PB-261 700

Validation of Test 21 for Entry-Level D.C. Police Officers

Civil Service Commission, Washington, D C

Aug 76

BIBLIOGRAPHIC DATA SHEET	1. Report No. TM 76-7	2.	3. Recipient's Accession No.
4. Title and Subtitle Validation of Test 21 for Entry-Level D.C. Police Officers			5. Report Date August 1976 6.
7. Author(s) Leon J. Wetrogan and Cynthia H. Cundiff			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Personnel Research and Development Center U.S. Civil Service Commission 1900 E Street N.W., Room 3H22 Washington, D.C. 20415			10. Project/Task/Work Unit No. 11. Contract/Grant No.
12. Sponsoring Organization Name and Address Same as box 9			13. Type of Report & Period Covered 14.

15. Supplementary Notes

16. Abstracts

This study was conducted to lend further support to the validity of Test 21, an 80 item verbal ability test for selecting entry level patrolmen in the Washington, D.C. Police Department. Test 21 scores, training performance measures and job performance measures were obtained on 263 recruits entering the D.C. Police training academy from December, 1973 to May, 1975. Test 21 was found to be significantly related to final grade average in training for the total sample and for the black and white subgroups separately. Further, the test was significantly correlated with weeks of training for the total sample and for the black sample. The results also indicated that Test 21 was significantly related to each of the cognitive dimensions included in the supervisor ratings of job performance. Tests for the significance of the difference between correlations indicated that the Test was equally valid for blacks and whites for both the training criteria and job performance criteria.

17. Key Words and Document Analysis. 17a. Descriptors

Public Personnel Testing
Test Validation
Test Statistics

17b. Identifiers/Open-Ended Terms

17c. COSATI Field Group

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 23
Unlimited	20. Security Class (This Page) UNCLASSIFIED	22. Price $4.00 -3.00

FORM NTIS-35 (REV. 10-73) ENDORSED BY ANSI AND UNESCO

Personnel Research and Development Center

Technical Memorandum 76-7

PB 261 700

Validation of Test 21 for Entry-Level D.C. Police Officers

United States
Civil Service Commission
Bureau of Policies and Standards

Technical Memorandum 76-7

VALIDATION OF TEST 21 FOR ENTRY LEVEL D. C. POLICE OFFICERS

Leon I. Wetrogan and Cynthia H. Cundiff

Applied Psychology Section
Personnel Research and Development Center
U. S. Civil Service Commission
Washington, D. C.
August 1976

VALIDATION OF TEST 21 FOR ENTRY LEVEL
D. C. POLICE OFFICERS

ABSTRACT

This study was conducted to lend further support to the validity of Test 21, an 80 item verbal ability test for selecting entry level patrolmen in the Washington, D.C. Police Department. Test 21 scores, training performance measures and job performance measures were obtained on 263 recruits entering the D.C. Police training academy from December, 1973 to May, 1975. Of the 263 patrolmen, 121 were white, 141 were black and 1 was Spanish-surnamed. The training criteria consisted of the number of weeks it took each recruit to complete the training program and final grade average in training. The job performance criterion consisted of the operational supervisor ratings used by the D.C. Police Department. Test 21 was found to be significantly related to final grade average in training for the total sample and for the black and white subgroups separately. Further, the test was significantly correlated with weeks of training for the total sample and for the black sample. The results also indicated that Test 21 was significantly related to each of the cognitive dimensions included in the supervisor ratings of job performance. Tests for the significance of the difference between correlations indicated that the Test was equally valid for blacks and whites for both the training criteria and job performance criteria.

CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

VALIDATION OF TEST 21 FOR ENTRY LEVEL
D. C. POLICE OFFICERS

Test 21 is an 80-item verbal ability test consisting of the following item types: a. vocabulary; b. reading; c. analogies; d. proverbs; and e. judgment. It has been used as a selection device for D.C. Police since 1948. The purpose of the test is to measure the ability of the applicant to complete the Police Academy training course. Forty correct answers are required for passing.

Previous Studies

In recent years several studies have been conducted to evaluate the effectiveness of Test 21. Weatherford (Note 1) found that means for alternate forms of the test showed little change from 1948 through early 1967, and that as the level of education increased so did the probability of an applicant's passing Test 21. She found the test to be sensitive not only to level of education, but also to quality of education. Competitors educated in southern and D.C. inner city high schools had a higher failure rate than those educated in Washington suburban and northeastern schools. However, most people passing Test 21 were able to complete the Police Academy training course.

Futransky (Note 2), in the second phase of the above study, found Test 21 to be a valid predictor of performance in recruit school, but found the relationship between Test 21 and performance ratings by supervisors to be somewhat less impressive. Racial data in this study indicated a low positive relationship for whites between recruit school performance and on the job supervisory ratings, however, even this small relationship was non-existent for black officers. He recommended that the "...Police Department should scrutinize the training program and its supervisory evaluation methods and precepts to be sure that its training program is relevant to police work."

Wilson (Note 3) found that Test 21 continued to be an effective predictor of performance in recruit school for both blacks and whites. Test 21 correlated .52 with final grade average in the Police Academy for the total group. The correlation for blacks, as a group, was not significantly different from that of whites.

At this time Wilson also explored the effectiveness of other tests. An analysis of the curriculum of the academy identified knowledges, skills, and abilities required for successful completion of the course. These included reading comprehension, writing ability and problem-solving ability. A factor analysis revealed that the vocabulary, reading comprehension, and proverb meaning items in Test 21 had very high loadings on the verbal factor, but none of the question types tested nonverbal reasoning, used in problem-solving. Wilson hypothesized that the test would be an even more effective selection device if items measuring nonverbal reasoning were included. Letter series and simple number series test items were found to measure this factor effectively. Wilson concluded that the selection test would be improved by the addition of these item types.

Between the 1967 and 1974 studies several changes were made in the recruit training program. The length of the training program was increased from twelve to seventeen weeks, recruits were required to take nine instead of eight courses and the criterion for passing was raised from a final grade average of 70 percent to 75 percent.

Training Program - Past and Present

In 1970 the training program was 17 weeks in length. During this period recruits were given instruction in nine subjects, including the following: a. Police Regulations; b. Police Manual; c. Traffic Regulations; d. Alcoholic Beverage Control; e. First Aid; f. Laws of Arrest; g. Rules of Evidence; h. District of Columbia Code; and i. Juvenile Delinquency.

To graduate from the Academy each recruit was required to satisfy the following criteria (Wilson, Note 3): a. a final exam score of at least 75.0 in each of the courses. Recruits were allowed to repeat the course exams a limited number of times in order to meet the criterion if necessary; b. a final grade average of at least 75.0. This was the arithmetic mean of the first final exam

scores in each of the nine subjects.

Since 1970 drastic changes have been made in the Police Academy training program. In 1972 the D. C. Police Department contracted Educational Systems for the Future of Langley Park, Maryland to develop and implement a training program for entry-level policemen, "...relevant to their performance on the street," (Educational Systems for the Future, Note 4). To accomplish this a task analysis was conducted employing subject matter experts (SME's) to define the tasks and subtasks making up terminal performance objectives (TPO's). The result was "...sequential listing of tasks and subtasks required for mastery of a TPO," (Educational Systems for the Future, 1972, p. 2) and the development of the Basic Officer Training System, the modular systems approach to learning now used at the Academy to train recruits.

The system is divided into four levels which are composed of thirteen major performance areas: a. Police Manual; b. Self Defense; c. Community Relations; d. Interviews; e. Vehicle Skills; f. First Aid; g. Arrest Procedures; h. Communications; i. Firearms; j. Court Procedures; k. Patrol Techniques; l. Water Safety; and m. Police Assistance. (See Appendix A.) These areas are further divided into modules, of which there are approximately one hundred. Criterion-referenced tests are given in each of the modular areas. The test items are constructed so that they correspond to a specific task or subtask at the modular level.

