DOCUMENT RESUME

ED 056 081 TM 000 885

AUTHOR Flaugher, Ronald L.
TITLE Minority Versus Majority Group Performance on an Aptitude Test Battery. Project Access Research Report No. 3.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO CEEB-RDR-71-72-1; RB-71-48
PUB DATE Aug 71
NOTE 23p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Academic Performance; *Aptitude Tests; Caucasians; *Cognitive Ability; *Comparative Analysis; *Group Intelligence Tests; Mathematics; Mexican Americans; *Minority Groups; Negroes; Sex Differences; *Testing Problems; Test Reliability; Test Validity; Verbal Ability
IDENTIFIERS Orientals; *Project Access

ABSTRACT

Black students from two cities and Mexican-American students, both male and female, showed small but consistent tendencies to perform better, relative to White groups, on three nontraditional measures: tests of inductive reasoning, spatial scanning, and associative memory. These measures showed somewhat less discrepancy between the groups than did tests of the more traditional verbal and mathematical aptitudes. The results seemed compatible with the suggestions of the Commission on Tests to expand the number of measures included in traditional testing programs. Other measures of these same aptitudes, and their validities, remain to be explored in this context.

COLLEGE ENTRANCE EXAMINATION BOARD
RESEARCH AND DEVELOPMENT REPORTS
RDR-71-72, NO. 1

RESEARCH BULLETIN
RB-71-48 AUGUST 1971

Project Access Research Report No. 3
Minority Versus Majority Group Performance on an Aptitude Test Battery

Ronald L. Flaugher

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY
BERKELEY, CALIFORNIA

PROJECT ACCESS RESEARCH REPORT #3

Minority Versus Majority Group Performance on an Aptitude Test Battery

Ronald L. Flaugher

Abstract

Black students from two cities and Mexican-American students, both male and female, showed small but consistent tendencies to perform better, relative to White groups, on three nontraditional measures: tests of inductive reasoning, spatial scanning, and associative memory. These measures showed somewhat less discrepancy between the groups than did tests of the more traditional verbal and mathematical aptitudes.

The results seemed compatible with the suggestions of the Commission on Tests to expand the number of measures included in traditional testing programs. Other measures of these same aptitudes, and their validities, remain to be explored in this context.

PROJECT ACCESS RESEARCH REPORT #3

Minority Versus Majority Group Performance on an Aptitude Test Battery[1]

Ronald L. Flaugher

The content and structure of large-scale assessment programs have been determined, like many other aspects of our society, by the services and benefits that accrue to the white majority population of the United States. It is frequently the case that the decisions which are made to maximize this benefit either ignore or actually conflict with the wishes and needs of ethnic minority groups. The purpose of the present study was to examine a multimeasure aptitude test battery for suggestions of revisions that would counteract this effect.

In an earlier report (Flaugher, 1971) data from this same test battery were compared to those from previous studies by Lesser, Fifer, and Clark (1965) and Stodolsky and Lesser (1967). In those studies, each ethnic group was treated as if it were equal in size to the other groups, which yielded interesting results, but was not representative of the numerical imbalances in the society. For this reason, the present analysis was conducted, in which each minority group was compared individually with data on the available samples of the white population. It is the differences in performance which

---

[1]Project Access Research Reports: #1, Reactions to a very difficult test by an inner-city high school population: A test and item analysis, by Ronald L. Flaugher and Lewis W. Pike, College Board Research and Development Reports 69-70, No. 4, and ETS Research Memorandum 70-11, June 1970; #2, Patterns of test performance by high school students of four ethnic identities, by Ronald L. Flaugher, College Board Research and Development Reports 70-71, No. 9, and ETS Research Bulletin 71-25, May 1971.

appear between any particular minority group and the majority that will likely be the significant ones for that minority and that will have the greatest implications for the appropriateness of particular assessment techniques.

Project Access, through which these data were collected, and the specific content of the test battery are described in detail in Appendix I.

Procedure

The analysis utilized here is a duplication of that employed in the previous study of these data (Flaugher, 1971), with the exception that just two rather than four groups are included, such that each minority group is compared only with the majority group. This method adjusts the two group means as if the average of those two means were 50 and the total standard deviation were 10. The advantage of this approach is that it permits a graphical presentation of group mean differences on a variety of tests [illegible] presentation which is independent [illegible] ues, and thus permits comparison of relative performance from test to test. The groups are once again treated as if they were of equal size, but the single comparisons of each minority group with the majority group is a more meaningful treatment than in the previous study.

Data from two cities were available for this analysis: Memphis, Tennessee and Los Angeles, California. Analyses were conducted separately by sex, and in each case the comparisons were made with the White group of the appropriate sex. Three minority ethnic groups were available in Los Angeles: Black, Mexican-American, and Oriental; in Memphis, just Black students were available in sufficient numbers to be compared meaningfully

to the White students of that city. This resulted, then, in eight independent comparisons with the appropriate White samples: (1) Black Los Angeles males, (2) Black Los Angeles females, (3) Black Memphis males, (4) Black Memphis females, (5) Mexican-American Los Angeles males, (6) Mexican-American Los Angeles females, (7) Oriental Los Angeles males, and (8) Oriental Los Angeles females.

