DOCUMENT RESUME

ED 069 685 TM 002 136

AUTHOR Ruch, William W.
TITLE A Re-analysis of Published Differential Validity Studies.
PUB DATE 6 Sep 72
NOTE 35p.; Presented at the symposium, "Differential Validation under EEOC and OFCC Testing and Selection Regulations," (American Psychological Association, Honolulu, Hawaii, Sept. 6, 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Aptitude Tests; Data; Employment Opportunities; Evaluation Techniques; Industrial Personnel; *Measurement Techniques; Multiple Regression Analysis; *Negro Employment; *Predictive Ability (Testing); Racial Differences; Racial Discrimination; *Task Performance; Technical Reports; Test Interpretation; Test Validity

ABSTRACT

A survey of recent literature was undertaken to locate validity studies of paper-and-pencil tests which met the following criteria: (1) Studies were conducted in a business or industrial (i.e. non-education, non-military) setting; (2) Separate statistics were available for blacks and whites; (3) Race was not confounded with some outside variable which would preclude meaningful interpretation; (4) Necessary data were reported to enable a test of homogeneity of regression between racial groups. For each of 20 studies which met these criteria, a homogeneity of regression analysis was conducted on each predictor-criterion pair to determine if there were significant differences between blacks and whites in standard errors, slopes, or intercepts of the regression lines. The number of significant differences in standard errors and in slopes was less than would be expected by chance, indicating that tests do not have differential validity between white and black groups. For intercepts, significant differences in excess of chance were obtained. The direction of the differences was such that job performance of blacks was overestimated by tests. (Author)

A RE-ANALYSIS OF PUBLISHED DIFFERENTIAL VALIDITY STUDIES

by

William W. Ruch

Psychological Services, Inc.

Los Angeles, California

Presented at

the Symposium on:

"Differential Validation under EEOC and OFCC Testing and Selection Regulations"

American Psychological Association

Honolulu, Hawaii

September 6, 1972

A B S T R A C T

A survey of the recent literature was undertaken to locate validity studies of paper-and-pencil tests which met the following criteria:

1. Studies were conducted in a business or industrial (i.e. non-educational, non-military) setting;
2. Separate statistics were available for blacks and whites;
3. Race was not confounded with some outside variable which would preclude meaningful interpretation;
4. Necessary data were reported to enable a test of homogeneity of regression between racial groups.

For each of 20 studies which met these criteria, a homogeneity of regression analysis was conducted on each predictor-criterion pair to determine if there were significant differences between blacks and whites in standard errors, slopes, or intercepts of the regression lines. The number of significant differences in standard errors and in slopes was less than would be expected by chance, indicating that tests do not have differential validity between white and black groups. For intercepts, significant differences in excess of chance were obtained. The direction of the differences was such that job performance of blacks was overestimated by tests.

A RE-ANALYSIS OF PUBLISHED DIFFERENTIAL VALIDITY STUDIES*

William W. Ruch
Psychological Services, Inc.
Los Angeles, California

A survey of the recent literature was undertaken to locate validity studies of paper-and-pencil tests which met the following criteria:

1. Studies were conducted in a business or industrial (i.e. non-educational, non-military) setting;
2. Separate statistics were available for blacks and whites;
3. Race was not confounded with some outside variable which would preclude meaningful interpretation;
4. Necessary data were reported to enable a test of homogeneity of regression between racial groups.

Studies which met some but not all of the above-stated criteria were excluded. Baehr's (1) study of Chicago policemen was excluded since the criteria were confounded with race due to the fact that, in the words of the author, "...Negro patrolmen tend more often to be assigned to predominately Negro districts, which would include the ghetto areas. Since the crime rate in these districts is usually higher, it is natural that more arrests will be made. It also seems likely that patrolmen who are constantly 'where the action is' will have more complaints made against them and, possibly, have more of

---

* Presentation in the symposium, "Differential Validation under EEOC and OFCC Testing and Selection Regulations," at the annual meetings of the American Psychological Association, Honolulu, Hawaii, September 6, 1972.

them sustained." This confounding is evidenced by the fact that in her Wave I sample, the average Negro officer made 81 arrests, versus 40 for whites; the average Negro officer had 2.6 disciplinary actions versus .8 for whites. If we view making arrests as good, and being subjected to disciplinary action as bad, it is obviously impossible for a test which is fair on one criterion to be fair on the other also.

