DOCUMENT RESUME

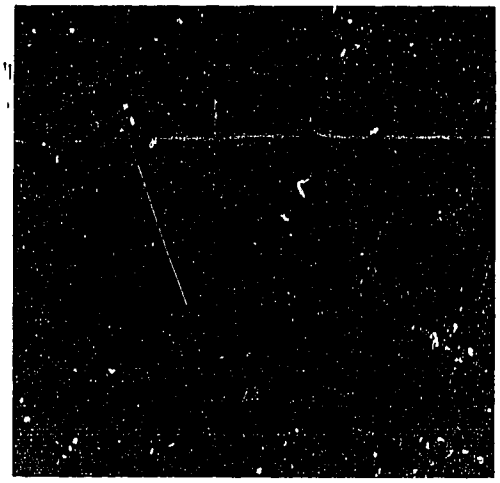
ED 280 854 TM 870 169

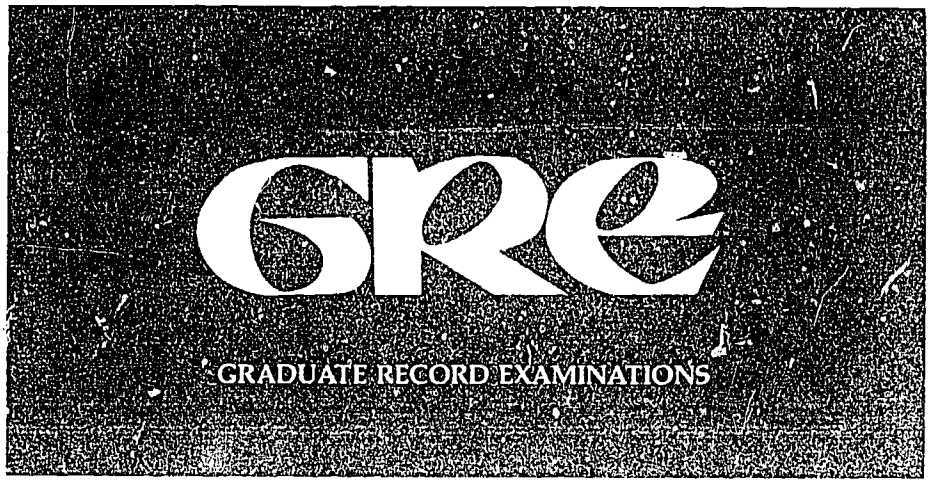
AUTHOR Wilson, Kenneth M.
TITLE The Relationship of Scores Based on GRE General Test Item Types to Undergraduate Grades: An Exploratory Study for Selected Subgroups.
INSTITUTION Educational Testing Service, Princeton, NJ. Graduate Record Examination Board Program.
REPORT NO ETS-RR-86-37; GREB-83-19P
PUB DATE Nov 86
NOTE 56p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Abstract Reasoning; College Entrance Examinations; *Discriminant Analysis; *Ethnic Groups; Grade Point Average; Grades (Scholastic); Higher Education; Majors (Students); *Predictor Variables; Reading Comprehension; *Sex Differences; *Test Items; Test Results; *Test Theory; Undergraduate Students; Verbal Ability; Vocabulary Skills
IDENTIFIERS *Graduate Record Examinations

ABSTRACT

This study used multiple discriminant analysis (MDA) to assess differences among subgroups defined in terms of sex and undergraduate major area, and ethnic group and undergraduate major area, with respect to patterns of performance on Graduate Record Examination (GRE) item-type part scores. Special subscores based on item types included in the current GRE General Test were derived for the study. The correlations of departmentally standardized scores on these subtests with a similarly standardized self-reported undergraduate GPA (SR-UGPA) criterion were analyzed. Multiple regression analysis was used to determine the relative contribution of the item-type part scores to prediction for the various subgroups. Predicted SR-UGPA means for subgroups, based on general major-area regression equations using GRE item-type part scores as predictors were compared with comparable predicted means using GRE total scores as predictors. The study was based on data from GRE files for 9,375 examinees in 12 fields of study, representing 437 undergraduate departments from 149 colleges and universities. Raw total number-right scores, similar raw scores based on the nine basic GRE General Test item types, a raw vocabulary score, and a raw reading comprehension score were computed. Item-type scores based on verbal and analytical ability item types provided more information about group differences than did total ability scores. (JAZ)





THE RELATIONSHIP OF SCORES BASED ON GRE GENERAL TEST ITEM TYPES TO UNDERGRADUATE GRADES: AN EXPLORATORY STUDY FOR SELECTED SUBGROUPS

Kenneth M. Wilson

GRE Board Professional Report GREB No. 83-19P
ETS Research Report 86-37

November 1986

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

The Relationship of Scores Based on GRE General Test Item Types to Undergraduate Grades:

An Exploratory Study for Selected Subgroups

Kenneth M. Wilson

GRE Board Report No. 83-19P

November 1986

Abstract

This study used multiple discriminant analysis (MDA) to assess differences among subgroups defined in terms of sex and undergraduate major area, and ethnic group and undergraduate major area, with respect to patterns of performance on GRE item-type part scores. Special subscores based on item-types included in the current GRE General Test were derived for the study. The correlations of departmentally standardized scores on these subtests with a similarly standardized self-reported undergraduate GPA (SR-UGPA) criterion were analyzed. Multiple regression analysis was used to determine the relative contribution of the item-type part scores to prediction for the various subgroups. Predicted SR-UGPA means for subgroups, based on general major-area regression equations using GRE item-type part scores as predictors were compared with comparable predicted means using GRE total scores as predictors.