The sizes of the groups utilized in this analysis are presented in Table 1.

Results

The data from the nine tests in the battery are presented here, grouped to conform to the presentation in the previous studies. Four of the tests, those which best represent the four aptitudes of Verbal, Reasoning, Mathematics, and Space, are presented together and in that order; the results from the remaining five tests are presented separately. The numerical values for all of the following graphs are presented in tabular form in Appendix II.

Figure 1 presents the comparison of Black students with White students on the first four tests, for both male and female students, in two cities. Thus, there are four independent comparisons of test performance represented on each of the four kinds of test material.

In order to assure proper interpretation of this graph and the ones to follow, it should be understood that there is a "mirror image" nature about the values represented in the graphed lines. That is, each pair of points indicating mean values on a particular test (e.g., Black and White Males, or Black and White Females) must necessarily have an average, between the two of them, of 50, by virtue of the conversion technique.

Table 1

Options and Responses for Students on the Item "How Do You Describe Yourself?"

	Male	Female	Total[a]
Los Angeles			
Black, Afro-American, Negro	1211	1923	3164
White	155	151	312
Mexican-American, Brown	512	562	1094
Oriental	207	200	411
Memphis			
Black, Afro-American, Negro	1120	1614	2773
White	864	950	1844

[a]Total equals number of males plus number of females plus number of students who did not indicate sex. A total of 196 students indicated ethnic categories other than those shown here, and 2861 students chose either to omit this item or to give multiple responses; these were not included.

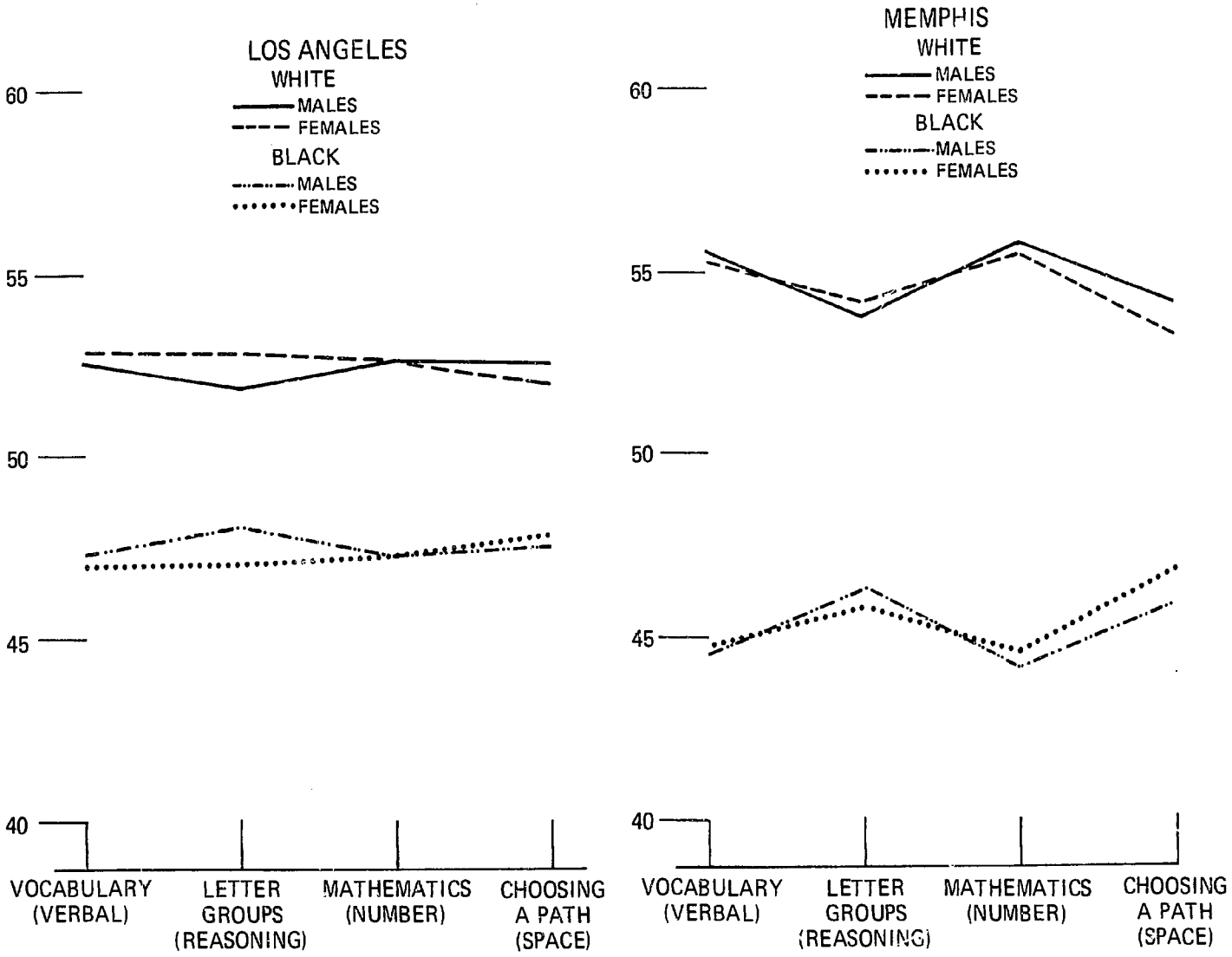

Figure 1. Contrasting patterns of performance on four tests, by sex, between Black and White students from two cities.