In order to complete the training course successfully the recruit must attain a grade average of at least 90% on the modular exams in each major performance area. Modular tests may be retaken to attain the 90% criterion, as long as the recruit maintains a major area average of at least 90%. The Department refers to this process as enrichment. On the other hand, if a recruit's area average falls below 90%, he runs the risk of being terminated from the Academy unless he retakes the modular exam and passess. The latter process is termed recycling. If a recruit fails a recycling of any one modular area or three major performance areas, he is recommended for termination from the Academy. It now takes recruits an average of 14 weeks to complete the program.

The mechanics of the system provide for continued evaluation and revision within the program. Little use is made of lectures. Learning aids consist of workbooks; skills laboratories taking the form of role-playing sessions and micro teaching experience, where the recruit's behaviors are video taped for individual critique; instructor demonstration and instruction, and audio visual aids. There are also on-the-job-training assignments in the field.

The system allows the recruit to plan his course of study with the aid of an instructor-manager in accordance with his learning needs, the recommended sequence of study as derived from the learning hierarchy, and the availability of materials. As the system now stands recruits begin their training at individual times and work through the areas at their own pace.

Purpose of the Study

The Basic Officer Training System became operational in late 1973. Since its implementation there have been no studies of the training program and how it related to performance on Test 21. Thus, it is the purpose of this study to investigate the effectiveness of Test 21 as a predictor of training and job performance criteria. To accomplish this task, the study specifically examined the following:

1. the relation between scores on Test 21 and time required to complete training;

2. the relation between Test 21 scores and final grade average at the Academy;

3. the relation between Test 21 scores and job performance ratings;

4. the relation between time to complete training and job performance;

5. the relation between training grade average and job performance.

Method

Subjects

Subjects for this study were 263 police recruits entering training from December, 1973 through May, 1975. Of this number there were 121 whites, 141 blacks and 1 Spanish-surnamed. Out of the population of 263 officers, 96 had been on duty long enough to receive their first administrative performance evaluation. Of this group, 50 were black and 45 were white. The mean age of the total group was 21.92 years. There was no significant difference between the mean age of the black and white officers. All officers had completed at least high school or its equivalent.

Predictor

Test 21, an 80-item verbal abilities test, was used as the predictor. Test scores were taken from personnel files of individual recruits where possible. However, in several cases it was necessary to obtain test scores from the Civil Service Commission's Washington Area Office.

Criteria

There were two major criteria in this study.

1. Supervisory Performance Ratings. The form used for this is the standard performance evaluation form used by the Department (See Appendix B for sample). Officers are rated on ten dimensions along a ten point scale ranging from unsatisfactory to outstanding. They are also given an overall evaluation in which each officer is compared only with all officers of the same grade. Where a folder contained more than one rating, the rating coinciding with the time nearest completion of the one year trial period was used.

2. Performance in Training represented by:

a. Time to completion of training or academic weeks. Because the expected time for completing the training program varied with when the recruits entered the program, each trainee's number of academic weeks were converted to standard scores according to the particular program in which he was enrolled. This variation in expected weeks was the result of content changes in the program.

b. Final grade average. Final grade average represented the average of final exam scores in each of the areas. For approximately 95% of the sample this figure represented scores achieved the first time the recruits attempted the exams.

Results

Correlation of Test 21 with Training Criteria

The means and standard deviations for Test 21, grade average and number of academic weeks for the two racial groups and the total sample are presented in Table 1. While the data on Test 21 and grade average reflect raw score means and standard deviations, the data on academic weeks represent the mean and standard deviation of standard scores.

Table 2 shows the validity coefficients for Test 21 for blacks, whites, and the total sample with the training measures. The data indicate that Test 21 is significantly correlated with grade average for the black, white, and total sample. Also, the Test is significantly correlated with number of academic weeks for the black and the total sample. The negative correlations between Test 21 and academic weeks suggest that higher test scores are associated with fewer weeks to the completion of training. Significance tests for the differences in correlations between the white and black samples indicated that the validities were not significantly different ($p>.05$).

In order to determine if the same regression line could be used to describe the predictor-criterion relationship for the two racial subgroups, the Gulliksen-Wilks test for differences between pairs of regression lines was performed. The Gulliksen-Wilks involves three successive tests: a. for differences in scatter or dispersion around the regression line (standard error of estimate); b. if test a is not significant, then for differences in slopes of the regression lines; and c. if test b is also not significant, for differences in intercept or level

TABLE 1

Means and Standard Deviations for Test 21, Grade Average in Training and Academic Weeks to Completion of Training for the Two Racial Groups and Total Sample

Racial Group	Test 21		Grade Average		Academic Weeks*	
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
White	84.13	7.66	94.64	1.63	-.34	.81
Black	76.13	5.92	93.53	1.61	.26	1.05
Total	79.85	7.85	94.03	1.68	0.00	1.00

*Data are in standard score form (z-scores)

TABLE 2

Correlation Coefficients for Test 21 for Whites, Blacks and Total Sample with Training Criteria

	Whites	Blacks	Total
Academic Weeks	(N=120) -.01	(N=141) -.17*	(N=262) -.26***
Grade Average	(N=121) .36***	(N=141) .45***	(N=263) .49***

* p<.05
** p<.01
*** p<.005
one tailed test
Note. The total sample includes one Spanish-surnamed.

of regression line. Figures 1 and 2 illustrate the white and black regression lines for the test with the two training criteria. In all figures the solid line represents the blacks and the dotted line the whites. Table 3 presents the chi squares and degrees of freedom for the Gulliksen-Wilks regression tests between Test 21 and the two training measures. The results indicated that the chi square for the standard error of estimate was significant for the relationship between Test 21 and Academic Weeks. However, when grade average was used as the criterion, the chi squares for errors, slopes, and intercepts were non-significant.