The study was based on data from GRE files for 9,375 examinees in 12 fields of study, representing 437 undergraduate departments from 149 colleges and universities. Students were classified by field of study into four major areas: primarily verbal fields (English, history, sociology, political science); primarily quantitative fields (chemistry, computer science, mathematics, engineering, and economics); fields of mixed quantitative/verbal emphasis (Q/V), namely, biology and agriculture; and education.

The GRE scores involved were (a) raw number-right total verbal, quantitative, and analytical ability scores, unequated across test forms, transformed to a common scale—that is, z-scaled by test form—labelled V*, Q*, and A*, respectively, to distinguish them from the corresponding GRE scaled scores ; and (b) similarly developed verbal, quantitative, and analytical ability item-type part scores. Verbal part scores were based on antonyms, analogies, sentence completion, and reading passage sets. Primary interest was in a vocabulary score (antonyms plus analogies) and a reading comprehension score (sentence completion plus reading passages). Quantitative part scores were based on quantitative comparison, regular mathematics, and data interpretation item types. Analytical ability part scores were based on analytical reasoning and logical reasoning item types.

The part scores for each test were treated as individual variables in multiple discriminant analyses (MDA) for students classified by sex and major area, and by ethnic-group membership and major area. For the MDA, the ethnic groups were American Indian, Black, Mexican American and other Hispanic, Puerto Rican, and Asian American; for regression analyses, by major area, the groups were Black, all Hispanic origin, Asian American, all Minority, White, male, and female.

For each test, the criterion groups were found to be differentiated significantly along both a general ability dimension (represented by the principal discriminant function of part scores, all positively weighted), and a secondary, bipolar dimension (defined by a second significant discriminant function, uncorrelated with the general ability dimension, that reflected differences in patterns of performance on the part scores).

The part scores (especially vocabulary and reading comprehension, and analytical reasoning and logical reasoning) were found to exhibit different patterns of correlations with the SR-UGPA criterion.

With respect to both patterns of part-score means (mean scores on the second discriminant function) and patterns of part-score correlations with SR-UGPA, major-area differences appeared to be stronger and more systematic than ethnic-group or sex differences. Major-area differences were more pronounced when verbal and analytical part scores were used as independent variables than when quantitative part scores were used. Systematic major-area differences in patterns of part-score/SR-UGPA correlation were more clearly evident for reading comprehension and vocabulary part scores, and for analytical reasoning and logical reasoning part scores, than for the quantitative ability part scores.

Using part scores rather than total scores did not result in different inferences regarding the relative standing of subgroups on the SR-UGPA criterion. Predicted subgroup SR-UGPA means based on general major-area regression equations were essentially the same when item-type part scores were used as predictors as when the three GRE section scores (V*, Q*, and A*) were used as predictors.

Study findings indicate that the item-type scores, especially scores based on verbal and analytical ability item types, provide more information about group differences than is provided by the total ability scores. Questions regarding the incremental predictive value of this information remain unresolved on the basis of the study findings. Sample size was limited for several of the subgroups, predictive equations were not cross-validated, and self-reported undergraduate grades rather than graduate grades were used as criteria. Resolution of these questions is a matter for further research. To be most useful, such research would involve graduate-level performance criteria and employ equated part scores.

Based on the overall pattern of findings, attention might most profitably be focused on the potential contribution of separate subscores for reading comprehension, vocabulary, analytical reasoning, and logical reasoning. Continued exploration of questions regarding the validity of item-type part scores should contribute to better understanding of the nature of the abilities being measured by the GRE General Test, within if not beyond the well established verbal and quantitative domains.

Acknowledgments

The study was made possible by the support of the Graduate Record Examinations Board. Richard Harrison provided assistance in data management and analysis. Robert Altman, Brent Bridgeman, Linda Wightman, Philip Oltman, and Cheryl Wild critically reviewed various drafts of the study report, and made numerous helpful suggestions. These contributions are gratefully acknowledged. However, responsibility for the contents of the report is exclusively that of the writer.

Table of Contents

The Relationship of Scores Based on GRE General Test Item Types to Undergraduate Grades: An Exploratory Study for Selected Subgroups

Kenneth M. Wilson
Educational Testing Service

Study Background

The Graduate Record Examinations (GRE) General Test, widely used in evaluating the academic qualifications of applicants for admission to graduate study, provides measures of verbal, quantitative, and analytical reasoning abilities (ETS, 1984). Each of these general ability measures is composed of several different item types, thought of as being different methods of measuring their respective constructs (e.g., Rock, Werts, & Grandy, 1982).

The verbal measure employs four types of questions or items, namely, antonyms, analogies, sentence completions, and reading passage sets. Antonym items are designed to test the ability to identify words that are opposite in meaning, and analogy items test the ability to identify words or phrases that are related to each other in the same way as other words or phrases. The sentence completion items test the ability to identify words that are logically and stylistically consistent with the sentence in which they appear. A fourth set of items is included to test the ability to recognize in a reading passage the main ideas, information explicitly provided, implied ideas, the attitude of the author, and the like.

Three item types are employed in the quantitative measure. Quantitative comparison items test the ability to reason quickly and accurately regarding the relative sizes of two quantities or to perceive that not enough information is available to make such a decision. Quantitative items that measure basic mathematical skills, or regular mathematics, are included. These items are balanced among questions requiring arithmetic, algebra, and geometry. They are designed to test basic mathematical skills and understandings of concepts at levels applicable to individuals who have not specialized in mathematics. Data interpretation items test the ability to synthesize information presented in tabular or graphic form, to select data appropriate for answering a question, and so on.