In other words, in Figure 1 the distance above the midpoint of 50 for the White group is necessarily matched by the distance below that midpoint for the Black groups.

With this understanding, then, interpretations can be made of the graphs. It can be seen that there are both similarities and differences between the left and right sides of the figure. The similarities are that the two graphs are rather more "flattened" than was the case when four ethnic groups, rather than two, were included (see Flaugher, 1971), but the present two graphs show a trend toward less discrepancy on the reasoning and spatial tests than on the verbal and mathematical tests.

The main difference between the two graphs is perhaps the wider separation between the Black and White groups from Memphis than those from Los Angeles. Rather than indicating that the respective students are somehow "better" or "worse" in one city than in another, it should be noted that this difference in separation is almost certainly due to the kind of selection for participation in Project Access that occurred in the two cities. It has been well documented that socioeconomic status is related to general level of test performance, and the students who participated in Project Access from the two cities differed a great deal on that variable. In Los Angeles, a selected group of inner city schools, characterized by low socioeconomic status, were participants in the study. In Memphis, however, virtually the entire school system participated, including a large number of White students from high socioeconomic backgrounds. The effect on the test scores was to create a wider separation between groups in Memphis than in Los Angeles, as shown in Figure 1.

In Figure 2, the data are presented from the sample of Mexican-American students in Los Angeles. On the left side of the figure is the graphing of the first four tests, and to the right are the results from the additional five tests. Within the graph of the first four tests, once again the two least discrepant tests are the reasoning and the spatial tests for the comparison of the male groups; for the female groups the reasoning test does not show the lessened discrepancy, but the spatial test does, which is similar to the pattern for the Black females of Los Angeles shown in Figure 1.

The additional five tests, graphed in the right side of Figure 2, show the Mexican-American students, both male and female, doing relatively better on Figure Analogies and Picture Number (associative memory) and, for males only, on Sentences.

In Figure 3, the results from both Memphis and Los Angeles Black students are presented for the additional five tests. This combining of results from the two cities into one graph was possible because of the differences between the cities in the socioeconomic level of the participants, as mentioned earlier, and the resultant differences in mean scores between Black and White groups. The Memphis results show little difference for any of the tests, except a dramatic reduction in differences for the Picture Number test. In the Los Angeles data there are similar results for the Picture Number test, and there is a smaller difference for the Black males on Reading and Sentences.

Finally, for completeness, the Oriental-with-White comparisons are graphed in Figure 4. The interpretation of these differences are a different matter, since it is the Oriental students who obtained higher