Also excluded on the basis of confounding was Kirkpatrick's (10) study on Nursing Students in which race was 100% confounded with institution. Another reason for omitting this study was that it is in an educational, rather than in an industrial setting. Kirkpatrick's (10) Study 5 on Clerical Insurance Employees was also omitted on the basis of a near 100% confounding of race and institution. Since criteria were standardized within institution there was no way to test for the significance of differences between standard errors, slopes or intercepts, since all of these require between-group comparisons of criterion means and/or standard deviations.

Since only continuous variables will fit the model employed, all dichotomous turnover criteria were deleted. This resulted in omitting the study by Ruda and Albright (16) in which turnover was the sole criterion and deleting the turnover criterion from several of the Farr and O'Leary studies (3). Also deleted from several of the Farr and O'Leary (3) studies were the "extension of probation" and the "promotion" criteria. While a study by Lefkowitz (11)

involved a continuous turnover criterion based upon the number of days employees remained with the company, with a maximum of 60, this study was omitted due to the markedly truncated and U-shaped nature of the distribution.

Studies by Lopez (4), Mitchell and Albright (14), and Wollowick (18) were omitted because all of the data necessary for the present analysis were not reported.

All of the military studies were omitted somewhat arbitrarily. These include 11 studies by Gordon (6), 10 by Guinn, Tupes and Alley(8), and 1 by Farr *et al* (4). As will be seen, these studies and the ones in the present report yield essentially the same results.

PROCEDURE

In most of the studies there were several predictors and several criteria. In order to look for patterns, each combination of predictor and criterion was treated separately. That is, if a study had 6 predictors and 10 criteria, 60 analyses were made for that study. It must be pointed out that the criteria are highly intercorrelated - not only for the usual reasons of halo, but also because they included many part-whole and corrected-uncorrected situations. The predictors are also highly intercorrelated, sometimes involving two scoring formulas and two or three time limits for the same test. Thus, elements in the test by criterion matrix certainly cannot be treated as independent observations.

The method of analysis used was that of significance tests of homogeneity of regression between whites and blacks as formulated by Gulliksen and Wilks (9). Since so few studies involved other minorities, these subjects were not included in the present analysis. Therefore, *all conclusions must be limited to white versus black differences*.

Within each study and for each predictor-criterion combination, three significance tests were run. First, the significance of the difference between standard errors of whites and of blacks was assessed. If this proved to be significant, the second two tests were not run. If it was not significant, the significance of differences between slopes of regression lines was assessed. If this was significant, the final test was not run. If it was non-significant, the significance of differences of intercepts was assessed. The 5% level of significance was used throughout.

If the standard errors were significantly different, the difference was indicated as <u>negative</u> if the standard error was smaller for whites and <u>positive</u> if it was greater. A significant difference between slopes was recorded as <u>negative</u> if the slope was smaller for blacks and <u>positive</u> if the slope was greater for blacks. If there was a significant difference between intercepts, the black test mean was plugged into the white regression equation to predict the black criterion mean and the actual black criterion mean was subtracted from the results. Thus, a <u>plus</u> means that the

test overpredicts for blacks - that is the test is unfair to whites - and a minus means that the test underpredicts for blacks - that is, it is unfair to blacks. Popular hypotheses which have been advanced are that tests are less valid for blacks and/or they are unfair to blacks. These hypotheses are symbolized in this analysis with a minus for standard error, a minus for slope and a minus for intercept.

RESULTS

Results for each of the 20 studies included in this report are attached. The first is Kirkpatrick's Study 1 on Female Clerical Workers in an Insurance Company. Note that there are four predictors and five criteria for a total of 20 validity regression equations. For each of these 20, the test of significance of difference in standard errors was conducted. Had any of these been significant the letters "SE" would have been placed in the appropriate element in the matrix. No significant differences were found so none is reported.