The 1981 revision of the analytical measure includes two item-types, namely, analytical reasoning items and logical reasoning items. Analytical reasoning items test the ability to understand a given structure of arbitrary relationships among fictitious entities, deduce new information from given relationships, and the like. Logical reasoning items test the ability to understand, analyze, and evaluate arguments, recognize the point of an argument or the assumptions on which it is based, analyze evidence, and the like.

Interest in the potential predictive role of part scores based on item types included in the GRE General Test, especially the GRE verbal ability measure, was prompted by the results of undergraduate-level validity studies conducted by the College Board Validity Study Service (VSS) at ETS. For several years, vocabulary and reading comprehension subscores have been routinely reported for the Scholastic Aptitude Test (SAT) verbal ability

measure. The SAT vocabulary subtest uses antonymn and analogie items, and the reading comprehension subtest uses sentence completion and reading passage items. These are parallel in type to those included in the GRE Verbal Test.

Based on internal analyses (Ramist 1981a, 1981b) of over 100 validity studies conducted by the College Board Validity Study Service (VSS), the vocabulary and reading comprehension subscores of the SAT verbal measure were found to differ in validity for predicting freshman grade point average criteria. For example, the SAT reading comprehension subscore (sentence completion plus reading passage items) tended to be a consistently better predictor than the vocabulary subscore (antonym and analogy items) and essentially as valid as the entire verbal score, including the vocabulary items.

An exploratory study (Wilson, 1984), sponsored by the Graduate Record Examinations Board, was undertaken to assess the relationship of scores based on GRE vocabulary and reading comprehension item types to a self-reported undergraduate GPA criterion; part scores based on the quantitative and analytical ability item types were also studied. The study was based on data from GRE files for samples of undergraduate-level GRE test takers classified according to undergraduate department (institution and field of study).

Findings involving GRE vocabulary and reading comprehension item types were generally similar to those reported for the parallel SAT verbal item types. The reading comprehension subtest tended to be correlated more highly than the vocabulary subtest with undergraduate grades, and in some fields was more closely related to grades than was the total verbal score. However, in some major field subgroups the vocabulary subscore, but not the reading comprehension score, was significantly weighted in predictive composites with total quantitative and analytical ability scores, suggesting a potentially useful role in prediction for both of the verbal item-type part scores. There were also major-field differences in patterns of average vocabulary and reading comprehension scores. Majors in verbal fields (such as English, history, sociology, or political science) tended to score higher on vocabulary than on reading comprehension, while the opposite was true for majors in quantitative fields (such as chemistry or computer science.)

With respect to the quantitative item types, the data interpretation items appeared to be measuring a somewhat different dimension of quantitative ability than that measured by the other two item types. Majors in the verbal fields, for example, tended to perform better on the data interpretation items than on the regular mathematics or quantitative comparison items, while the opposite was true for majors in quantitative fields.

With respect to the analytical ability measure, the component represented by the analytical reasoning items appeared to differ from that being measured by the logical reasoning items along a quantitative versus verbal dimension. Analytical reasoning item-type part scores tended to exhibit "quantitative" characteristics while the logical reasoning part scores exhibited "verbal characteristics." Analytical reasoning items, for example, tended to be more predictive of undergraduate grades in quantitative than in verbal fields,

while logical reasoning items were more predictive in verbal fields. Also, majors in quantitative fields tended to have higher scores on analytical reasoning than on logical reasoning items, while the opposite was true for majors in verbal fields, and so on.

The Present Study

The present study analyzed GRE item-type part score data for students classified by sex and ethnic group membership as well as by undergraduate major area, using the data set developed for the original study. One aim of the study was to assess the patterns of relationships between part scores and undergraduate grades for the subgroups under consideration.

A second aim of the study was to assess systematically, using the method of multiple discriminant analysis (e.g., Klecka, 1975), the "dimensionality" of differences among subgroups (defined in terms of sex or ethnic group and major undergraduate area) with respect to performance on item-type part scores within each of the general ability measures. The issue of "dimensionality" calls for some elaboration.

The total score on each ability may be thought of as being made up of the sum of scores on subtests (part scores) based on the respective item types, weighted roughly according to the lengths of the respective subtests. The total analytical ability score, for example, may be defined as a linear composite (AR + LR) of scores on two subtests, namely, an analytical reasoning subtest, AR (38 items), and a logical reasoning subtest, LR (12 items).

Subgroups defined in terms of, say, major field are known to differ significantly along the single (total analytical score) dimension specified by AR + LR. Such subgroups may also differ systematically in analytical reasoning ability relative to logical reasoning ability. If so, in order to account for (or describe) the subgroup differences, it is necessary to think in terms of two dimensions. The general analytical ability dimension (the total score) is represented by AR + LR. However, a second dimension is needed to account for group differences in relative performance on the two subtests. This dimension is represented by AR - LR.

The central question regarding dimensionality in the present study is whether subgroups differ in both level of general analytical (verbal, quantitative) ability and relative level of performance on item-type part scores within the respective ability measures. Analytically, more than one linear combination of the two analytical (four verbal, three quantitative) item-type part scores may be required to account for differences among subgroups. If so, this would mean that the analytical (verbal, quantitative) item-type part scores provide more information about group differences than is provided by their summarization in a single total score. This information may prove to be useful for prediction, diagnosis, or guidance.

A third aim of the study was to compare predicted level of undergraduate

grades for subgroups based on GRE item-type part scores with predicted level based on GRE total scores.