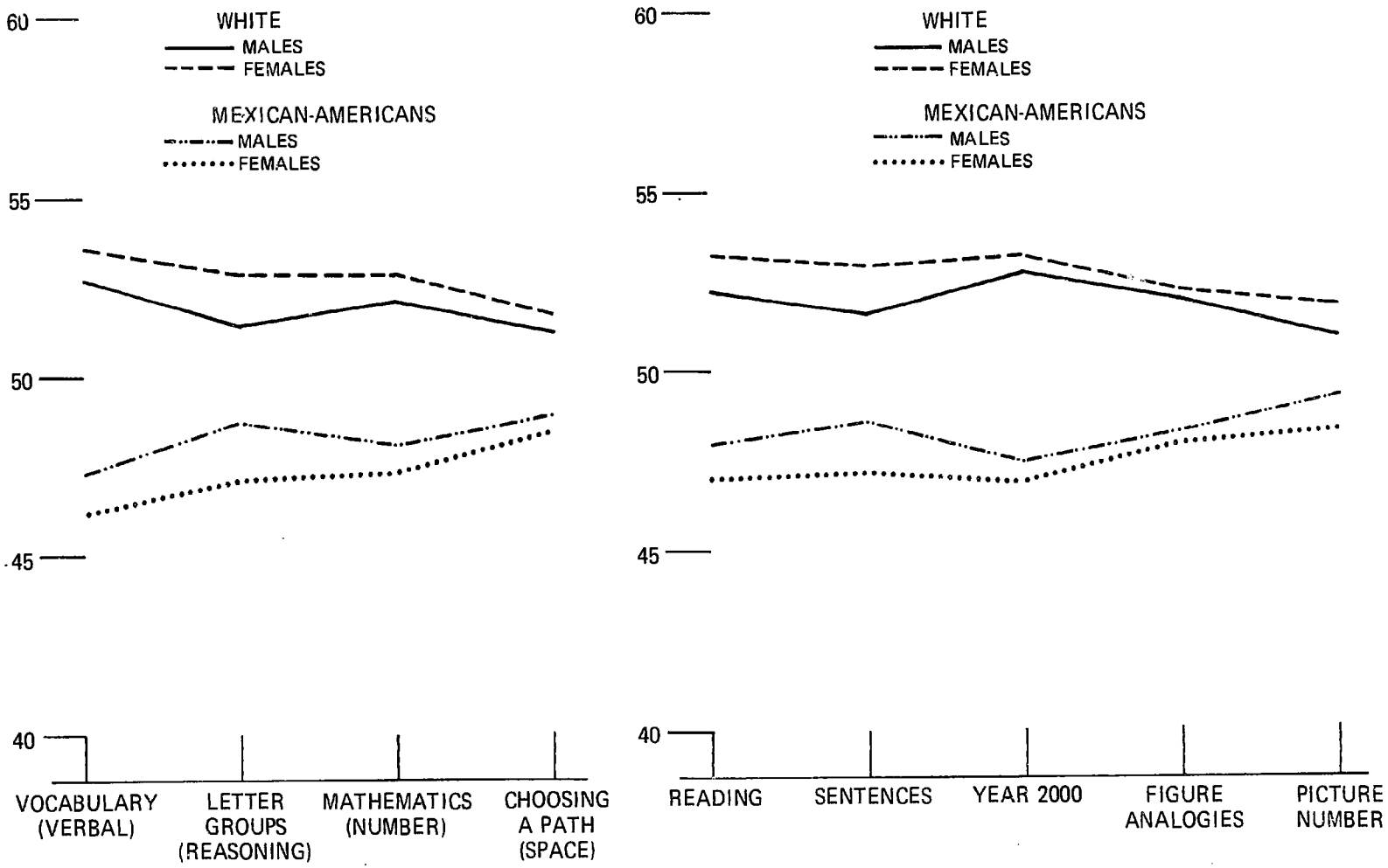

Figure 2. Contrasting patterns of performance on four tests, and five "additional" tests, by sex, between Mexican-American and White students.