The next test was for significance of difference in slopes. None was found and none is reported. The final test was for significance of intercepts. We see that all four tests showed significantly different intercepts when predicting the merit rating criterion. Since all of these are indicated as minus, we know that the criterion was underpredicted for blacks. That is, the tests would be called unfair to blacks. Out of 20 significance tests made for differences in intercepts, four were found to be significant. Since this is greater

than the 5% which would be expected by chance, it is so indicated in the summary table, with a minus which indicates the direction of the difference. The zero for the standard error and the zero for the slope indicate that there were fewer significant findings than would be expected by chance. For each of the 20 studies included in this report, standard error, slope and intercept are each designated as whether they show significance in one direction, have fewer than chance significant findings, or are significant in the other direction.

I would like to point out in this first study that although it is tabulated as a situation in which tests are unfair to blacks, there is perhaps a better reason to suspect the criterion. If a test is unfair - that is has a cultural bias - we would expect it to be unfair against many or all of the criteria. That is, we would expect to have significant findings in rows corresponding to the biased tests. However, in this instance, we have significant findings in a column, perhaps indicating a biased criterion in which blacks are rated spuriously high. Note that this is the only subjective criterion in Study 1.

Let's look at one more table. Study 2 is Kirkpatrick's second study of Female Clerical Employees in an Insurance Company. Out of 28 significance tests we find 16 significant differences in standard errors, far more than would be expected by chance. Note that they are all in the same direction, with the standard error being greater for whites than for blacks. In every instance when more

than one significant finding was found in a study, all were in the same direction. This is due to the high intercorrelation of tests and intercorrelation of criteria, which, incidentally, make a joint probability test of significance impossible. For the 12 situations in which the standard error was not significant, the slope test was run, yielding one significant finding. This is indicated in the summary table as greater than chance, since one out of 12 is greater than 5%. Of all 11 tests of intercepts run, 3 were significant, also greater than chance. The study is scored as plus, minus, plus indicating that all three tests were in excess of chance with the standard error being greater for whites, the slope being smaller for blacks, and the tests being unfair to whites.

Under the null hypothesis that tests work exactly the same for blacks as they do for whites, in the long run we would expect to find exactly one-half of our studies showing statistical significance in excess of chance and exactly one-half of our studies showing statistical significance fewer times than would be predicted by chance. Thus, if chance alone were operating in the 20 studies, we would expect 10 studies in which fewer significant findings occurred than would be expected by chance, and 10 in which more significant findings occurred than would be expected by chance. Of these 10 in excess of chance, we would expect 5 to be in the plus direction and 5 to be in the minus direction. That is, for the 20 studies in this report, if only chance were operating we would expect the distribution of minus, zero and plus to be 5, 10 and 5.

The summary on page 33 gives these tabulations. We see that for tests of standard error, we get 13 with fewer than chance significance, 2 with more than chance significance and the standard error smaller for whites, and 5 with more than chance significance and the standard error greater for whites. There is no evidence whatsoever that the standard errors differ between the two groups. The summary of the tests of slopes is also a chance distribution. There is no evidence of differential validity. The tests for intercepts do show a significant pattern as evidenced by a chi-square of 6.40, significant at the 5% level. However, the criterion scores of blacks are overestimated by tests, not underestimated. That is, interpreted in the same manner for blacks as they are for whites, tests are unfair to whites.

Another way to summarize all of the enclosed studies is simply to count up the total number of significant findings and to divide by the total number of significance tests run. This has the undesirable effect of weighting studies by the number of validity coefficients run (number of predictors times number of criteria). Also, since the tests and the criteria are so highly intercorrelated, there is no way to assess the significance of any departure from the 5% expected under the null hypothesis.

Of the 618 tests of significance of differences between standard errors, 72 (12%) were significant at the 5% level. Of the 546 tests for slopes, 64 (12%) were significant. Of the 482 tests for intercepts 87 (18%) were significant. While these are somewhat in excess of

chance, it should be pointed out that 30 of the 72 significant standard error tests were all from study 11, and 39 of the significant 64 slopes tests were all from study 13. Both of these studies were of the same set of predictors which included multiple scoring formulas and multiple time limits for the same tests, thus making repeated significance tests on practically the same data.