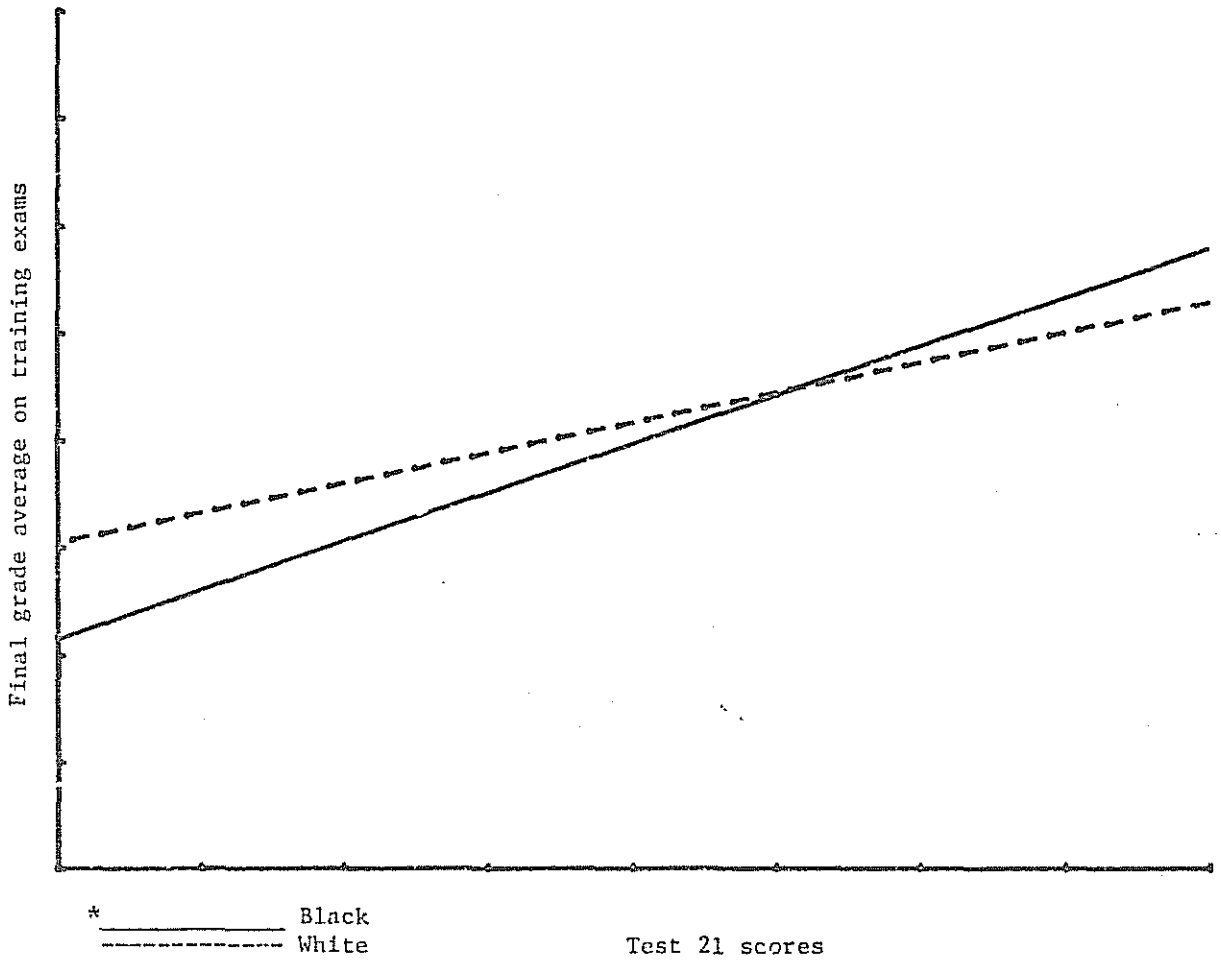

Figure 1.* Regression lines of Test 21 on grade average for black and white officers.

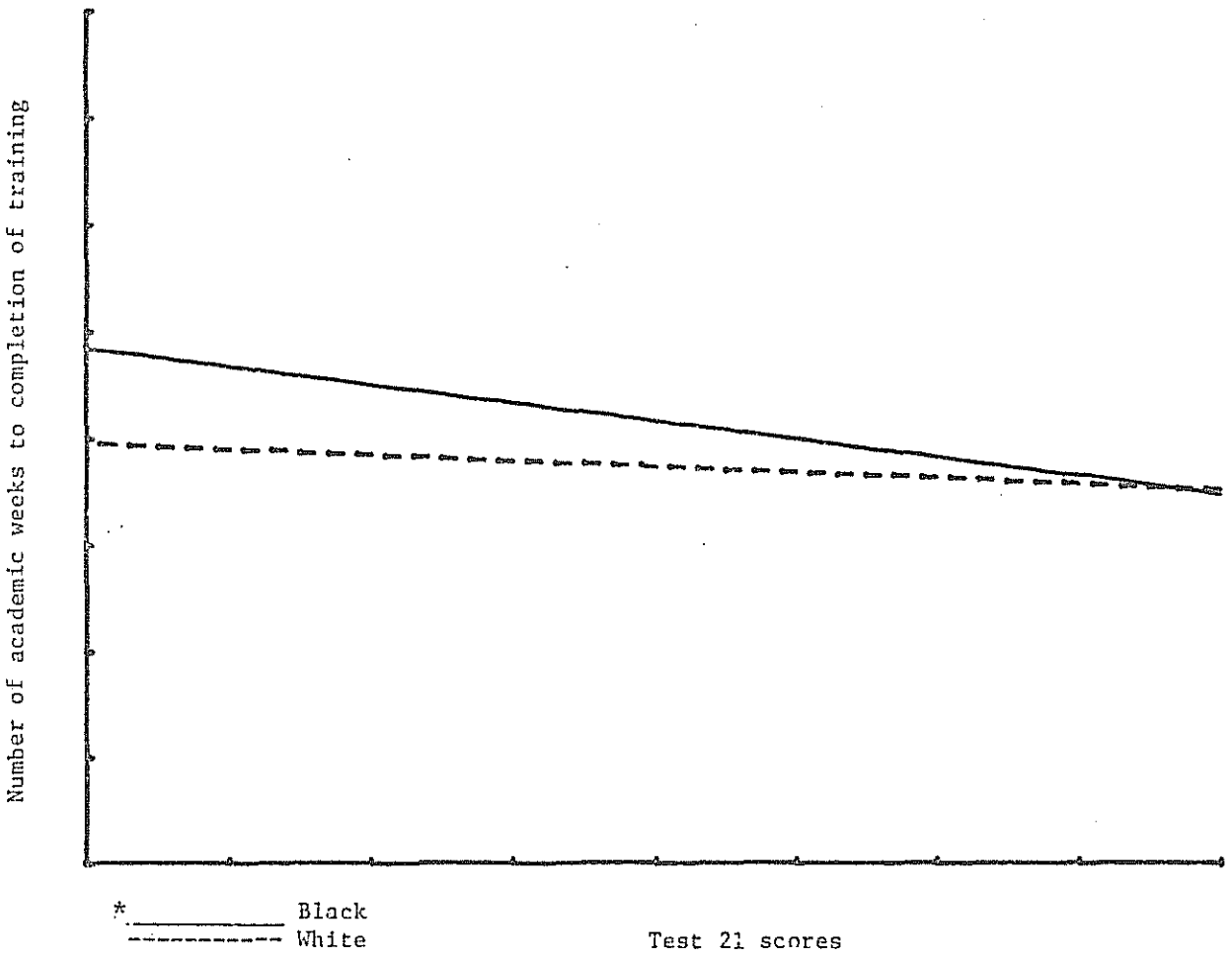

Figure 2.* Regression lines of standard scores for academic weeks on Test 21 for black and white officers.

TABLE 3

Gulliksen-Wilks Regression Analysis for Equality of Standard Error, Slopes and Intercepts for Whites and Blacks for Test 21 with Training Criteria

Training Measure	X^2	df
Academic Weeks	7.47***	1
Error		
Slopes		
Intercepts		
Grade Average		
Error	0.46	1
Slopes	2.98	1
Intercepts	2.61	1

*** $p<.01$

Correlation of Test 21 with on the Job Performance Criteria

Table 4 shows the means and standard deviations for the supervisor ratings of job performance for the whites, blacks and total sample. Scores on performance dimensions one through nine represent a rating on each of the nine single scales. Scores on performance dimension 10 represent the sum of each officer's ratings on dimensions one through nine. Scores on performance dimension eleven represent the sum of ratings on the major cognitive dimensions reflecting what the test was intended to predict. The four dimensions included in performance dimension 11 were: 3. Learning Ability, 4. Knowledge and Skill, 6. Written Communication and 7. Oral Communication. These four are predominantly cognitive in nature.

Prior to correlating Test 21 with the performance dimensions, the internal consistency reliabilities of the composite criteria, performance dimensions 10 and 11 were computed. The reliabilities were determined using the Kuder-Richardson 20 (KR-20) formula. This formula assumes that: a. all items in a scale measure the same factor; b. the correlations between all items in a scale are equal; and c. the variances of items in a scale are equal. The KR-20 reliability estimates for performance dimensions 10 and 11 were .90 and .93.

Table 5 presents the validity coefficients for Test 21 with the relevant job performance ratings for whites, blacks and the total sample. The data indicate that Test 21 correlated significantly with 4. Knowledge and Skill performance dimension for the white sample. For the black sample Test 21 correlated significantly with: 3. Learning Ability, 4. Knowledge and Skill, 9. Overall Performance and composite criteria, 10 and 11. In the case of the total sample Test 21 correlated significantly with all of the performance dimensions except 8. Performance of Duty.