The Basic Data Set

The data employed are from GRE files for a sample of 9,375 examinees who took the GRE General Test between October 1981 and June 1982, inclusive, as enrolled undergraduates or recent graduates not yet enrolled in graduate school. In addition, they were were U.S. citizens who reported English as the better language of communication. Only test takers who reported the undergraduate institution attended, the major field of enrollment, and the undergraduate grade point average in the major field and over the last two years of undergraduate study were included in the sample. The sum of the two self-reported undergraduate GPA variables (SR-UGPA) was employed as the academic performance criterion.

These test takers were from a total of 479 undergraduate departments (major field by designated undergraduate institution combinations), in 12 fields of study, from undergraduate institutions that are the major sources of GRE General Test takers. The fields involved were as follows:

(1) four fields judged to be primarily verbal (English, history, political science, and sociology;

(2) five fields judged to be primarily quantitative (chemistry, computer science, mathematics, electrical engineering, and economics);

(3) two fields judged to be of mixed verbal and quantitative emphasis (agriculture and biology, or biosciences); and

(4) education (a large field including students with a variety of subject- matter backgrounds).

Table 1 shows the distribution of the undergraduate departmental samples, by size and field of study. The percentage of students reporting membership in any ethnic minority group and the percentage of males is reported in the table, by field.

The Test Variables

Raw total number-right scores, and similar raw scores based on the nine basic GRE General Test item types, were computed for each member of the study sample; a raw vocabulary score (sum of scores on antonym and analogy items) and a raw reading comprehension score (sum of scores on sentence completion and reading passage items) were also computed. These scores and their acronyms, the number of items included in each score, and estimates of reliability are listed below.

Table 1

Distribution of Undergraduate Departmental Samples Included in the Study, by Size and Field

Sample size	Number of undergraduate departmental samples by field												
	English[a]	History[b]	Sociology[b]	Pol. Sci.[a]	Chemistry[b]	Comp. Sci.[b]	Mathematics[c]	Elec. Eng.[a]	Economics[b]	Agriculture[d]	Biology[d]	Education[d]	All fields
100+										1		1	2
90-99										-		-	-
80-89										-		2	2
70-79										1		-	1
60-69										2		3	5
50-59	1									4	2		8
40-49	-	1		1				3		1	5	10	21
30-39	2	-		2		2		6	2	2	11	4	31
20-29	16	5	2	6	7	5		12	6	13	33	19	124
10-19	24	33	24	16	38	34	13	15	36	-	-	-	233
<10	-	-	-	-	-	-	10	-	-	-	-	-	10
No. of depts	43	39	26	25	45	41	23	36	44	24	51	40	437
No. of students	884	584	364	545	644	647	251	850	663	976	1318	1649	9375
Male(%)	34.2	54.8	25.8	57.2	67.2	69.6	62.5	88.3	62.9	59.7	45.9	12.0	49.4
Minority(%)	11.0	13.9	29.2	18.8	14.6	17.9	11.0	11.6	15.4	7.3	14.9	9.0	14.1

Note. An undergraduate departmental sample includes individuals naming a designated undergraduate major field and a designated undergraduate school who were taking the GRE General Test during 1981-82 as either (a) enrolled undergraduates or (b) nonenrolled bachelor's degree holders no more than two years beyond the bachelor's.

[a]Minimum N = 15; [b]Minimum N = 10; [c]Minimum N = 9; [d]Minimum N = 20

Test	Typical form reliability and number of items		Acronym
Verbal Test (Total)	.90+	(76 items)	V*
Antonyms	.75+	(22 items)	ANT
Analogies	.70+	(18 items)	ANA
Sentence completion	.60+	(14 items)	SC
Reading passages	.80+	(22 items)	RC
Vocabulary (ANT + ANA)	.80+	(40 items)	VO
Reading Comprehension (SC + RD)	.80+	(36 items)	RC
Quantitative Test (Total)	.90	(60 items)	Q*
Quantitative comparison	.80+	(30 items)	QC
Regular mathematics	.75+	(20 items)	RM
Data interpretation	.60+	(10 items)	DI
Analytical Test (Total)	.85+	(50 items)	A*
Analytical reasoning	.80+	(38 items)	AR
Logical reasoning	.60+	(12 items)	LR

Unlike the operational, GRE-scaled total verbal, quantitative, and analytical ability scores, neither the raw total number-right scores (V*, Q*, and A*) nor the raw item-type part scores are equated across test forms. Six different forms were used during the 1981-82 testing year. The various item types are not necessarily parallel in difficulty within a given test form or across forms. Equating procedures were not feasible for the exploratory study. The raw item-type part and total scores were simply transformed to a common scale, by form. For each test form, the raw part and total scores of individuals were expressed as deviations from the mean of all examinees taking the form, in form standard deviation units. Thus, for each test, in the total sample (N = 9,375), the grand mean of the transformed scores was set equal to 0.0, and the standard deviations was 1.0.

Subgroup Distribution

A distribution of the number of students by ethnic group and by sex is provided in Table 2, for each field and for each of the four area classifications outlined above. The minority sample was made up of 346 Orientals or Asian Americans (AA), 361 Blacks (BL), 191 Puerto Ricans (PR), 100 Mexican Americans (MX) or Chicanos (CH), 83 Other Hispanics (OH), and 52 American Indians (AI).