The problems caused by these spuriously intercorrelated predictors as well as the problem of weighting findings by total number of validity coefficients can be offset to a great extent by determining the median percent of significant findings across the 20 studies. For the significance of the difference between standard errors, the median percent significant was zero - certainly less than the 5% expected by chance. For slopes, the median percent significant was between 2% and 3%, and for intercepts between 4% and 14%.

While this method of summarizing the results of the 20 studies has weaknesses it is consistent with the summary on page 33.

Certainly these 20 studies do not tell the whole story. The evidence that they do provide is that there is no such thing as differential validity but there is a tendency of tests to overestimate black job performance. This is exactly what has been found in most of the studies in the military. Guinn, Tupes and Alley (8) summarized 10 studies by stating that: "Assuming that the performance criterion was unbiased, results indicate that when statistically significant differences in levels of regression lines were found the performance of Negroes and high school non-graduates tended to be overestimated."

If we follow the OFCC and EEOC Guidelines, and conduct validation studies separately for blacks, we are likely to find that between-group differences in test scores do not correspond to between-group differences in job performance. If we then follow the Guidelines and adjust cutoff scores "so as to predict the same probability of job success in both groups," we will have to raise, not lower, the passing scores for blacks. Thus, following the OFCC and EEOC Guidelines will reduce, not increase, the employment opportunities of blacks.

REFERENCES

1. Baehr, M. E. *et al* Psychological Assessment of Patrolman Qualifications in Relation to Field Performance. Superintendent of Documents, U. S. Government Printing Office: Washington, D. C., November 5, 1968.

2. Campbell, J. T., Pike, L. W., & Flaugher, R. L. *Prediction of job performance for Negro and white medical technicians - A regression analysis of potential test bias: Predicting job knowledge scores from an aptitude battery.* (Educational Testing Service Rep. PR-69-6) Princeton, New Jersey: Educational Testing Service, 1969.

3. Farr, J. L., O'Leary, B. S., & Bartlett, C. J. Ethnic group membership as a moderator of the prediction of job performance. *Personnel Psychology*, 1971, 24, 609-636.

4. Farr, J. L. *et al* *Ethnic group membership as a moderator in the prediction of job performance: An examination of some less traditional predictors.* (AIR Technical Rep. 2) Washington, D. C.: American Institutes for Research, September, 1971.

5. Gael, S., & Grant, D. L. Employment test validation for minority and nonminority telephone company service representatives. *Journal of Applied Psychology*, 1972, 56, 135-139.

6. Gordon, M. A. A study in the applicability of the same minimum qualifying scores for technical schools to white males, WAF and Negro males. (Technical Rep. 53-34) Lackland AFB, Texas: Human Resources Research Center, November, 1953.

7. Grant, D. L., & Bray, D. W. Validation of employment tests for telephone company installation and repair occupations. *Journal of Applied Psychology*, 1970, 54, 7-14.

8. Guinn, N., Tupes, E. C., & Alley, W. E. Cultural subgroup differences in the relationships between Air Force aptitude composites and training criteria. (Technical Report 70-35) Lackland Air Force Base, Texas: Human Resources Research Center, September, 1970.

9. Gulliksen, H., & Wilks, S. S. Regression tests for several samples. *Psychometrika*, 1950, 15, 91-114.

10. Kirkpatrick, J. J. *et al* *Testing and fair employment.* New York: New York University, 1968.

11. Lefkowitz, J. Differential validity: Ethnic group as a moderator in predicting tenure. *Personnel Psychology*, 1972, 25, 223-240.

12. Lopez, F. M., Jr. Current problems in test performance of job applicants. *Personnel Psychology*, 1966, 19, 10-18.

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

14. Mitchell, M. D., Albright, L. E., & McMurry, F. D. Biracial validation of selection procedures in a large southern plant. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, 3, 575-576. (Summary)

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

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

17. Tenopyr, M. L. Race and socioeconomic status as moderators in predicting machine-shop training success. Paper presented at the meeting of the American Psychological Association, Washington, D. C., September, 1967.