TABLE 4

Means and Standard Deviations for Job Performance Ratings for Whites, Blacks and Total Sample

	Performance Dimension	Whites (N=45)		Blacks (N=50)		Total (N=96)	
		$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
1	Bearing and Behavior	6.00	0.88	5.76	0.94	5.88	0.91
2	Human Relations	5.84	0.77	5.78	0.76	5.81	0.76
3	Learning Ability	6.04	0.77	5.04	1.21	5.53	1.1[illegible]
4	Knowledge and Skill	5.27	0.72	4.74	0.87	5.00	0.85
5	Acceptance of Responsibility	5.98	0.84	5.24	1.00	5.59	0.99
6	Written Communication	5.64	0,86	5.00	0.95	5.31	0.95
7	Oral Communication	5.64	0.74	5.18	0.77	5.41	0.79
8	Performance of Duty	5.95	0.77	5.20	1.07	5.56	1.00
9	Overall Performance	5.82	0.65	5.20	0.90	5.50	0.85
10	Sum of Dimensions (1-8)	46.38	4.41	41.94	5.98	44.09	5.71
11	Sum of Dimensions (3), (4), (6) and (7)	22.60	2.51	19.96	3.28	21.25	3.22

Note. The total sample includes one Spanish-surnamed.

TABLE 5

Correlation Coefficients for Test 21 and Relevant Job Performance Ratings for the White, Black and Total Sample

	Performance Dimension	Whites (N=45)	Blacks (N=50)	Total (N=96)
3	Learning Ability	.110	.293*	.375***
4	Knowledge and Skill	.265*	.369**	.407***
6	Written Communication	.042	.179	.256**
7	Oral Communication	.221	.227	.329***
9	Overall Performance	.178	.275*	.362***
10	Sum of Dimensions (1-8)	.195	.239*	.362***
11	Sum of Dimensions (3), (4), (6), (7)	.189	.312*	.397***

* $p<.05$
** $p<.01$
*** $p<.005$

Note. Differences between the correlations for the two racial groups were nonsignificant for all performance dimensions, $p \geq .05$.

In order to test for differential validity, the significance of the differences between correlations for the two racial subgroups was computed. The results indicated that for each of the performance dimensions the correlations with the test were not significantly different, $p > .05$.

Figures 3 through 9 present the regression lines for whites and blacks on each of the relevant performance dimensions. Table 6 shows the chi squares for the Gulliksen-Wilks regression tests between Test 21 and the relevant performance dimensions. The table indicates that the chi squares for the Standard Errors of Estimate were significant when the criteria were: 3. Learning Ability, 9. Overall Performance and 10. the composite of all nine dimensions. Differences in the intercepts of the regression lines were significant for the racial subgroups only when the criteria were: 6. Written Communication and 11. the composite of dimensions 3, 4, 6, and 7.

Correlation of Training Measures with Supervisory Ratings of Job Performance

Table 7 shows the correlations between the two training measures and the performance dimensions for the whites, blacks and total sample. The data point out that academic weeks correlated significantly with: 1. Bearing and Behavior, 4. Knowledge and Skill, 5. Acceptance of Responsbility, 7. Oral Communication and the two composite performance dimensions for the white sample. In the case of the blacks, Academic Weeks correlated significantly with all of the performance dimensions except 2. Human Relations and 5. Acceptance of Responsibility. For the total sample, academic weeks correlated significantly with all performance dimensions except 2. Human Relations.

The pattern of correlations between Grade Average and the performance dimensions were slightly different from that between academic weeks and the performance dimensions. For the white sample, grade average correlated significantly only with 1. Bearing and Behavior whereas for the black sample, grade average correlated significantly with 3. Learning Ability; 7. Oral Communication; 8. Performance of Duty; 9. Overall Performance; and 10. and 11. the two composite criteria. In the case of the total sample, grade average correlated significantly with all of the performance dimensions except 2. Human Relations.

In order to determine if significant differences existed in the relationship between performance in training and performance on the job for the two racial subgroups, the differences between the correlations were tested. The data point out that all differences of correlations between the training measures and the job performance dimensions for the two racial subgroups were nonsignificant, $p > .05$.

Testing for Fairness of Test 21

While significant differences in slopes suggest differential validity, differences in intercepts indicate potential fairness. This raised the question of the effect of using a common regression line to predict criterion performance for the two racial subgroups versus using separate regression lines developed to minimize the error of prediction for each subgroup. In order to test for unfairness, the mean test scores for the two racial subgroups were used in the common regression equation to predict mean criterion performance for each subgroup. This analysis was performed for predicting each criterion in which the Gulliksen-Wilks analysis yielded significant differences in intercepts. It should be pointed out that differences in intercepts may have been present where the standard errors were significant. However, due to the sequential nature of the Gulliksen-Wilks test, such differences could not be examined where the errors were found to be significantly different.

Table 8 presents the actual and predicted mean criterion scores for the blacks and whites for Written Communication and for the Sum of Dimensions 3, 4, 6 and 7 performance ratings. For the black sample, the actual and predicted mean criterion scores on Written Communication were 5.00 and 5.20, while on the sum of the four dimensions the means were 19.96 and 20.64, respectively. The actual and predicted mean criterion scores for the whites on Written Communication were 5.64 and 5.45, while on the sum of the four dimensions the means were 22.60 and 21.92, respectively.

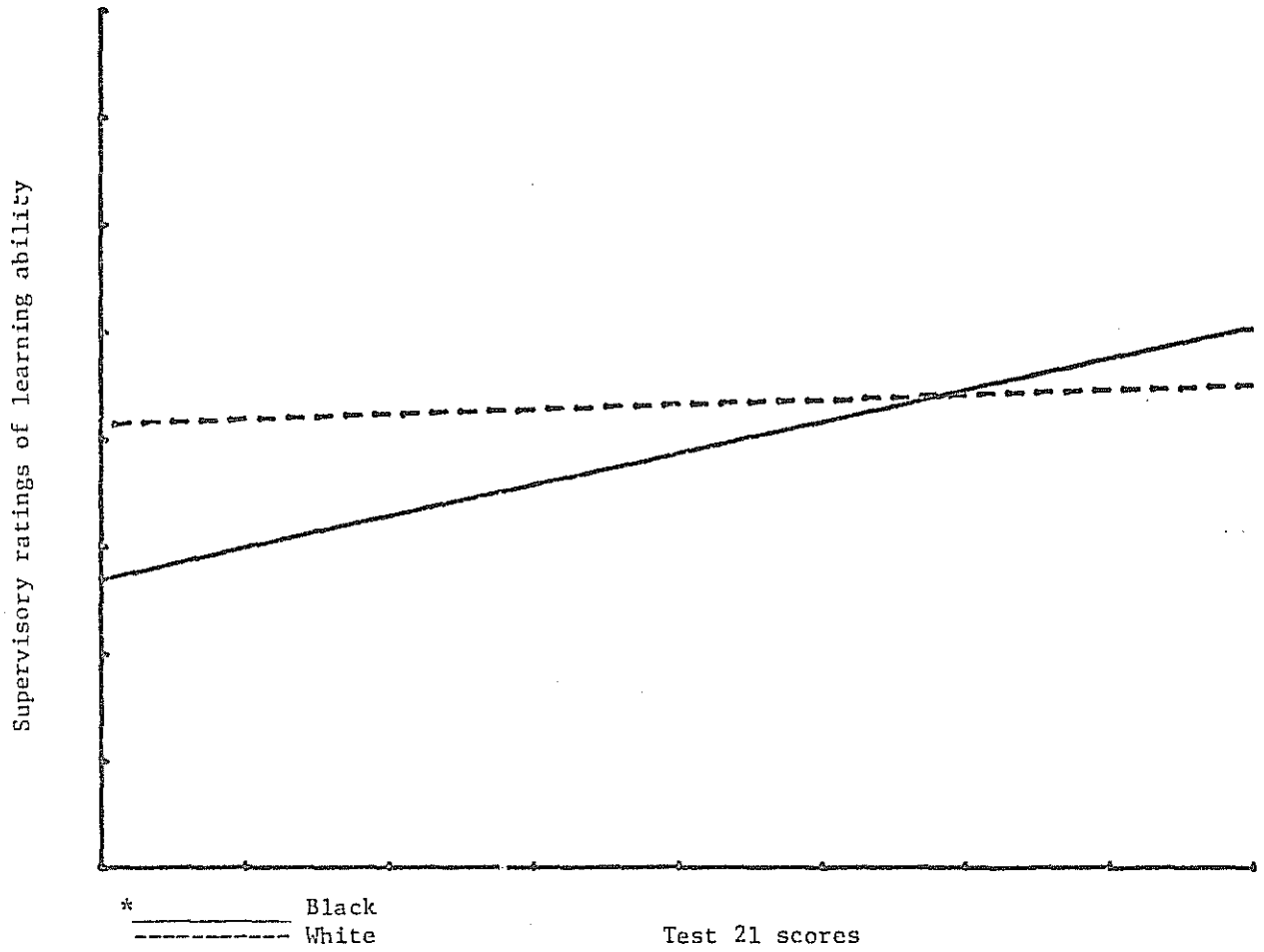

Figure 3.* Regression lines of Test 21 on learning ability for black and white officers.