Group differences in the percentage distribution of students by field and major area are apparent in Table 3. For example, more than half of the Puerto Ricans as compared to 18 percent of the Chicanos were in biology or agriculture. Almost half (46 percent) of the American Indian students were in

Table 2

Distribution of the 1981-82 Part-score Sample by Field, Ethnic Group, and Sex

Field	N*	Ethnic group								Sex		
		AI	BL	CH	PR	OH	AA	OMIN	WH	Total**	Women	Men	Total***
English	884	10	31	7	1	3	14	29	772	867	578	301	879
History	584	7	16	15	1	5	19	17	496	576	261	316	577
Sociology	364	3	56	9	2	8	20	7	254	359	268	93	577
Pol. Sci.	545	4	36	12	2	10	17	19	432	532	230	307	537
All Verbal	**2377**	**24**	**139**	**43**	**6**	**26**	**70**	**72**	**1954**	**2334**	**1337**	**1017**	**2354**
Chemistry	644	-	12	1	36	5	27	12	544	637	208	426	634
Computer Sci.	647	-	40	5	3	6	47	12	520	633	196	448	644
Mathematics	251	-	12	-	1	1	11	2	219	246	94	157	251
Elec. Engin.	850	3	30	9	17	18	89	12	646	824	99	748	847
Economics	663	7	19	4	16	8	30	16	548	648	245	416	661
All Quant	**3055**	**10**	**113**	**19**	**73**	**38**	**204**	**54**	**2477**	**2988**	**842**	**2195**	**3037**
Biology	1318	6	23	12	79	9	3	9	1093	1285	707	601	1308
Agriculture	976	7	10	6	29	6	3	9	893	963	390	578	969
All Q/V #	**2294**	**13**	**33**	**18**	**108**	**15**	**47**	**28**	**1986**	**2248**	**1097**	**1179**	**2276**
Education	**1649**	**5**	**76**	**20**	**4**	**4**	**25**	**12**	**1483**	**1629**	**1439**	**197**	**1636**
All Fields	**9375**	**52**	**361**	**100**	**191**	**83**	**346**	**166**	**7900**	**9199**	**4715**	**4588**	**9303**

Note: AI = American Indian, BL = Black, CH = Chicano or Mexican American, PR = Puerto Rican, OH = Other Hispanic, AA = Asian American, OMIN = other Minority, WH = White.

* Number with GRE scores and SR-UGPA.
** Number responding to question on ethnic group membership.
*** Number of examinees classifiable by sex.
# Q/V = Bioscience fields with balanced quantitative and verbal emphasis.

Table 3

Percentage Distribution of Members of Ethnic Groups, by Undergraduate Field

Field	Ethnic group*								
	AI	BL	CH	PR	OH	AA	OMN	WH	Total
English	19.2	8.6	7.0	0.5a	3.6a	4.0	17.5	9.8	9.4
History	13.5	4.4	15.0	0.5a	6.0	5.5	10.2	6.3	6.2
Sociology	5.8a	15.5	9.0	1.0a	9.6	5.8	4.2	3.2	3.9
Political Sci.	7.7a	10.0	12.0	1.0a	12.0	4.9	11.4	5.5	5.8
All Verbal	46.2	38.5	43.0	3.1	31.3	20.2	43.4	24.7	25.3
Chemistry	-	3.3	1.0a	18.8	6.0	7.8	7.2	6.9	6.9
Computer Sci.	-	11.1	5.0	1.6a	7.2	13.6	7.2	6.6	6.9
Mathematics	-	3.3	-	0.5a	1.2a	3.2	1.2a	2.8	2.7
Elec. Engin.	5.8a	8.3	9.0	8.9	21.7	25.7	7.2	8.2	9.1
Economics	13.5	5.3	4.0a	8.4	9.6	8.7	9.6	6.9	7.1
All Quantitative	19.2	31.3	19.0	38.2	45.8	59.0	32.5	31.4	32.6
Biology	11.5	6.4	12.0	41.4	10.8	12.7	11.4	13.8	14.0
Agriculture	13.5	2.8	6.0	15.2	7.2	0.9a	5.4	11.3	10.4
All Q/V**	25.0	9.1	18.0	56.6	18.1	13.6	16.9	25.1	24.5
Education	9.6	21.1	20.0	2.1a	4.8a	7.2	7.2	18.8	17.6
All Fields (N)	(52)	(361)	(100)	(191)	(83)	(346)	(166)	(7900)	(9199)

Note: Column totals should equal 100 percent within limits of rounding.

* AI (American Indian); BL (Black); CH (Chicano or Mexican American); PR (Puerto Rican); OH (Other Hispanic); AA (Asian American); OMN (Other Minority); WH (White).

a Percentage based on less than five cases.

** Q/V = Bioscience fields with balanced quantitative and verbal emphasis.

verbal fields, while Asian Americans were heavily concentrated in quantitative fields (59 percent), and so on.

The basic classifications employed in the study were as indicated below.

Acronym	and Group	Total	Verbal fields	Quantitative fields	Mixed fields	Education
AI*	American Indian	52	24	10	13	5
BL**	Black	361	139	113	33	76
MX/O*	Mexican American + Other Hispanic	183	69	57	33	24
PR*	Puerto Rican	191	6	73	108	4
HISP-T**	Hispanic (Total)	374	75	130	141	28
AA**	Asian American or Oriental	346	70	204	47	25
MIN-T**	Minority (Total) (AI through AA)	1133	308	457	234	134
WH**	White	7900	1954	2477	1986	1483

Subgroups with single asterisks (*) following the acronyms were treated separately in analyses of group differences in part-score performance, but not in regression analyses. The regression analyses were based on the groups denoted by double asterisks (**). Thus, for example, American Indians were treated as a separate group in analyses of group differences in performance on item-type part scores, but as part of Minority Total in regression analyses; Puerto Ricans were included in Hispanic Total in regressions but treated separately in analyses of group differences, and so on.

The distribution of students by sex and and major area was as indicated below.