18. Wollowick, H. B., Greenwood, J. M., & McNamara, W. L. Psychological testing with a minority group population. *Proceedings of the 77th Annual Convention of the American Psychological Association*, 1969, 4, 609-610. (Summary)

STUDY 1

Kirkpatrick (10): Study 1 - Female Clerical Employees in an Insurance Company

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Merit Rating	Job Grade	Current Pay	Pay (controlled for tenure)	Pay Increase (controlled for tenure)
SET Verbal	I-				
SET Numerical	I-				
SET Clerical	I-				
SET Total	I-				

(Sample Sizes: 100 Whites; 23-26 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/20 = 0%	Less	0
Slope	0/20 = 0%	Less	0
Intercept	4/20 = 20%	More	- (Blacks underpredicted)

STUDY 2

Kirkpatrick (10): Study 2 - Verbal and Non-Verbal Tests with Female Clerical Employees in an Insurance Company

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Quantity of Work	Quality of Work	Overall Performance	Overall Effectiveness
Vocabulary		SE+	SE+	SE+
Numerical	S-	SE+		I+
Checking		SE+	SE+	SE+
Coding		SE+	SE+	SE+
Total Score		SE+	SE+	SE+
SRA - Non Verbal		SE+		I+
DAT Abstract Reasoning		SE+	SE+	I+

(Sample Sizes: 36-38 Whites; 22-31 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	16/28 = 57%	More	+ (Greater for Whites)
Slope	1/12 = 8%	More	- (Smaller for Blacks)
Intercept	3/11 = 27%	More	+ (Blacks Overpredicted)

STUDY 3

Kirkpatrick (10): Study 3 - Male Trainees for Low-Level Manual Occupations (General Maintenance)

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Performance Task
Numerical Ability	
Reading Ability	

(Sample Sizes: 23 Whites; 39 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/2 = 0%	Less	0
Slope	0/2 = 0%	Less	0
Intercept	0/2 = 0%	Less	0

STUDY 4

Kirkpatrick (10): Study 3 - Male Trainees for Low-Level Manual Occupations (Heavy Vehicle)

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Verbal Proficiency Test
Numerical Ability	
Reading Ability	

(Sample Sizes: 39-40 Whites; 38-39 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/2 = 0%	Less	0
Slope	0/2 = 0%	Less	0
Intercept	0/2 = 0%	Less	0

STUDY 5

Farr and O'Leary (3): Toll Collectors

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Attendance (Days Absent)	Attendance* (Periods absent in 3 months)	Dollar Accuracy	Axle Accuracy
Clerical Checking				
Arithmetic Reasoning		S+		

* Reflected

(Sample Sizes: 83-97 Whites; 33-40 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/8 = 0%	Less	0
Slope	1/8 = 13%	More	+ (Greater for Blacks)
Intercept	0/7 = 0%	Less	0

STUDY 6

Farr and O'Leary (3): Correctional Officers

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Test	Attendance* (Days Absent)	Rating by by Supervisor
California Test of Mental Maturity	S-	

*. Reflected

(Sample Sizes: 207 Whites; 41 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/2 = 0%	Less	0
Slope	1/2 = 50%	More	- (Smaller for Blacks)
Intercept	0/1 = 0%	Less	0

STUDY 7

Farr and O'Leary (3): Toll Facility Officers

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Test	Attendance* (Days Absent)	Attendance (Periods Absent)	Rating by by Supervisor
Otis Quick-Scoring	SE+		SE+

* Reflected

(Sample Sizes: 51-56 Whites; 16-18 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	2/3 = 67%	More	+ (Greater for Whites)
Slope	0/1 = 0%	Less	0
Intercept	0/1 = 0%	Less	0

STUDY 8

Farr and O'Leary (3): Home Office Clerical

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Quickness		Accuracy		Numerical Ability		Verbal Ability		Judgment		Overall Mental Ability		Promotion Potential	
	Off Mgr	Imm Sup	Off Mgr	Imm Sup	Off Mgr	Imm Sup	Off Mgr	Imm Sup	Off Mgr	Imm Sup	Off Mgr	Imm Sup	Off Mgr	Imm Sup
Original TMA	I+		I+	I+			I+				I+	I+	I+	I+
New TMA	I+		I+	I+			I+				S+	I+	I+	I+
Picture Selection Test	I+		I+				I+				I+	I+	I+	I+