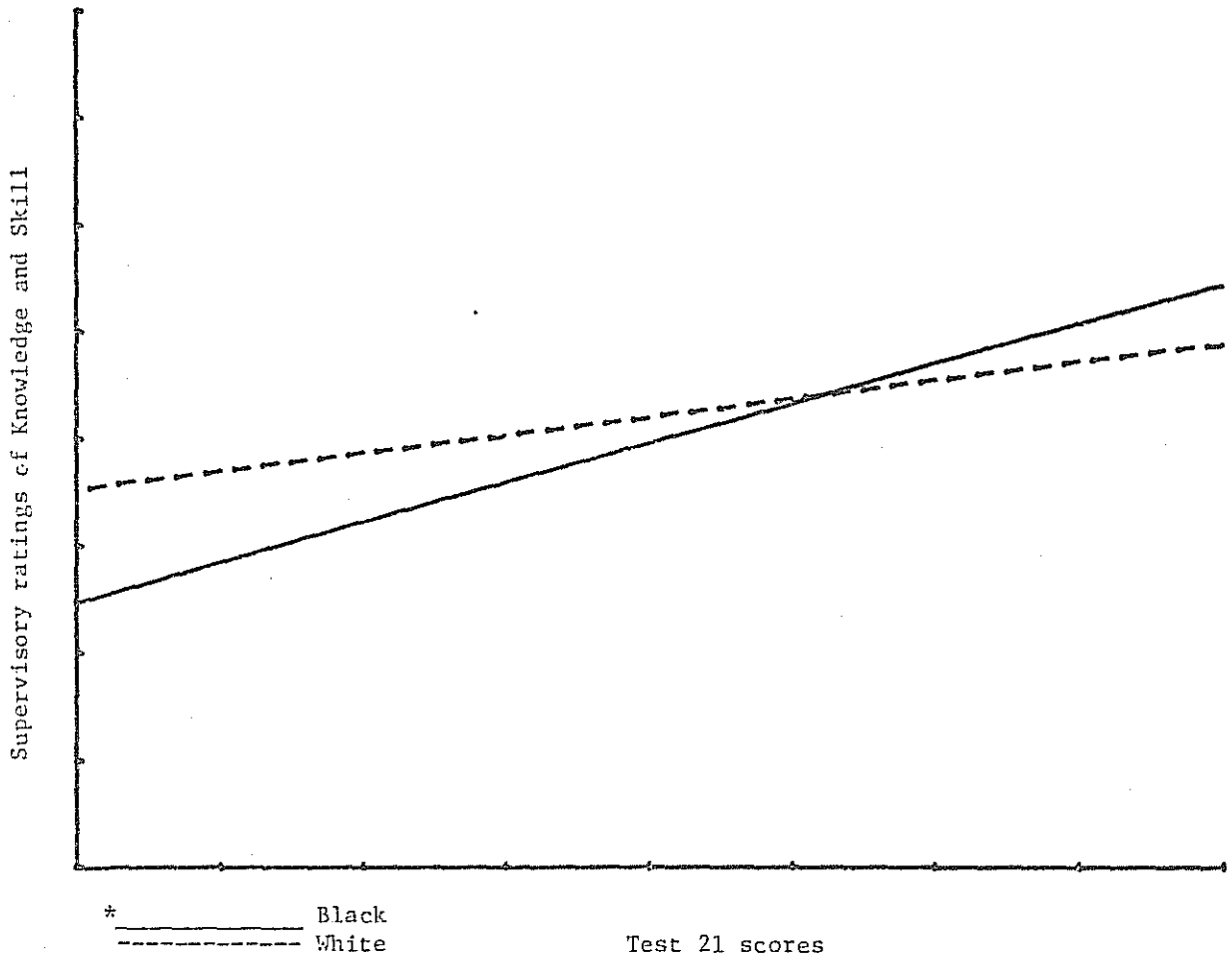

Figure 4.* Regression lines of Test 21 on knowledge and skill for black and white officers.

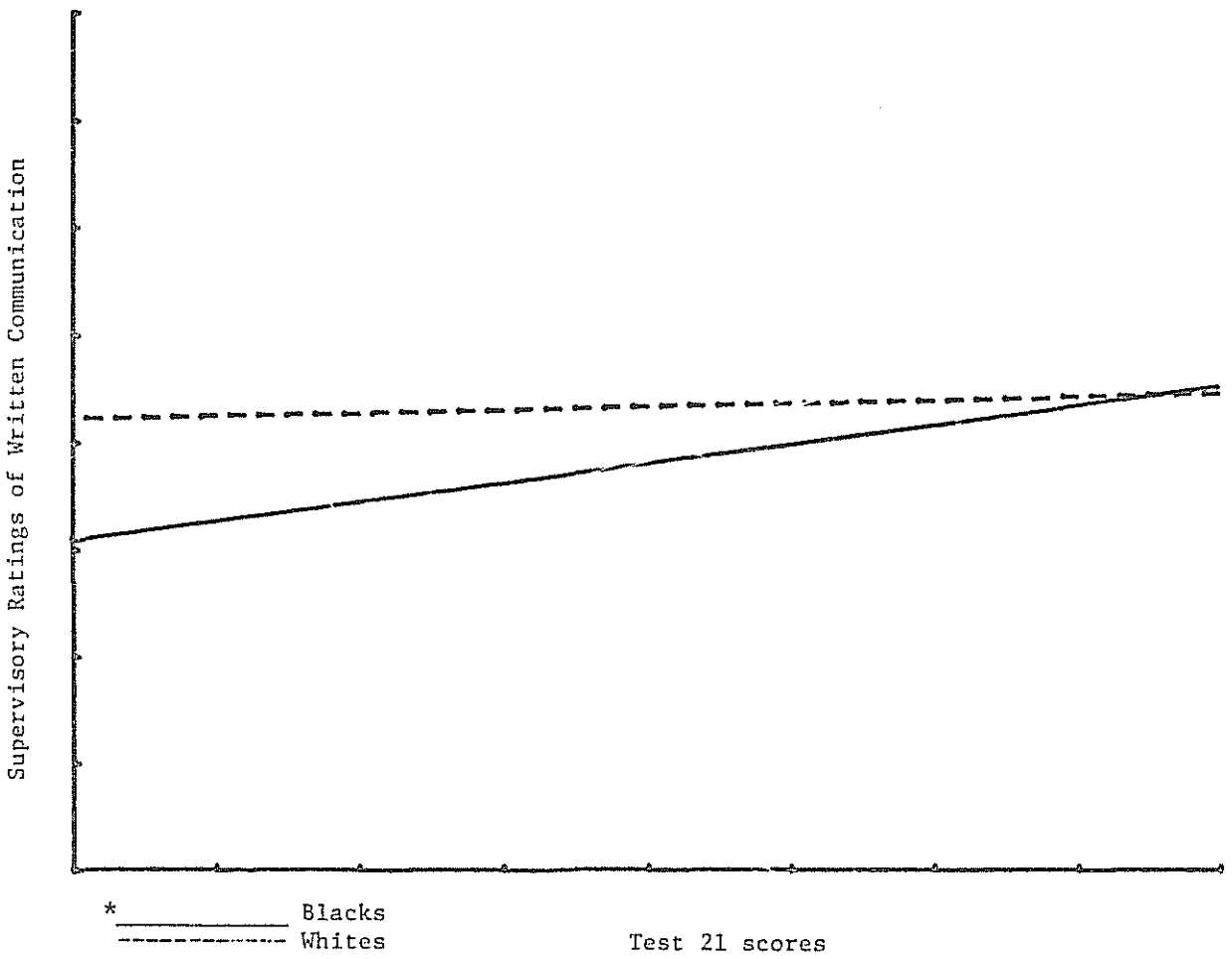

Figure 5.* Regression lines of Test 21 on written communication for black and white officers.