Sex	Total	Major areas Verbal	Quantitative	Mixed	Education
Female	4715	1337	842	1097	1439
Male	4588	1017	2195	1179	197
Total	9303	2354	3037	2276	1636

Study Methods and Procedures

Analysis of Subgroup Differences

For the study of subgroup differences, multiple discriminant analysis (MDA) was the principal method employed, using Statistical Package for the Social Sciences (SPSS) routines (Klecka, 1975). MDA was used to assess the dimensionality of observed differences among subgroups with respect to performance on verbal, quantitative, and analytical item-type part scores.

Given observations on p test or other variables for members of G groups,

multiple discriminant analysis yields either p or G-1 statistically uncorrelated) linear combinations of the p variables, whichever is smaller. Functions are derived in such a way that the first function (weighted composite of test scores on other variables) accounts for the largest percentage of among-group differences, the second function accounts for the second largest percentage, and so on. Each function is uncorrelated with other functions.

Evaluation of standardized discriminant function coefficients (weights for the test or other variables involved) and comparison of group discriminant-score means (means of linear composites of the variables, weighted as specified by the analysis), provide a basis for interpreting results of the MDA (see, for example, Klecka, 1975, 443 ff.).

Standardized discriminant coefficients (ignoring signs) reflect the relative contribution of the variables to discrimination among the groups involved on the particular discriminant function under consideration, analogous to the interpretation of beta weights (standard partial regression coefficients) in multiple regression analysis.

The discriminant solution also yields raw-score weights corresponding to the standardized weights for each function. The raw-score weights may be used to compute a discriminant (function) score for each individual on each function. Comparison of the mean discriminant scores of subgroups on the principal and secondary functions provides an additional basis for interpretation of MDA outcomes. Ordinarily, the discriminant scores are standardized—that is, z-scaled with reference to the grand mean for all groups—for comparison of group means.

In the present study, the multiple discriminant analyses were designed specifically to determine whether or not only one significant linear function of the item-type part scores for a given ability measure would be required to account for significant differences among selected groups. If only one function is significant, this would mean that the groups differ only with respect to general verbal, quantitative, or analytical ability. However, a second (or other) function may also prove to be significant in any given analysis. This outcome would indicate that the item-type part scores involved provided information about group differences beyond that provided by the principal function—corresponding to the total score on the measure involved.

For example, certain subgroups that differ in, say, total verbal or analytical ability may also perform significantly better on vocabulary than on reading comprehension, or may have high analytical reasoning relative to logical reasoning ability. Such differences cannot be indicated in the total verbal or analytical ability scores. The differences would become apparent only if the item types were scored separately.

Groups are expected to differ primarily in the total ability under consideration. The principal discriminant function may be thought of as representing general ability. The component item-type part scores should be positively weighted. If the second (or other) discriminant function is significant, this would indicate that the groups under consideration differ

not only in general ability, but also in relative development of the abilities represented by the part scores under consideration. On the second function, differences in relative performance on item types within a given ability would be indicated by differences in the signs of the weights for the item-type part scores under consideration.

Results of the original study (Wilson, 1984) indicated different patterns of part-score performance by major area. The prevalence of major-area differences would complicate the interpretation of MDA findings for groups that might be defined in terms of sex or ethnicity alone. The distribution of these subgroups by major area is not random. Accordingly, groups were defined in terms of sex and major area, and ethnicity and major area, rather than by sex or ethnicity alone.

Because of the disproportionately large size of the sample of White students, the multiple discriminant analyses involving ethnic groups were based on data for five minority ethnic groups only: American Indian, Black, Mexican American and Other Hispanics, Puerto Ricans, and Asian Americans.

Thus, in each MDA involving ethnic minorities the number of groups was 20—5 ethnic groups classified by 4 undergraduate major areas. In each MDA involving sex, the number of groups was 8—males and females classified by 4 undergraduate major areas. The number of independent variables ranged from 2 to 4. The number of independent variables in each analysis (two, three, or four item-type part scores) was smaller than the number of groups. Therefore, the total number discriminant functions that could be extracted in each analysis was equal to the number of independent variables.

Analyzing the Relationship of Part Scores to SR-UGPA

In studying the relationship of the GRE part scores and total scores to undergraduate grades (self-reported UGPA or SR-UGPA), the transformed raw part and total scores, as well as SR-UGPA, were expressed as deviations from department-level grand means in departmental standard deviation units. A department-level grand mean is defined as the mean for all individuals majoring in a given field, such as English or mathematics, at a given undergraduate institution, without regard to subgroup membership.

The department-level, z-scaled data were pooled for regression analysis by the four broad major areas described earlier. Each pooled coefficient for a broad area is equivalent to a weighted average of the department-level coefficients for the variables involved, for the departments (fields) included in a broad area. The coefficients may be thought of as approximating population values around which the department-level coefficients will vary due to sampling and other considerations. In this study, the approximations to population values represented by the pooled-sample coefficients are of primary interest.

Multiple regression analyses (MRA) were conducted for students classified by sex within each of the four major areas, and for selected ethnic-group

classifications within the four major areas, as follows: Black, Hispanic Total (Mexican American, Puerto Rican, other Hispanic), Asian Americans, Minority (total), and White. Two general regression equations were used to generate expected subgroup standing within each major area. One equation used V*, Q*, and A* total scores, and the other used VO, RC, QC, PM, DI, AR, and LR (GRE item-type part scores) as predictors (see list, page 6). The general equations were based on data for all students in each undergraduate major area without regard to their subgroup membership.

These analyses were designed to permit assessment for each test, by subgroup, of (a) the pattern of simple correlations between part scores and SR-UGPA, (b) the relative contribution of the item-type part scores in part-score/total-score composites, (c) the likelihood that separate treatment of part scores in this way would yield increments in multiple correlation over the basic total score composite (V*Q*A*), and (d) the possibility that use of item-type part scores rather than total scores as predictors might lead to improved inferences regarding the probable performance of subgroups within the respective major areas.