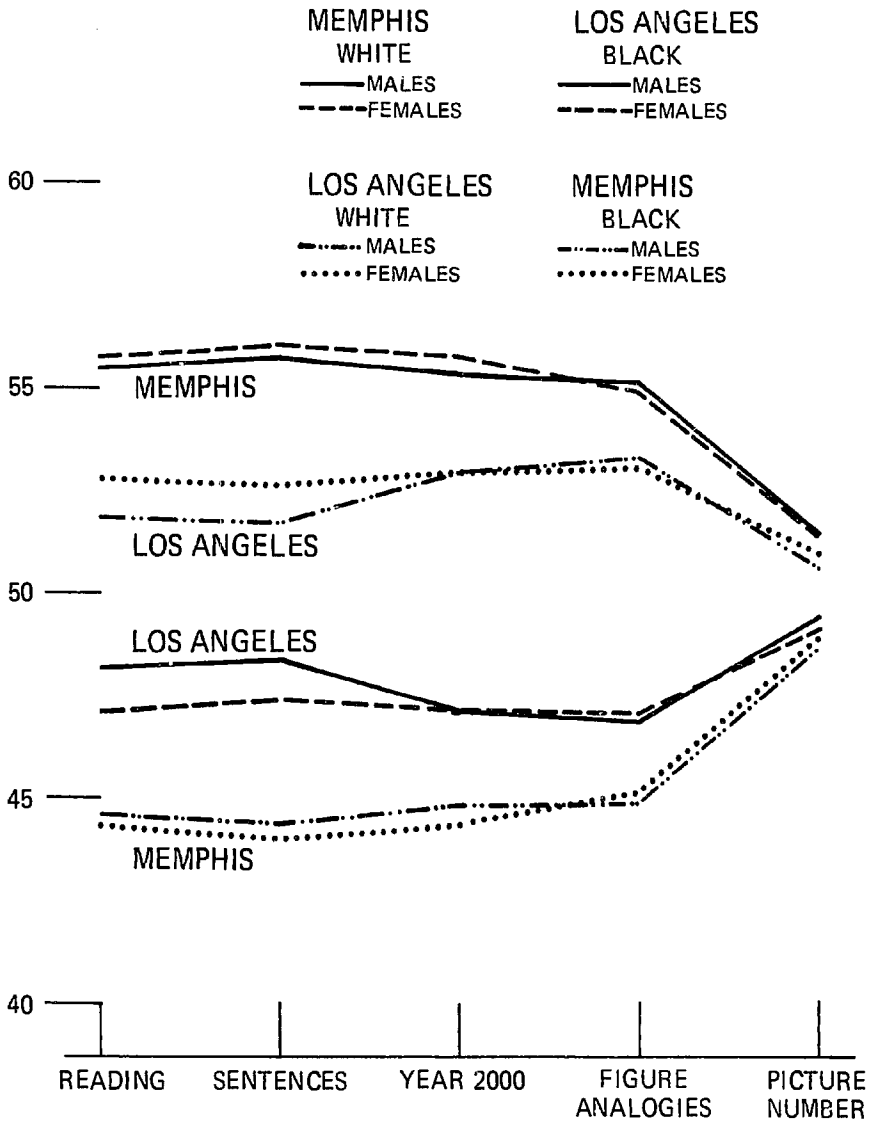

Figure 3. Contrasting patterns for five "additional" tests, by sex, between Black and White students from two cities.

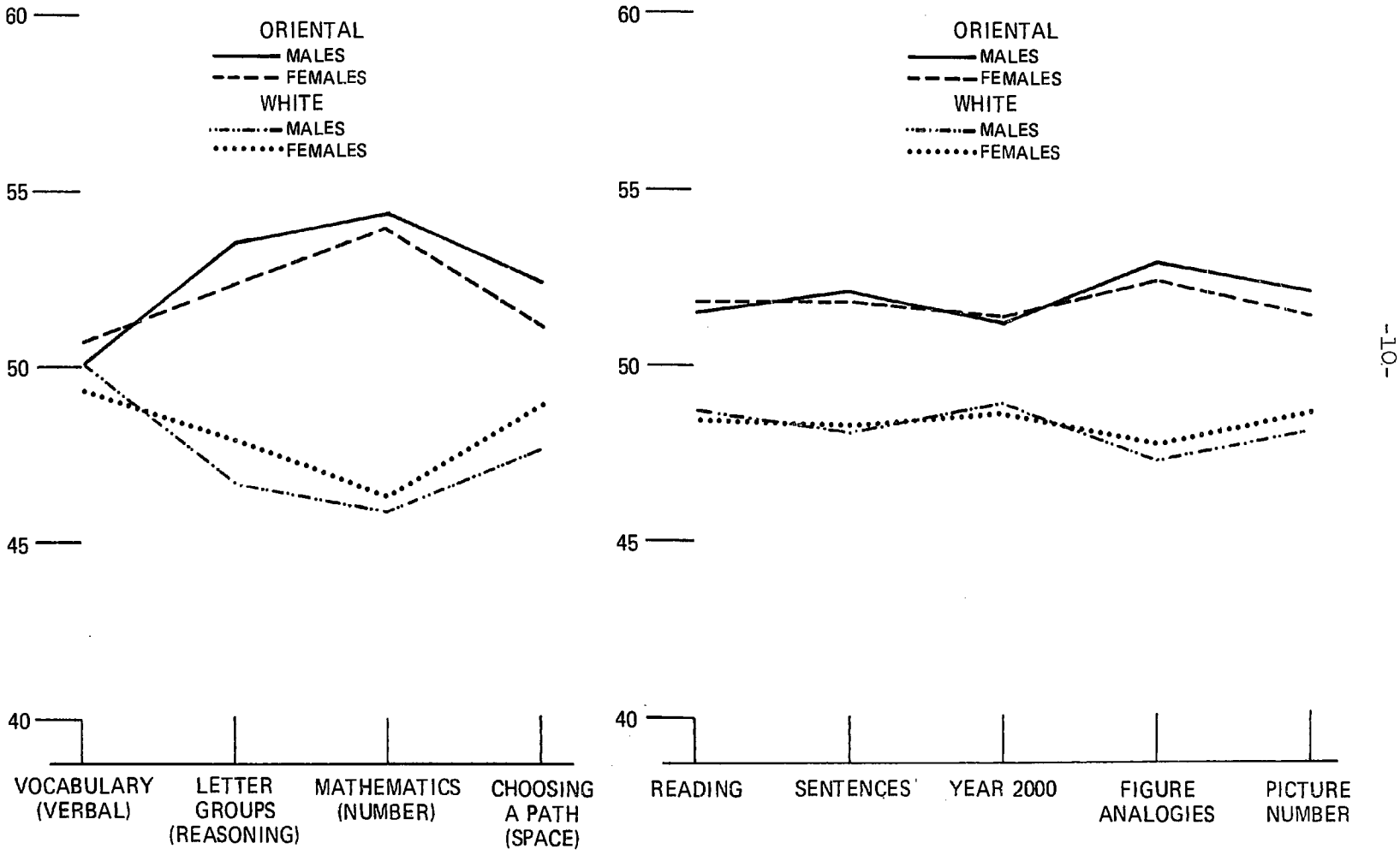

Figure 4. Contrasting patterns of performance on four tests, and five "additional" tests, by sex, between Oriental and White students.

scores, and consequently their graphed lines appear above the midline and the White students' lines appear below. It happens that there is some evidence from the background questionnaire data that the Oriental students were of a higher socioeconomic level than the White students in the same schools, and this might well be the explanation of these quite different results.[2] For the purposes of the present discussion, however, these data are presented here only for completeness and will not be included in the interpretations of the value of particular types of aptitude test.