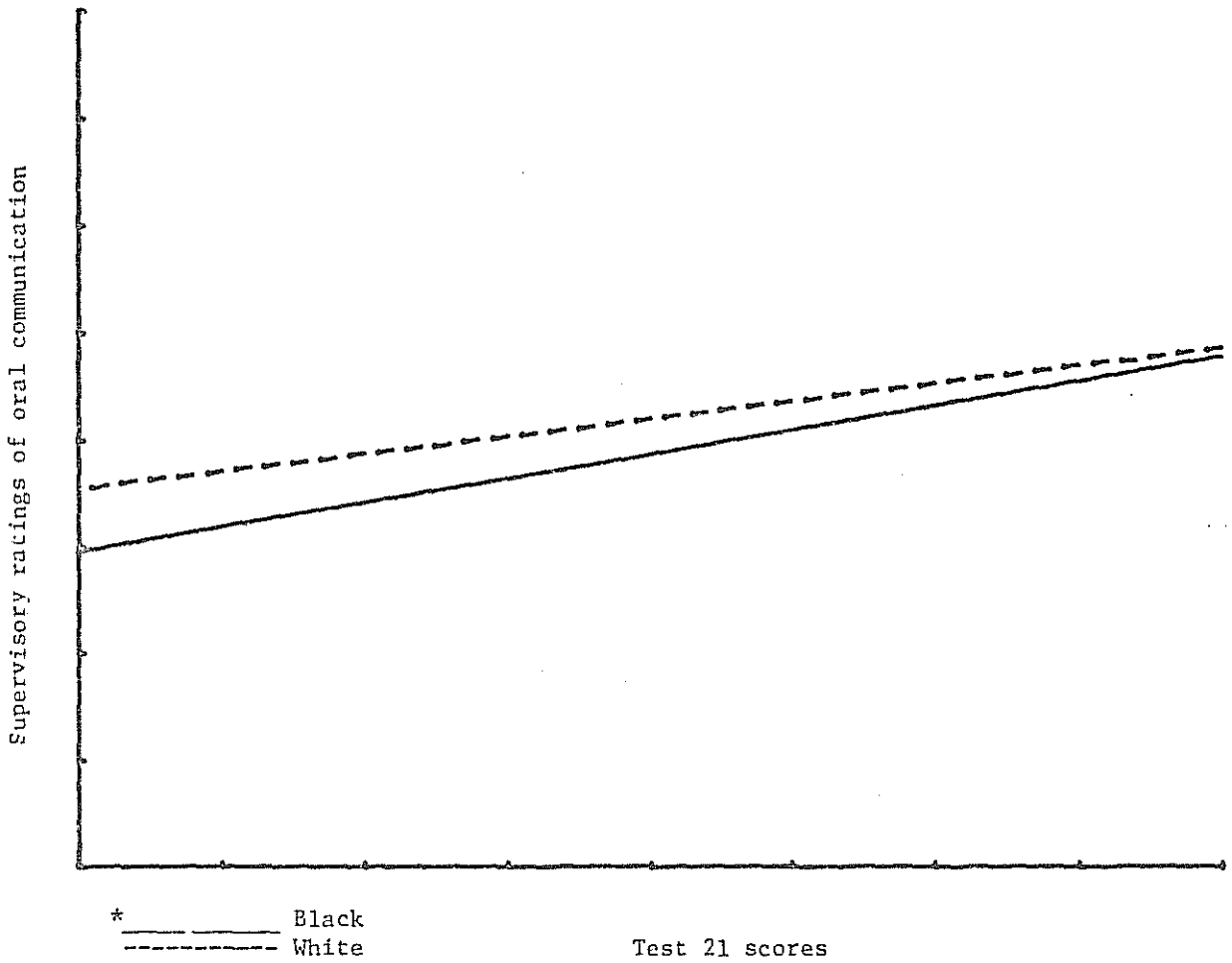

Figure 6.* Regression lines of Test 21 on oral communication for black and white officers.

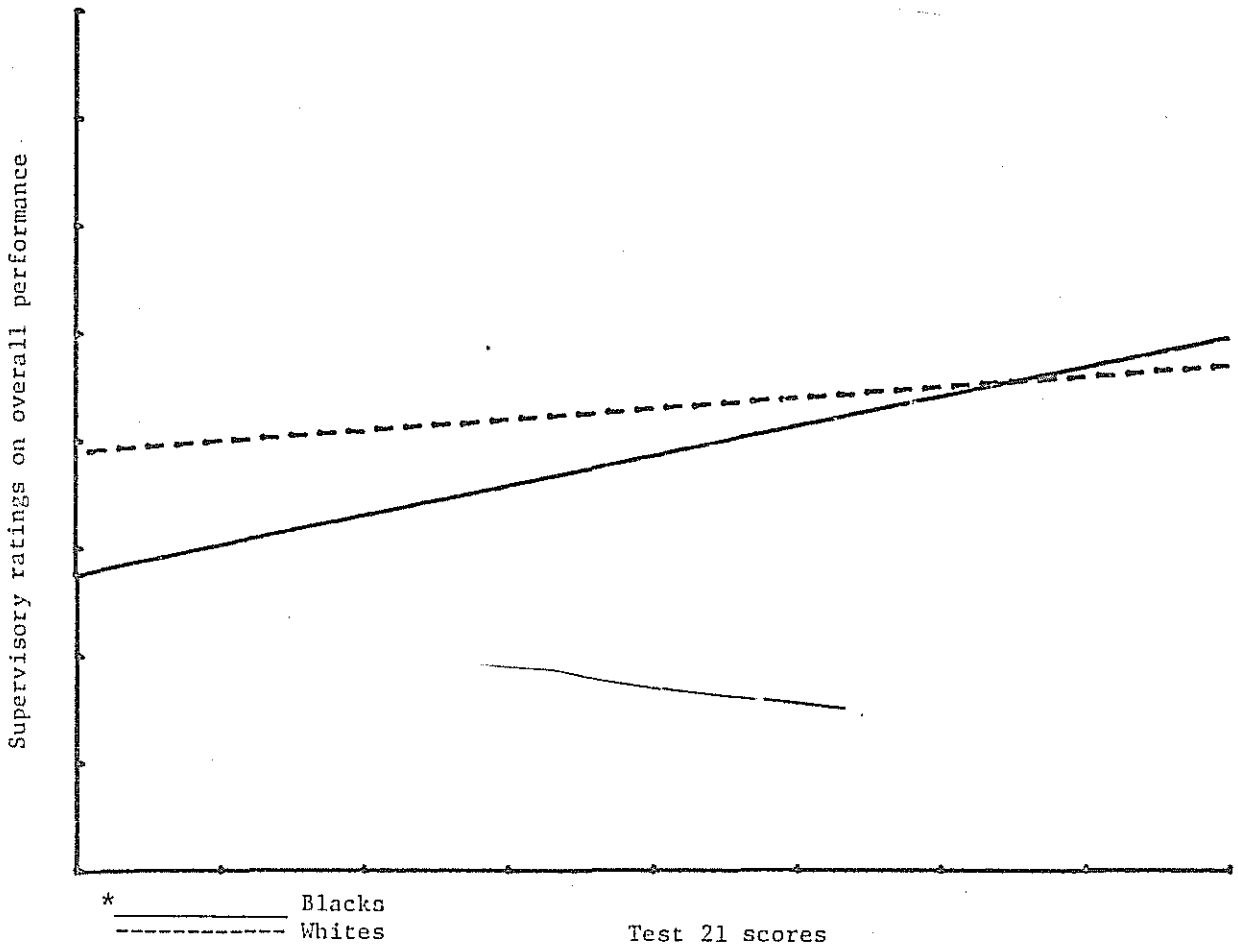

Figure 7.* Regression lines of Test 21 with overall performance for black and white officers.