In evaluating the relationship of part scores to the SR-UGPA criterion it should be kept in mind that the part scores are based on different numbers of items. They represent subtests of unequal length and reliability. If two subtests measuring the same ability differ in length, the longer subtest would be expected to demonstrate somewhat higher validity than the shorter subtest due to greater reliability of measurement. For example, the analytical reasoning (AR) subtest includes 38 items while the logical reasoning (LR) subtest includes only 12 items. If these two item types are measuring the same ability, the validity of the 38-item AR subtest should tend to exceed that of the 12-item LR subtest, because the AR subtest is more reliable. However, if validity coefficients for the LR subtest tended to be equal to or higher than those for the AR subtest, this would indicate that the item types are tapping somewhat different abilities. Similarly, if the correlations of a given item-type part score with the SR-UGPA criterion are equal to or higher than those of the corresponding total score, factors other than differences in reliability clearly must be considered in order to explain the finding.

Findings

Results of the Multiple Discriminant Analyses

For each GRE ability measure, the MDA results indicated that the various subgroups were differentiated significantly along both (a) a major general ability dimension (defined by the principal discriminant function of item-type part scores, all positively weighted), and (b) a secondary bipolar dimension (defined by a combination of positively weighted and negatively weighted part scores). Two significant discriminant functions were obtained in every analysis but one—that involving five minority-ethnic groups classified by four major areas, with vocabulary and reading comprehension as the independent variables. However, when the four basic verbal item-type part scores were used, two significant functions were obtained.

The fact that a second discriminant function was significant in these analyses indicates that part scores based on verbal, quantitative, and analytical ability item types provided information regarding sex, ethnic group, and/or major-area differences that was not provided by the corresponding total scores. Put another way, this result indicates differential development within individuals of the skills or abilities being measured by the various item types as well as differential development in level of the corresponding general ability.

MDA Involving Verbal Test Part Scores

Table 4 shows results of the MDA involving two sets of verbal item-type part scores as independent variables. Set 1 analyses were those in which the four verbal item-type part scores (antonyms, analogies, sentence completions, and reading passages—ANT, ANA, SC, RD) were used. Four discriminant functions were derived. Set 2 analyses were those in which vocabulary (ANT + ANA) and reading comprehension (SC + RD) were the independent variables. Two discriminant functions were derived.

Standardized discriminant function coefficients are shown for the principal and second functions derived in each analysis. For example, in the Set 1 results for the 20 ethnic groups, all four verbal part scores were positively weighted on the principal function. Highest weights were associated with analogies (.41) and reading passages (.36), but the other verbal item types also contributed positively to this function (general verbal ability). On the second function, reading passages (1.12) and antonyms (.45) were contrasted with sentence completions (-.77) and analogies (-.77).

The percentage of total among-groups variance accounted for by each function is also shown in Table 4. The sum of the percentages (not tabled) indicates the percentage of total among-groups differences accounted for by the two discriminant functions. Thus, for example, continuing with the Set 1, ethnic group, illustration, the principal function accounted for 86.5 percent of group differences, and the second function accounted for 7.1 percent. The two functions (only the first two were significant in this analysis) accounted for 94.6 percent of the information about differences among these particular groups on the four verbal part scores.

In analyses of differences among the sex-by-major subgroups and the ethnic-by-major subgroups, respectively, with respect to the four verbal item-type part scores, two discriminant functions were significant. However, when vocabulary and reading comprehension scores were the independent variables, two functions were significant only in the sex-by-major analysis.

On the second function in the sex-by-major analysis, the weighting of antonyms and analogies (the two vocabulary item types) relative to the weighting for sentence completion and reading passages (the two reading comprehension itemvtypes) was consistent with the weighting of the vocabulary and reading comprehension partvscores. Thus, the two vocabulary component item-types (ANT and ANA), negatively weighted, were contrasted with the two reading comprehension component types (SC and RD), positively weighted.

Table 4

Standardized Discriminant Function Coefficients for Verbal Test Analyses

Part scores employed	Criterion groups for analysis: Ethnic group by major area (20 groups)		Sex by major area (8 groups)*	
Set 1	Principal function	Second function	Principal function	Second function
ANT (Antonyms)	.26	.45	.38	-.93
ANA (Analogies)	.41	-.77	.26	-.20
SC (Sentence Completions)	.18	-.81	.37	.06
RD (Reading Passages)	.36	1.12	.18	1.17
Percent variance	86.5	7.1	81.8	16.4
Significance	p <.001	p <.001	p <.001	p <.001
Set 2				
Vocabulary (ANT + ANA)	.59	-1.22	.60	-1.31
Reading Comprehension (SC + RD)	.51	1.26	.48	1.47
Percent variance	97.1	2.9	86.4	13.6
Significance	p <.001	p >.630	p <.001	p <.001

Note. The standardized discriminant function coefficients (ignoring signs) reflect the relative contribution of the part-scores to the respective functions. Interpretation of these weights is analogous to interpretation of standard partial regression (beta) weights in multiple regression analysis (see, for example, Klecka, 1975, pp. 443 ff.). Functions are derived in such a way that the first or principal function accounts for the greatest amount of among-groups variance, the second, statistically uncorrelated function accounts for the second greatest amount, and so on.

* In the sex-by-major-area analysis involving the four basic verbal item-type part scores (Set 1), a third discriminant function was significant ($p < .001$); it accounted for only 1.4 percent of the total among-groups variance.