Discussion

The differences from test to test in Figures 1, 2, and 3 are not large for the most part, but there is a certain amount of consistency in the results that would seem to deserve attention. For example, the comparison of Black and White students in Figure 1 shows that the Letter Groups and the Choosing A Path tests reduce the discrepancy between the races slightly over the discrepancy found in the Vocabulary and the Mathematics tests. In Figure 2, showing results for the Mexican-American students, the same trend is again present, but is small. However, the fact that the trend did occur in most of the situations where it was possible would indicate that some nonchance influence is at work and that these differences deserve further consideration. The same may be said of the consistent and sharp reduction in differences on the Picture Number test, as shown in Figures 2 and 3. (For a discussion of the influence of test reliabilities on these differences see Appendix III.)

---

[2] The same caveat applies, of course, in the interpretation of any of the mean differences between any of the groups. That is, the SES can be expected to influence the level of test scores, but not the relative ranking of the tests within a group (Flaugher, 1971).

Besides the consistency, another justification for the interpretation of these rather small differences between the relative test scores is the potential importance of even slight alterations in test performance by a large group of students. In national testing programs, because of the very large numbers of students involved, even slight differences in test score can be of consequence to a great many students, particularly in settings of competition for postsecondary educational opportunities. For example, on the basis of these results, the consequences of utilizing Letter Groups in a testing program, either in addition to the traditional verbal and math measures or as a substitute for one of them, would be higher scores for many Black and Mexican-American students.

It is likely that for any given subgroup of the population, a particular test could be designed on which that subgroup would score higher than any other. The viability and the usefulness of such a measure would remain to be proved, however. Such was not the case with the measures employed in the present study, in that each measure is either an example of a well-documented cognitive factor or an established measure of academic aptitude.

Questions of current validity are of interest, of course, and such information is presently being sought for these same students. But documentation of reduced discrepancies in test performance would seem justified in itself, since it reveals the existence of aptitudes which, because they are not typically documented in the traditional test batteries, are not acknowledged to exist in any great quantity. If they are not acknowledged, it is unlikely that they are being utilized in the traditional processes of education, and in such a situation, calculated validities would necessarily be low. Rather than reject the use of those particular aptitude measures,

however, it would seem more appropriate to consider the alteration of the curriculum to utilize more fully those talents whose existence is documented. In such an ideal circumstance, validity coefficients might then be considered an index of the success of the curriculum, rather than a measure of the worth of a particular aptitude measure.

The data presented here acquire additional significance when considered in the light of the recent "Report of the Commission on Tests" (College Board, 1970), whose charge was to review the testing functions of the College Entrance Examination Board and to consider possibilities for fundamental changes in tests and their use. In particular, John Carroll suggested an expansion of the number of component parts of the SAT to four or five reported scores. He suggested separation of the measured abilities into purely verbal, purely mathematical, and an isolation of the reasoning component from both the verbal or mathematical components. In discussing the separation of reasoning from the verbal scores, he stated that "presumably, the verbal scores would be largely a function of the individual's education and general reading experience, while the reasoning score would be less influenced by these factors and be more predictive of success for individuals with educational disadvantages" (p. 5).

Certainly a comparable case could be made to justify the separation of the mathematics score from the reasoning score. Mathematical test performance is even more clearly a function of the amount and quality of the formal education that the student has been exposed to, and there is ample evidence that these minority groups have not received a high quality education. Therefore, to attempt to judge such a student's reasoning

capacity by examining his score on a mathematics test would seem to be unwise and perhaps quite unfair in effect.

The results of the present study, then, seem compatible with the suggestions of the Commission on Tests to expand the number of measures included in traditional testing programs. Further, if it were necessary to choose the particular measures that should be included, then the measures of inductive reasoning, spatial aptitude, and associative memory could be recommended. Existing measures of other cognitive factors need to be explored, however, as do other measures of these same factors.

References

College Entrance Examination Board. Report of the Commission on Tests, Vol. II, "Briefs." New York: College Board, 1970.

Dressel, P. L. Some remarks on the Kuder-Richardson reliability coefficient. Psychometrika, 1940, 5, 304-310.

Flaugher, R. L. Patterns of test performance by high school students of four ethnic identities. Research Bulletin 71-25. Princeton, N. J.: Educational Testing Service, 1971.

Lesser, G. S., Fifer, F., & Clark, H. Mental abilities of children from different social-class and cultural groups. Monographs of the Society for Research in Child Development, 1965, 30(Whole issue No. 4).

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

APPENDIX I

Description of Project Access and Its Test Battery

Project Access is a program sponsored by the College Entrance Examination Board and has as its goal the providing of access to post-secondary education for minority/poverty students, while applying the knowledge gained in this endeavor to other testing and measurement activities that affect these students. In the 1969-70 school year, a nine-test Project Access battery was administered, together with a Biographical Inventory, to approximately 18,000 junior-year students in three cities: Los Angeles, Memphis, and Washington, D. C. Not all schools in each city participated. They were chosen within the city on the basis of interest and willingness to participate.