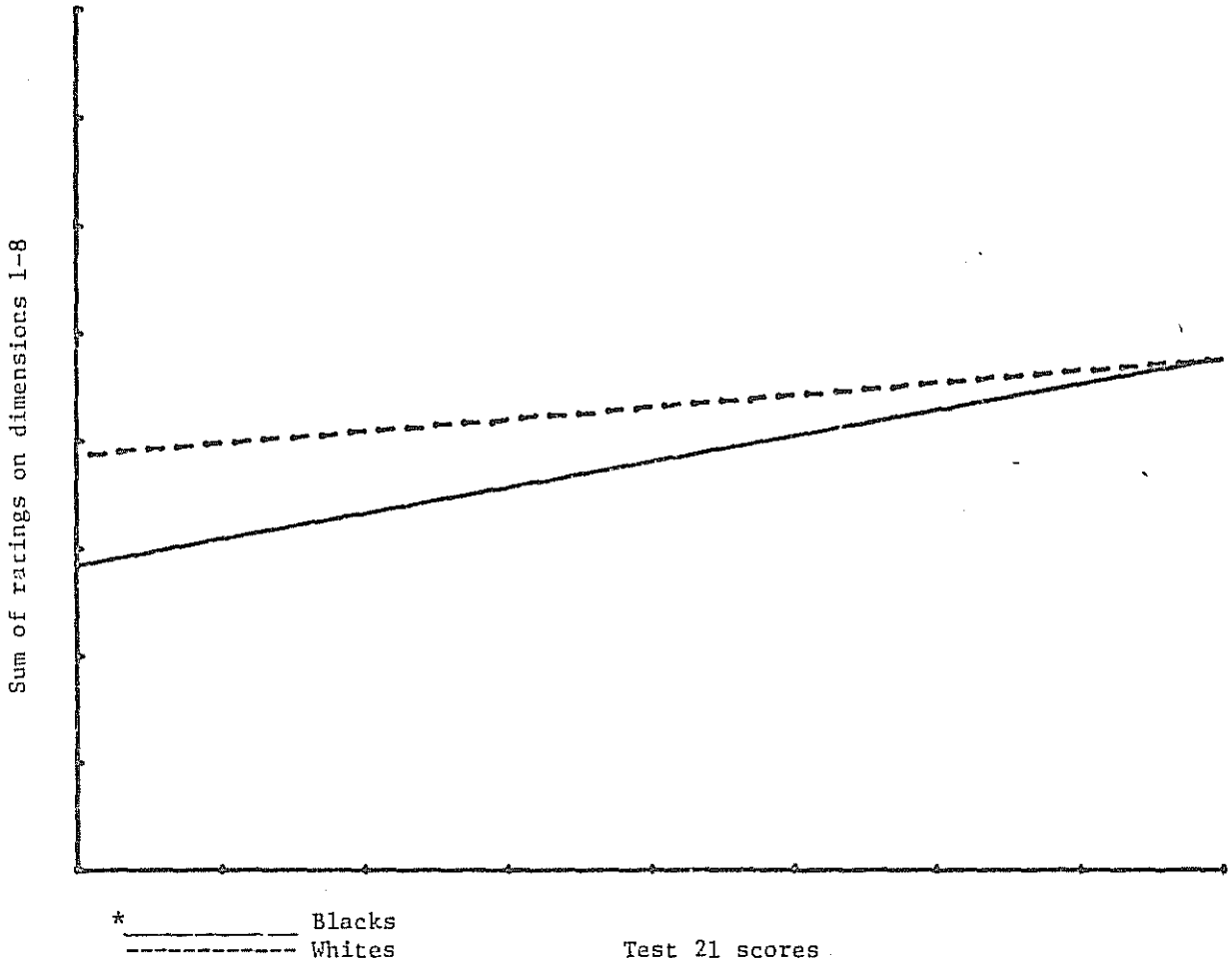

Figure 8.* Regression lines of Test 21 with sum of dimensions (1-8) for black and white officers.

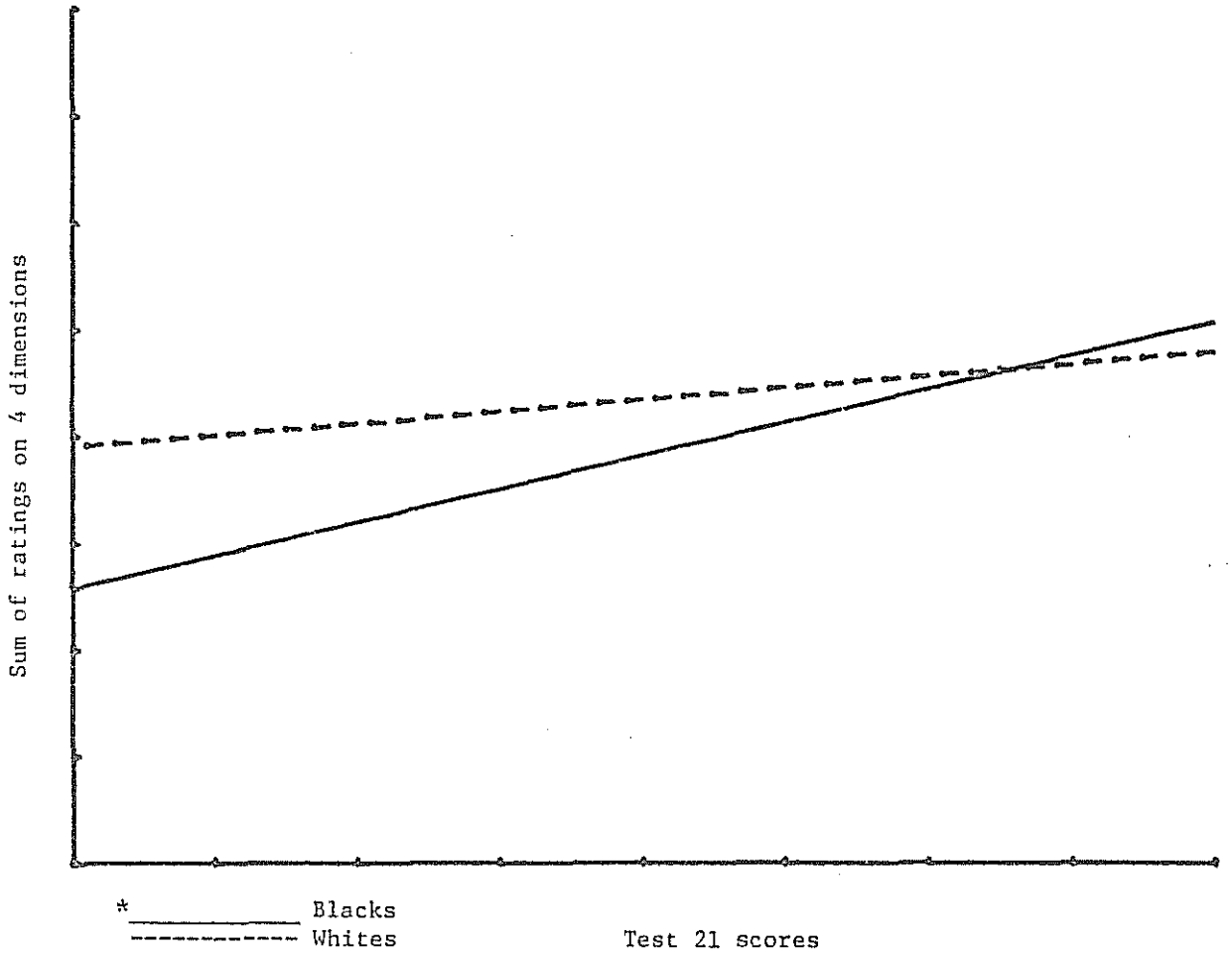

Figure 9.* Regression lines of Test 21 with sum of dimensions 3, 4, 6, 7 for black and white officers.