(Sample Sizes: 233-328 Whites; 21-43 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/42 = 0%	Less	0
Slope	1/42 = 2%	Less	0
Intercept	22/41 = 54%	More	+(Blacks Overpredicted)

STUDY 9

Farr and O'Leary (3): Keypunch Operators

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Concentration	Learning Ability	Work Sharing	Error Detection	Social Interaction	General Overall Effectiveness	Keypunch Speed	Error Percentage
Verbal	SE+							SE+
Quantitative	I+							SE+
Clerical	I+							SE+
Arithmetic	I+							SE+
Active								SE+
Vigorous								SE+
Impulsive								SE+
Dominant								SE+
Emotional Stability			S+					SE+
Sociable								SE+
Reflective								SE+

(Sample Sizes: 60-101 Whites; 13-24 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	12/88 = 14%	More	+ (Greater for Whites)
Slope	1/76 = 1%	Less	0
Intercept	3/75 = 4%	Less	0

STUDY 10

O'Leary et al (13): Catalog Order Plant - Material Handler I

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Accuracy	Accuracy Under Pressure	Work Speed	Learning Ability	Human Relations	Overall Effectiveness
Verbal Reasoning						
Arithmetic Reasoning						
Clerical I - 5 minutes						
Clerical I - 10 minutes						
Clerical II - 5 minutes						
Clerical II - 10 minutes						
Clerical I (R-W) - 5 minutes						
Clerical I (R-W) - 10 minutes						
Clerical II (R-W) - 5 minutes	S+					
Clerical II (R-W) - 10 minutes	S+					

(Sample Sizes: 86 Whites; 84 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/60 = 0%	Less	0
Slope	2/60 = 3%	Less	0
Intercept	0/58 = 0%	Less	0

STUDY 11

O'Leary et al (13): Catalog Order Plant - Material Handler II

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Accuracy	Accuracy Under Pressure	Work Speed	Learning Ability	Human Relations	Overall Effectiveness
Verbal Reasoning			SE+		SE+	SE+
Arithmetic Reasoning			SE+		SE+	SE+
Clerical I - 5 minutes			SE+		SE+	SE+
Clerical I - 10 minutes			SE+		SE+	SE+
Clerical II - 5 minutes			SE+	I+	SE+	SE+
Clerical II - 10 minutes			SE+		SE+	SE+
Clerical I (R-W) - 5 minutes			SE+		SE+	SE+
Clerical I (R-W) - 10 minutes			SE+		SE+	SE+
Clerical II (R-W) - 5 minutes			SE+		SE+	SE+
Clerical II (R-W) - 10 minutes			SE+		SE+	SE+

(Sample Sizes: 122 Whites; 125 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	30/60 = 50%	More	+(Greater for Whites)
Slope	0/30 = 0%	Less	0
Intercept	1/30 = 3%	Less	0

STUDY 12

O'Leary et al (13): Catalog Order Plant - Clerical I

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Accuracy	Accuracy Under Pressure	Work Speed	Learning Ability	Human Relations	Overall Effectiveness
Verbal Reasoning			SE+			
Arithmetic Reasoning						
Clerical I - 5 minutes						
Clerical I - 10 minutes			SE+			SE+
Clerical II - 5 minutes						
Clerical II - 10 minutes			SE+			
Clerical I (R-W) - 5 minutes						
Clerical I (R-W) - 10 minutes			SE+			SE+
Clerical II (R-W) - 5 minutes						
Clerical II (R-W) - 10 minutes						

(Sample Sizes: 99 Whites; 22 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	6/60 = 10%	More	+(Greater for Whites)
Slope	0/54 = 0%	Less	0
Intercept	0/54 = 0%	Less	0

