations(a^1), and scaled meaningfulness (m^1) for the 2100 CVC combinations of the English alphabet. Psychological Reports, 1961, 8, 487-521.

O'Leary, B. S., Farr, J. L., and Bartlett, C. J. Ethnic group membership as a moderator of job performance. Washington, D.C.: American Institutes for Research, 1970.

Potthoff, R. F. Statistical aspects of the problem of biases in psychological tests. Chapel Hill, N.C.: University of North Carolina, Institute of Statistics Mimeo Series No. 479, May 1966.

Rapier, J. L. Measured intelligence and the ability to learn. Acta Psychology, 1962, 20, 1-7.

Ruda, E. and Albright, L. E. Racial differences on selection instruments related subsequent job performance. Personnel Psychology, 1968, 21, 31-41.

Saunders, D. R. Moderator variables in prediction. Educational and Psychological Measurement, 1956, 16, 209-222.

Schoer, L. A. An introduction to statistics and measurement. Boston: Allyn and Bacon, Inc., 1966.

Semler, I. J., Jr. and Iscoe, I. Comparative and developmental study of the learning abilities of Negro and white children under four conditions. Journal of Educational Psychology, 1963, 54, 38-44.

Sheppard, H. L., and Striner, H. E. Civil rights, employment, and the social status of American Negroes. Kalamazoo, Michigan: The W. E. Upjohn Institute for Employment Research, 1966.

Stabler, J. R. and Perry, O. B. Learning and retention as a function of instructional method and race. Journal of Psychology, 1967, 67, 271-276.

Stolurow, L. M. Teaching by Machines. Cooperative Research Monograph, No. 6, Washington, D.C.: U. S. Office of Education, 1961.

Stodolsky, S. S. and Lesser, G. Learning patterns in the disadvantaged. Harvard Educational Review, 1967, 37, 546-593.

Travers, R. M. W. Learning measures and psychometric variables. In R. M. Gagné, Learning and individual differences. Columbus, Ohio: Charles E. Merrill Books, Inc., 1967.

Wardrop, J. L. A learning test approach to predicting classroom performance. Office of Naval Research Report. St. Louis: Washington University, 1967.

Wonderlic, E. F. and Associates, Inc. The Wonderlic Personnel Test. Northfield, Illinois, 1969.

APPENDIX A - CRITERION TEST

MEASUREMENT TEST

Blacken in the correct answer on the Answer Sheet provided. DO NOT write in the test booklet.

1. A measure of central tendency which depends on every score in the distribution is the:

   (a) mean

   (b) median

   (c) mode

   (d) all of the above

2. What is the median of the following set of scores?
   27, 28, 29, 30, 31, 31, 31

   (a) 30

   (b) 31

   (c) 28.1

   (d) 27.5

3. What percentile rank corresponds to a score one standard deviation above the mean?

   (a) 16

   (b) 34

   (c) 50

   (d) 84

4. John has computed the dispersion of a first grade class on their final spelling test. Unknowingly, John has included the teacher's test score in the measure. If John eliminates the teacher's test score, which one of the following measures will not be affected?

   (a) standard deviation

   (b) inclusive range

   (c) interquartile range

   (d) percentile rank

5. One extreme score can effect which of the following measures?

   (a) mean

   (b) inclusive range

   (c) both a and b

   (d) none of the above

6. Assume that IQ scores are normally distributed with a mean of 100 and a standard deviation of 15. What proportion of people have scores above 115.?

   (a) 16

   (b) 34

   (c) 50

   (d) 84

7. In a normal distribution:

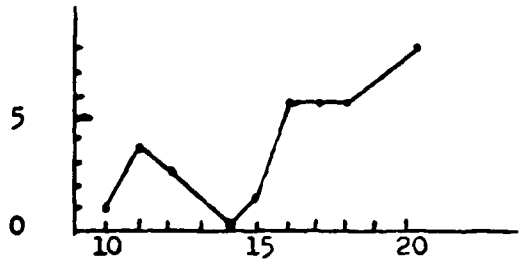
   (a) at least 75% of the cases lie on one side of the mean

   (b) 68% of the cases lie between plus and minus one standard deviation from the mean

   (c) the median must fall either above or below the mean

   (d) all scores lie within one standard deviation from the mean

8. Which number(s) represent(s) the mode in the following frequency distribution?



   (a) 11

   (b) 14

   (c) 16, 17, 18

   (d) 20

9. Which of the following measures is defined as the distance from the highest to lowest score plus one?

(a) standard deviation

(b) inclusive range

(c) quartile

(d) semi-quartile range

10. Which of the following is a measure of the dispersion of a set of scores?

(a) range

(b) standard deviation

(c) interquartile range

(d) all of the above

11. If a distribution is normal, the approximate per cent of cases falling within two standard deviations from the mean in either direction is:

(a) 100

(b) 92

(c) 68

(d) 50

12. Find the inclusive range of the following set of scores.

100,000 ; 50,000 ; 10,000 ; 1,000

(a) 99,001

(b) 40,000

(c) 20,000

(d) none of the above

13. If 60 per cent of the students score less than 24 points on a test, a score of 24 has a _____ of 60.

(a) percentile

(b) percentile rank

(c) quartile

(d) none of the above

14. When a score in the distribution is changed, the _____ will always change.

(a) median

(b) mean

(c) mode and median

(d) mode, median, and mean

15. A bowling score which occurs most often in the course of a season of bowling is the bowler's _____.

(a) median

(b) mode

(c) mean

(d) either mean or median

16. The median is the:

(a) midpoint between the largest and smallest score

(b) midpoint of the score class that contains 50% of the cases

(c) score point that has as many cases below it as above it

(d) average of the larger and smaller scores

17. In a normal distribution the:

(a) mean > mode > median

(b) mean = mode = median

(c) mean > median > mode

(d) mode > median > mean

18. A deviation score is a measure of how much a single score deviates from the _____ of a group.

(a) mean

(b) median

(c) mode

(d) none of the above

19. If a score of 20 has a percentile rank of 40, it means that _____per cent of the students scored fewer than _____ points on the test.

(a) 20, 40

(b) 40, 20

(c) 60, 80

(d) can not be determined from the above information

20. The score value at the most frequently occurring score in the distribution may be the _____, but must be the _____.

(a) mode, median

(b) mode and mean, median

(c) mode, mean and median

(d) mean and median, mode

21. Joe's percentile rank was as follows on 4 tests taken at the beginning of school.

Math	60	Georgraphy	50
English	65	History	55

Joe was above the mean in:

(a) Math, English, and Georgraphy

(b) Math, English and History

(c) all of the above

(d) not enough information to determine

22. The measure which is based on only two scores in the distribution is:

(a) inclusive range

(b) standard deviation

(c) coefficient of correlation

(d) median

23. Compute the interquartile range of the following set of scores

Score	Percentile Rank
13	89
12	75
11	60
10	42
9	25
8	13
7	5

(a) 3

(b) 6

(c) 50

(d) 84

24. The mode is:

(a) a point on the score scale below which lie one-half the scores

(b) the largest score in a set of data

(c) the sum of the scores d vided by the number of correct answers

(d) the score with the largest frequency

25. Given the following set of scores, what is the percentile rank of a score of 30?

10 ; 20 ; 30 ; 40 ; 50

(a) 10

(b) 20

(c) 40

(d) 60

26. A teacher decides to fail 16% of her class. Final class grades are roughly normally distributed with a mean of 72 and a standard deviation of 6. What mark must a student make to pass the course?

(a) 60

(b) 66

(c) 68

(d) 72

27. The third quartile is equal to the:

(a) 25 th percentile

(b) 50 th percentile

(c) 75 th percentile

(d) none of the above

28. If several mathematical wizards (college professors) are included in a high school mathematics class, their scores will affect which measure of central tendency the most:

(a) mean

(b) median

(c) mode

(d) depends on the number of students in the class

29. The median is equal to the

(a) 25% tile

(b) 50% tile

(c) 75% tile

(d) none of the above

30. What is the mean of the following set of data?
1, 3, 3, 4, 5, 6, 6, 7, 7, 8, 9

(a) 3.5

(b) 7

(c) 6 and 7

(d) 5.4

APPENDIX B - PRINCIPLE LEARNING TEST

Name _____

TEST OF YOUR GENERAL KNOWLEDGE OF MEASUREMENT

For each kind of measurement listed below decide whether the scale has the characteristics most like a NOMINAL or ORDINAL scale. Place an N beside those which are NOMINAL and an O beside those which are ORDINAL.

1. _____ Postal zones (zip codes)
2. _____ Dewey Decimal Classification in library
3. _____ Sex
4. _____ Win, place, show at horse race
5. _____ Telephone numbers
6. _____ Starting position in automobile race
7. _____ Goodness of handwriting
8. _____ Political affiliation (Republican, Democrat, etc.)
9. _____ Number of years of education
10. _____ Amount of weight lost on diet
11. _____ Medical diagnosis (Cancer, T. B., Polio, etc.)
12. _____ Major league baseball standings
13. _____ Class standing at graduation
14. _____ Grades in school
15. _____ Race of employee
16. _____ License plate members
17. _____ Socioeconomic status
18. _____ Religious denomination
19. _____ Major in college (Engineering, Liberal Arts, Education, etc.)
20. _____ Route numbers on highway signs
21. _____ Girls ranked in terms of beauty
22. _____ Marital status
23. _____ Mean shoe size
24. _____ Age in years
25. _____ Ratings of job performance
26. _____ Make of car (Ford, Plymouth, Chevrolet, etc.)
27. _____ Congressional district numbers
28. _____ Library of Congress Classification system in library
29. _____ Social security numbers
30. _____ Preference for types of ice cream
31. _____ County of birth
32. _____ Income level of college professors
33. _____ Rank in military service (Major, Captain, Private, etc.)
34. _____ Growth in inches from ages 10 to 15
35. _____ Intelligence test scores
36. _____ Psychiatric diagnosis (Schozphrenic, Paranoid, Neurotic, etc.)
37. _____ Numbers on football players' jersey
38. _____ Subjective pitch (apparent highness or lowness of tone)
39. _____ Texonomy in biology (phyla, genera, species, etc.)
40. _____ Attitude toward Vietnam war (Favorable, Neutral, Unfavorable)