TABLE 6

Gulliksen-Wilks Regression Analysis for Equality of Standard Errors, Slopes and Intercepts for Whites and Blacks for Test 21 with Performance Ratings

Performance Dimension	X^2	df
Learning Ability		
Error	7.93*	1
Slope		
Intercept		
Knowledge and Skill		
Error	1.19	1
Slope	1.60	1
Intercept	1.66	1
Written Communication		
Error	.34	1
Slope	.75	1
Intercept	5.97*	
Oral Communication		
Error	.10	1
Slope	.14	1
Intercept	2.26	1
Overall Performance		
Error	4.22*	1
Slope		
Intercept		
Sum of Dimensions (1-8)		
Error	4.03*	1
Slope		
Intercept		
Sum of Dimensions (3), (4), (6), (7)		
Error	2.59	1
Slope	1.62	1
Intercept	6.79*	1

* $p < .05$

TABLE 7

Correlations between Training Measures and Supervisory Ratings
for the White, Black and Total Sample

Performance Dimensions	Academic Weeks			Grade Average		
	Whites N=45	Blacks N=50	Total N=96	Whites N=45	Blacks N=50	Total N=96
Bearing Behavior	-.254*	-.265*	-.284***	.268*	.214	.271***
Human Relations	-.083	-.076	-.085	.101	.085	.103
Learning Ability	-.196	-.384***	-.421***	.102	.298*	.349***
Knowledge and Skill	-.378**	-.237*	-.351***	.198	.229	.311***
Acceptance of Responsibility	-.260*	-.012	-.218*	-.001	.137	.212*
Written Communication	-.179	-.279*	-.327***	.198	.047	.246**
Oral Communication	-.248*	-.273*	-.331***	.148	.271*	.301***
Performance of Duty	-.220	-.249*	-.340***	.059	.245*	.287***
Overall Performance	-.216	-.352**	-.396***	.228	.293*	.368***
Sum of Dimensions (1-8)	-.325*	-.290*	-.391***	.195	.247*	.342***
Sum of Dimensions (3), (4), (6), (7)	-.303*	-.350***	-.420***	.200	.248*	.352***

* $p < .05$
** $p < .01$
*** $p < .005$

Note. Differences between all of the correlations for blacks and whites was nonsignificant.

In order to determine if the differences between actual and predicted mean criterion scores on the two dimensions for the two racial subgroups were significant, t-tests for the differences between means of correlated samples were computed. The t-values for the black sample on the sum of four dimensions and Written Communication were 1.57 and 1.54 respectively, which with df=49 were nonsignificant, $p > .05$. For the white sample, the t-values for the differences between actual and predicted mean criterion scores for the sum of four dimensions and Written Communication were 1.77 and 1.45 respectively, which with df=44 were nonsignificant, $p > .05$.

TABLE 8

Actual and Predicted Mean Criterion Scores for the Black and White Samples Using Test 21 as the Predictor and Using Regression Equations Developed for the Combined Sample

	Black		White	
	Actual	Predicted	Actual	Predicted
Written Communication	5.00	5.20	5.64	5.45
Sum of Dimensions 3, 4, 6 and 7	19.96	20.64	22.60	21.92

Discussion

The validity of an instrument, written test, performance test, interview, etc., is the degree to which the instrument measures what it is intended to measure. In the selection situation, validity is the extent to which the selection instruments measure those knowledges, skills and/or abilities (KSA's) necessary for successful performance in training and/or on the job.

In order not to leave any doubt about the usefulness of Test 21, three sets of relationships were explored. The first involved a determination of the relationship between Test 21 and success in training. The second concerned the validity of Test 21 with on-the-job performance measures as criteria. The third involved the relationship between performance in training and performance on the job.

The Validity of Test 21 For Predicting Training Performance

The relationship between Test 21 and grade average was significant and appears to be quite strong for these types of measures. Not only is the test valid for the total sample but it is valid for both the white and black groups separately. Further, the regression analysis suggests that a common regression line can be used to describe the predictor-criterion relationship for both subgroups. Consequently, candidates can be ranked and compared without any adjustments in the scoring for the separate racial groups.

The correlation between Test 21 and academic weeks was also significant. The data point out that the Test is valid for the total sample and for the Black sample. The correlation between scores on Test 21

and number of academic weeks for the white sample was not significant.

However, since the differences in correlations between racial subgroups were not significant it can be concluded that the test is equally valid for both groups. Inspection of the regression analysis data points up that the standard errors were significantly different, suggesting that the validities may be increased by using separate regression equations for the two racial subgroups.

One possible reason for the substantially lower correlations of the test with academic weeks than with grade average relates to what is being measured in the criterion. In the case of grade average the criterion seems to be mainly measuring performance of those skills or knowledges needed for job success. However, the variance in academic weeks may not only be due to the KSA's needed for success but also to level of motivation. Consequently different applicants with the same test score might complete the training program in considerably different amounts of time due to motivational differences. This would serve to attenuate the validity of the test since it was developed primarily as a cognitive measure rather than as a measure of characteristics such as motivation.

The Validity of Test 21 For Predicting Job Performance Ratings

It was pointed out earlier that the validity of an instrument is the extent to which the instrument measures what it was intended to measure. In the case of Test 21 the abilities measured are predominately cognitive in nature. However, the job performance ratings include cognitive as well as noncognitive characteristics. The data indicate that Test 21 is a valid predictor of each of the cognitive criteria. Also, the validities were not significantly different between the racial subgroups. This would suggest that the test is equally valid as a predictor of job performance for the two racial groups. While the test does not exhibit differential validity, it would appear on the basis of the regression analysis that, except for the job performance dimensions of Knowledge and Skill and Oral Communication, a common regression line may be inappropriate. However, when a common regression line was used to predict mean criterion scores for the two performance dimensions with significant intercept differences, Written Communication and the sum of four dimensions, there was no systematic overprediction or under-prediction for either racial subgroup. This result lends support to the fairness of Test 21, and the use of a common regression line for both racial groups.

For the whites, the meaning of the regression analysis is somewhat uncertain, since the correlations between Test 21 and both Written Communication and the Sum of the Four Dimensions were nonsignificant for that subgroup. This would indicate that for the whites, the regression lines describing the predictor-criterion relationships may be unstable and therefore, the intercepts for the white regression lines may be unreliable. However, since the correlations between the predictor, Test 21, and the criteria were significant for the combined sample and for the black sample, it seems reasonable to conclude that the common regression equations accurately predict ratings on Written Communication and the Sum of the Four Dimensions for the blacks.

If it were the case that the regression line for the white sample actually varied to the extent that it was significantly different from the common regression line, the common regression line would then underpredict white performance. This would make it necessary to add points to the test scores of whites in order to be fair. The net effect might be that more whites and fewer blacks would be hired as police officers.

The Relationship Between Training Performance and Job Performance

One reason that training performance is frequently not accepted as a relevant criterion for predicting job performance is that in most studies, data are not presented to indicate the relationship between training performance and job success. In the present study, such a link was demonstrated. The training measures were found to be significantly related to all but one dimension of job performance - human relations. This consistent pattern of relationships clearly demonstrates that training performance is closely linked to success on the job.