STUDY 13

O'Leary et al (13): Catalog Order Plant - Machine Clerical

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Accuracy	Accuracy Under Pressure	Work Speed	Learning Ability	Human Relations	Overall Effectiveness
Verbal Reasoning					I+	I+
Arithmetic Reasoning					S-	I+
Clerical I - 5 minutes	S-	S-	S-	S-	S-	
Clerical I - 10 minutes	S-	S-	S-	S-	S-	S-
Clerical II - 5 minutes		S-	S-	S-	S-	S-
Clerical II - 10 minutes		S-		S-	S-	I+
Clerical I (R-W) - 5 minutes	S-	S-	S-	S-	S-	S-
Clerical I (R-W) - 10 minutes	S-	S-	S-	S-	S-	S-
Clerical II (R-W) - 5 minutes		S-		S-	S-	I+
Clerical II (R-W) - 10 minutes		S-		S-	S-	S-

(Sample Sizes: 60 Whites; 31 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/60 = 0%	Less	0
Slope	39/60 = 65%	More	- (Smaller for Blacks)
Intercept	5/21 = 24%	More	+ (Blacks Overpredicted)

STUDY 14

O'Leary et al (13): Catalog Order Plant - Miscellaneous Clerical

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANCE RESULTS

Tests	Accuracy	Accuracy Under Pressure	Work Speed	Learning Ability	Human Relations	Overall Effectiveness
Verbal Reasoning						
Arithmetic Reasoning						
Clerical I - 5 minutes						
Clerical I - 10 minutes						
Clerical II - 5 minutes						
Clerical II - 10 minutes						
Clerical I (R-W) - 5 minutes						
Clerical I (R-W) - 10 minutes						
Clerical II (R-W) - 5 minutes						
Clerical II (R-W) - 10 minutes						

(Sample Sizes: 106 Whites; 24 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/60 = 0%	Less	0
Slope	0/60 = 0%	Less	0
Intercept	0/60 = 0%	Less	0

STUDY 15

Farr et al (4): Health Insurance (Miscellaneous Clerical)

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANCE RESULTS

Tests	Quantity of Work	Quality of Work	Accuracy	Knowledge of Job	Job Aptitude	Flex-ability	Overall Effectiveness
Verbal				SE-	I+		S-
Quantitative	I+	S-		S-	I+		
Total	S-	S-		SE-	I+		S-
Pictorial Reasoning	S-			SE-	I+		I+

(Sample Sizes: 157-158 Whites; 51 Blacks)

SUMMARY

	Percent Significant	More or Less than Chance	Direction
Standard Error	3/28 = 11%	More	-(Smaller for Whites)
Slope	7/25 = 28%	More	-(Smaller for Blacks)
Intercept	6/18 = 33%	More	+(Blacks Overpredicted)

STUDY 16

Farr et al (4): Health Insurance (Clerk, Clerk-Typist)

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Quantity of Work	Quality of Work	Accuracy	Knowledge of Job	Job Aptitude	Flex-ability	Overall Effectiveness
Verbal				SE-	I+		
Quantitative		S-		S-	I+		
Total		S-					S-
Pictorial Reasoning		S-		S-	I+		

(Sample Sizes: 95 Whites; 31 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	1/28 = 4%	Less	0
Slope	6/27 = 22%	More	-(Smaller for Blacks)
Intercept	3/21 = 14%	More	+(Blacks Overpredicted)

Tenopyr (17): Machine Shop Training

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Final Grade General	Final Grade Machine	Final Grade Total	Machine Rating - 8 Week	Performance Rating - 16 Week	Short Performance Rating - 16 Week
EAS 1 - Verbal Comprehension	I+	I+	I+	I+	I+	I+
EAS 2 - Numerical Ability	I+	SE-	I+	I+	I+	
EAS 5 - Space Visualization	I+	SE-	I+	I+	I+	I+

(Sample Sizes: 71 Whites; 44-63 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	2/18 = 11%	More	-(Smaller for Whites)
Slope	0/16 = 0%	Less	0
Intercept	15/16 = 94%	More	+(Blacks Overpredicted)

STUDY 18

Campbell et al (2): Medical Technicians

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Job Knowledge Test
Subtraction and Multiplication	I+
Vocabulary	I+
Hidden Figures	I+
Numerical Comprehension	S+
Necessary Arithmetic	I+
Gestalt Completion	I+
Picture Numbers	I+
Paper Folding	I+

(Sample Sizes: approximately 168 Whites; approximately 297 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/8 = 0%	Less	0
Slope	1/8 = 13%	More	+(Greater for Blacks)
Intercept	7/7 = 100%	More	+(Blacks Overpredicted)