The nine aptitude tests included in the battery were as follows:

Vocabulary, 30 synonym items, 15 minutes. Knowledge of word meaning is measured.

Letter Groups, 25 items, 15 minutes. A measure of inductive reasoning. Five sets of four letters each are presented, the task being to find the rule which relates four of the sets to each other and to mark the one which does not fit the rule.

Mathematics, 50 items, 30 minutes. The quantitative comparison type of item was employed, involving the comparing of pairs of mathematical quantities to determine equality, or direction of inequality.

Choosing A Path, 18 items, 10 minutes. Spatial scanning is measured. Each item consists of a network of lines as in an electrical-circuit diagram having many intersecting and intermeshed wires with several sets of terminals. The task is to trace the lines and to determine for which pair of terminals there is a complete circuit. There is some orderliness in the layout to encourage comprehension of the pattern by scanning rather than by simple visual pursuit of lines.

Reading, 35 items, 25 minutes. Reading comprehension items on passages from short stories and articles.

Sentences, 40 items, 20 minutes. Detection of errors in written passages, involving language, punctuation, and capitalization.

Year 2000, 20 items, 10 minutes. Ability to follow complex directions, demonstrated by finding certain days on a calendar for the year 2000.

Figure Analogies, 25 items, 10 minutes. Analogy items composed of small geometric designs and figures.

Picture-Number, 15 items in each of two parts, 10 minutes. A two-part measure of associative memory between drawings and two-digit numbers. For each part, three minutes are allowed for study of the pairs, then two minutes for recalling the numbers when presented with the drawings.

In addition to the tests, a 15-minute biographical inventory asked information about ethnic identity, sex, aspirations, school grades, interests, and socioeconomic status.

APPENDIX II

Table A

Means and Sigmas of Raw Scores on Each Test, Converted Scores, and Majority-Minority Differences

	VOCABULARY 30 items				LETTER GROUPS 25 items				MATHEMATICS 50 items			
	Raw Scores		Converted Scores	Dif.	Raw Scores		Converted Scores	Dif.	Raw Scores		Converted Scores	Dif.
Los Angeles	$\bar{X}$	σ			$\bar{X}$	σ			$\bar{X}$	σ		
Black Male	16.8	6.2	47.4	.52	13.3	5.2	48.1	.38	27.6	10.6	47.3	.54
White Male	20.2	6.7	52.6		15.3	5.3	51.9		33.5	10.7	52.7	
Mex.Am. Male	16.5	5.7	47.2	.56	13.8	5.2	48.6	.28	29.2	10.3	48.0	.40
White Male	20.2	6.7	52.8		15.3	5.3	51.4		33.5	10.7	52.0	
Oriental Male	20.2	6.2	50.0	.00	18.9	4.4	53.4	.68	41.0	9.0	54.3	.86
White Male	20.2	6.7	50.0		15.3	5.3	46.6		33.5	10.7	45.7	
Black Female	15.7	5.5	47.1	.58	13.5	5.0	47.1	.58	24.7	9.6	47.3	.54
White Female	19.3	6.1	52.9		16.7	5.2	52.9		30.5	10.8	52.7	
Mex.Am. Female	14.8	5.0	46.3	.74	13.6	5.0	47.1	.58	24.5	9.0	47.1	.57
White Female	19.3	6.1	53.7		16.7	5.2	52.9		30.5	10.8	52.8	
Oriental Female	20.0	6.2	50.6	.12	18.8	4.4	52.2	.44	38.6	9.0	53.8	.76
White Female	19.3	6.1	49.4		16.7	5.2	47.8		30.5	10.8	46.2	
Memphis												
Black Male	15.6	5.2	44.5	1.10	12.8	5.2	46.3	.74	25.4	9.6	44.2	1.16
White Male	22.2	4.9	55.5		16.8	4.7	53.7		38.3	8.5	55.8	
Black Female	15.6	4.9	44.7	1.06	13.7	5.0	45.9	.82	23.7	8.8	44.6	1.08
White Female	21.6	4.6	55.3		17.8	4.3	54.1		35.4	9.3	55.4	