STUDY 19

Grant and Bray (7): Telephone Installation and Repair

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Highest Level Passed
SCAT Total	
SCAT Quantitative	
SCAT Verbal	I+
Mechanical Comprehension	
BS Qualification Test III	I+
Abstract Reasoning	S-

(Sample Sizes: 219 Whites; 211 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/6 = 0%	Less	0
Slope	1/6 = 17%	More	-(Smaller for Blacks)
Intercept	2/5 = 40%	More	+(Blacks Overpredicted)

STUDY 20

Gael and Grant (5): Telephone Service Representative

HOMOGENEITY OF REGRESSION ANALYSIS - TABLE OF SIGNIFICANT RESULTS

Tests	Job Knowledge Review	Filing Errors*	Verbal Contact	Record Preparation	Composite Performance Index
Bell System Qualification Test - I		I+			
Arithmetic	I+	I+			I+
Number Comparison	I+	I+			I+
Filing	I+	I+			I+
Number Transcription	I+	I+			I+
Spelling	S-	S-			S-
Space Relations Aptitude	I+	I+			I+

* Reflected

(Sample Sizes: 184-193 Whites; 97-106 Blacks)

SUMMARY

	Percent Significant	More or Less than chance	Direction
Standard Error	0/35 = 0%	Less	0
Slope	3/35 = 9%	More	-(Smaller for Blacks)
Intercept	16/32 = 50%	More	+(Blacks Overpredicted)

Summary of Significance Tests of Differences in Regression Equations Between Whites and Blacks

Study	St. Err.	Slope	Intercept
1. Kirkpatrick et al (10) - Study 1 - Clerical - Insurance	0	0	-
2. Kirkpatrick et al (10) - Study 2 - Clerical - Insurance	+	-	+
3. Kirkpatrick et al (10) - Study 3 - Man. Occ. - Gen. Maint.	0	0	0
4. Kirkpatrick et al (10) - Study 3 - Man. Occ. - Heavy Load	0	0	0
5. Farr et al (3) - Toll Collectors	0	+	0
6. Farr et al (3) - Correctional Officers	0	-	0
7. Farr et al (3) - Toll Facility Officers	+	0	0
8. Farr et al (3) - Home Office Clerical	0	0	+
9. Farr et al (3) - Keypunch Operators	+	0	0
10. O'Leary et al (13) - Catalog Order Plant - Mat'l Hndlr I	0	0	0
11. O'Leary et al (13) - Catalog Order Plant - Mat'l Hndlr II	+	0	0
12. O'Leary et al (13) - Catalog Order Plant - Clerical I	+	0	0
13. O'Leary et al (13) - Catalog Order Plant - Machine Clerical	0	-	+
14. O'Leary et al (13) - Catalog Order Plant - Misc. Clerical	0	0	0
15. Farr et al (4) - Health Insurance (Misc. Clerical)	-	-	+
16. Farr et al (4) - Health Insurance (Clerk, Clerk-Typist)	0	-	+
17. Tenopyr (17) - Machine Shop	-	0	+
18. Campbell et al (2) - Medical Technicians	0	+	+
19. Grant and Bray (7) - Telephone Installation and Repair	0	-	+
20. Gael and Grant (5) - Telephone Service Representatives	0	-	+

	Expected Under Null Hypothesis	Obtained	
STANDARD ERROR			
(-) More than chance significant, SE smaller for whites	5	2	$\chi^2 = 2.70$, (df = 2)
(0) Fewer than chance significant	10	13	$.40 < p < .50$
(+) More than chance significant, SE greater for whites	5	5	
SLOPE			
(-) More than chance significant, slope smaller for blacks	5	7	$\chi^2 = 2.70$ (df = 2)
(0) Fewer than chance significant	10	11	$.40 < p < .50$
(+) More than chance significant, slope greater for blacks	5	2	
INTERCEPT			
(-) More than chance significant, unfair to blacks	5	1	$\chi^2 = 6.40$ (df = 2)
(0) Fewer than chance significant	10	10	$.02 < p < .05$
(+) More than chance significant, unfair to whites	5	9	