Table B

	PATH 18 items				READING 35 items				SENTENCES 40 items			
	Raw Scores		Converted Scores	Dif.	Raw Scores		Converted Scores	Dif.	Raw Scores		Converted Scores	Dif.
Los Angeles	$\bar{X}$	σ			$\bar{X}$	σ			$\bar{X}$	σ		
Black Male	6.9	3.6	47.5	.50	17.8	7.4	48.1	.38	18.8	6.9	48.4	.32
White Male	8.8	4.1	52.5		20.7	7.8	51.9		21.2	8.1	51.6	
Mex.Am. Male	7.9	4.0	48.8	.24	17.5	6.9	47.9	.42	19.0	6.9	48.5	.30
White Male	8.8	4.1	51.2		20.7	7.8	52.1		21.2	8.1	51.5	
Oriental Male	10.9	4.5	52.4	.48	22.8	7.2	51.4	.28	24.4	7.4	52.0	.40
White Male	8.8	4.1	47.6		20.7	7.8	48.6		21.2	8.1	48.0	
Black Female	5.3	2.6	47.9	.42	16.7	6.6	47.6	.48	20.0	6.6	47.4	.52
White Female	6.6	3.3	52.1		20.3	7.7	52.4		23.8	7.5	52.6	
Mex.Am. Female	5.6	2.8	48.4	.32	15.8	6.2	46.9	.62	19.7	6.5	47.2	.56
White Female	6.6	3.3	51.6		20.3	7.7	53.1		23.8	7.5	52.8	
Oriental Female	7.4	3.4	51.2	.24	22.7	7.3	51.6	.32	26.6	7.7	51.8	.36
White Female	6.6	3.3	48.8		20.3	7.7	48.4		23.8	7.5	48.2	
Memphis												
Black Male	6.8	3.4	45.9	.82	16.3	6.5	44.6	1.08	18.4	6.5	44.4	1.12
White Male	11.2	4.2	54.1		24.6	6.3	55.4		27.5	6.7	55.6	
Black Female	5.3	2.5	46.8	.64	16.2	6.0	44.4	1.12	21.1	6.3	44.0	1.20
White Female	7.3	3.2	53.2		24.3	5.9	55.6		30.2	5.6	56.0	

Table C

	YEAR 2000 (20 items)				FIGURE ANALOGIES (25 items)				PICTURE-NUMBER (30 items)			
	Raw Scores		Converted Scores	Dif.	Raw Scores		Converted Scores	Dif.	Raw Scores		Converted Scores	Dif.
Los Angeles	$\bar{X}$				$\bar{X}$				$\bar{X}$			
Black Male	8.2	3.7	47.1	.59	10.4	5.9	46.9	.61	17.1	8.2	49.5	.10
White Male	10.7	4.2	53.0		14.1	5.7	53.0		18.0	7.2	50.5	
Mex.Am. Male	8.6	3.7	47.4	.52	12.0	4.0	48.2	.36	16.8	7.4	49.2	.16
White Male	10.7	4.2	52.6		14.1	5.7	51.8		18.0	7.2	50.8	
Oriental Male	11.6	4.2	51.1	.22	17.4	5.6	52.8	.56	21.0	7.2	52.0	.40
White Male	10.7	4.2	48.9		14.1	5.7	47.2		18.0	7.2	48.0	
Black Female	8.2	3.5	47.1	.58	9.4	5.4	47.0	.60	18.2	8.3	49.1	.19
White Female	10.6	4.3	52.9		12.9	6.1	53.0		19.8	8.2	51.0	
Mex.Am. Female	7.9	3.6	46.8	.64	10.3	5.6	47.8	.44	17.2	7.2	48.3	.34
White Female	10.6	4.3	53.2		12.9	6.1	52.2		19.8	8.2	51.7	
Oriental Female	11.8	4.2	51.4	.28	15.8	5.4	52.4	.48	22.0	6.9	51.4	.28
White Female	10.6	4.3	48.6		12.9	6.1	47.6		19.8	8.2	48.6	
Memphis												
Black Male	8.1	3.5	44.8	1.04	9.2	5.3	44.9	1.02	18.6	8.3	48.6	.28
White Male	12.5	3.7	55.2		15.6	5.5	55.1		20.7	7.2	51.4	
Black Female	8.3	3.6	44.4	1.12	9.1	5.0	45.1	.98	20.4	8.2	48.8	.24
White Female	13.0	3.3	55.6		14.8	5.0	54.9		22.3	6.8	51.2	

APPENDIX III

Test Reliabilities and Their Influence on Mean Differences

If the two "traditional" tests, Vocabulary and Mathematics, have higher reliabilities than the "new" tests, Letter Groups, Choosing A Path, and Picture-Number, then there exists the possibility that the lesser mean differences between the groups on the new tests are due to this, rather than because of some qualitative differences in the tests. Such an explanation assumes that reduced reliability would act to reduce mean differences between the groups.

The reliabilities of seven of the nine tests were estimated by Dressel's modification of the Kuder-Richardson formula 20, which takes formula scoring into account (Dressel, 1940). The median obtained reliability estimates for the first four tests were: Vocabulary, .83; Letter Groups, .85; Mathematics, .91; and Choosing A Path, .74. Thus it appears that such an explanation is possible, but by no means a certainty, for the Choosing A Path results, since it has the lowest reliability of the four; such an explanation is not appropriate, however, for the Letter Groups results. As for the third "new" test with noticeable and consistent reductions in discrepancy, namely Picture-Number, the median reliability estimate was .93 and thus unlikely to be a factor in the results.