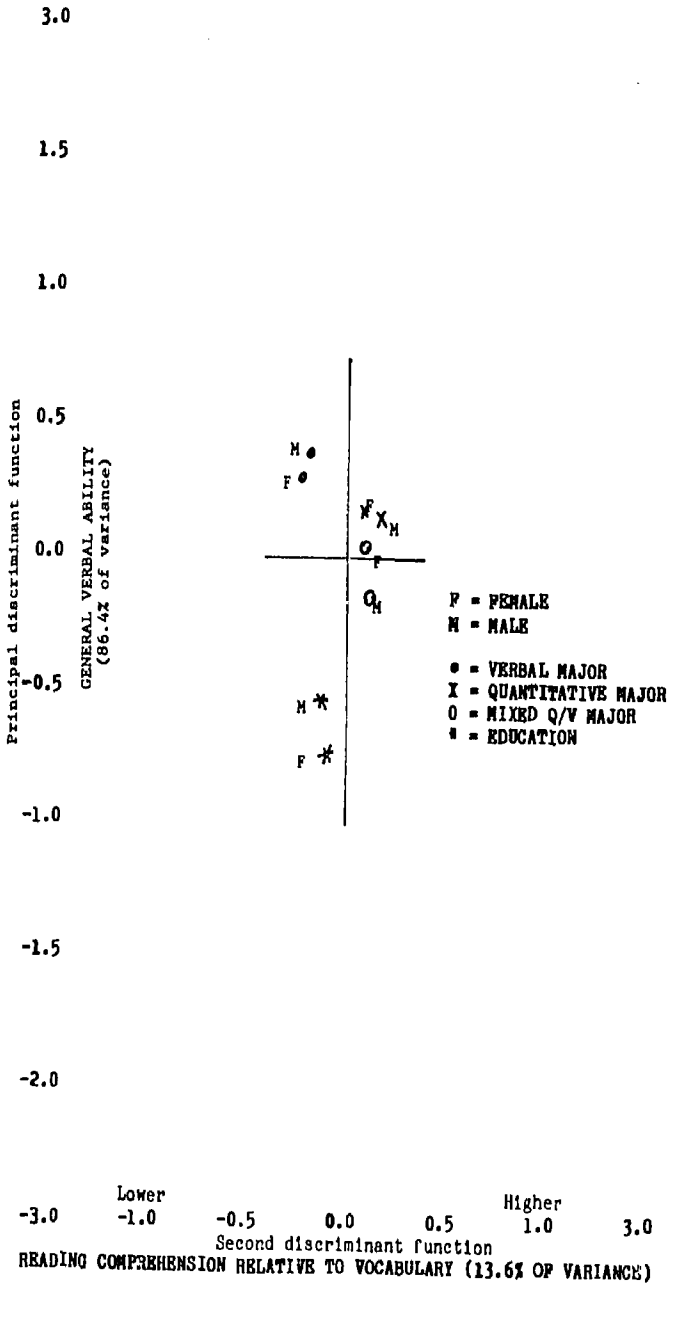
This contrast indicates that the relative performance of subgroups on antonyms and analogies, scored separately, relative to their performance on sentence completions and reading passages, similarly scored, was consistent with their relative performance on the vocabulary and reading comprehension subtests. Put another way, the two vocabulary component item-types exhibited similar properties and the two reading comprehension component item-types exhibited similar properties, different from those of the vocabulary set.

In the ethnic-by-major subgroup analysis, however, this consistency in relative performance was not found. The antonyms subscore (a VO component) was paired by sign with the reading passage subscore (an RC component). However, the analogies subscore (a VO component) and the sentence completion subscore (an RC component) were paired by the opposite sign. The second function was actually defined primarily by differential subgroup performance on the two item types that made up the reading comprehension part score. Reading passages received the highest positive weight (1.12), and sentence completions received the highest negative weight (-.81). This outcome indicates that when minority ethnic groups were classified by major area, the subgroups thus defined differed somewhat in relative performance on item types within the vocabulary and reading comprehension part scores.

On the principal function (representing a general verbal ability dimension), in both the two-score and the four-score analyses, without regard to the specific set of groups involved, all scores were positively weighted, as expected. There were some modest differences in the relative weighting of the four part scores on the principal function: for example, in the ethnic-group-by-major analysis, analogy and sentence completion item types were more heavily weighted than antonyms and reading. When sex groups were involved, reading passage item-type scores were less heavily weighted than the other three item-type part scores on the first function.

Group discriminant-score centroids. Two discriminant (function) scores (linear combinations of scores on the relevant verbal item-type part scores, weighted relatively as indicated in Table 4) were derived for each individual. Two discriminant score means were computed for each subgroup. The two discriminant-score means were used to locate points representing the joint performance of subgroups on the two discriminant functions. These points are called called group centroids. In evaluating the findings, one should keep in mind that the 8 sex-by-major subgroups were comparatively much larger than the 20 ethnic-by-major subgroups. Thus, centroids for ethnic-by-major subgroups are subject to greater sampling fluctuation than those for sex-by-major subgroups.

Centroids of the sex-by-major subgroups on the first and second verbal part-score discriminant functions are plotted in Figure 1. Centroids are shown only for groups with $N > 19$. Thus, for American Indians only a verbal-area centroid is shown; for Puerto Ricans, centroids are shown for quantitative-area majors and mixed quantitative/verbal area majors only. The left frame shows centroids for analyses involving vocabulary and reading comprehension part scores, and the right frame shows group centroids based on linear discriminant functions of antonym, analogy, sentence completion, and reading passage part scores, with relative weighting as indicated in Table 4.



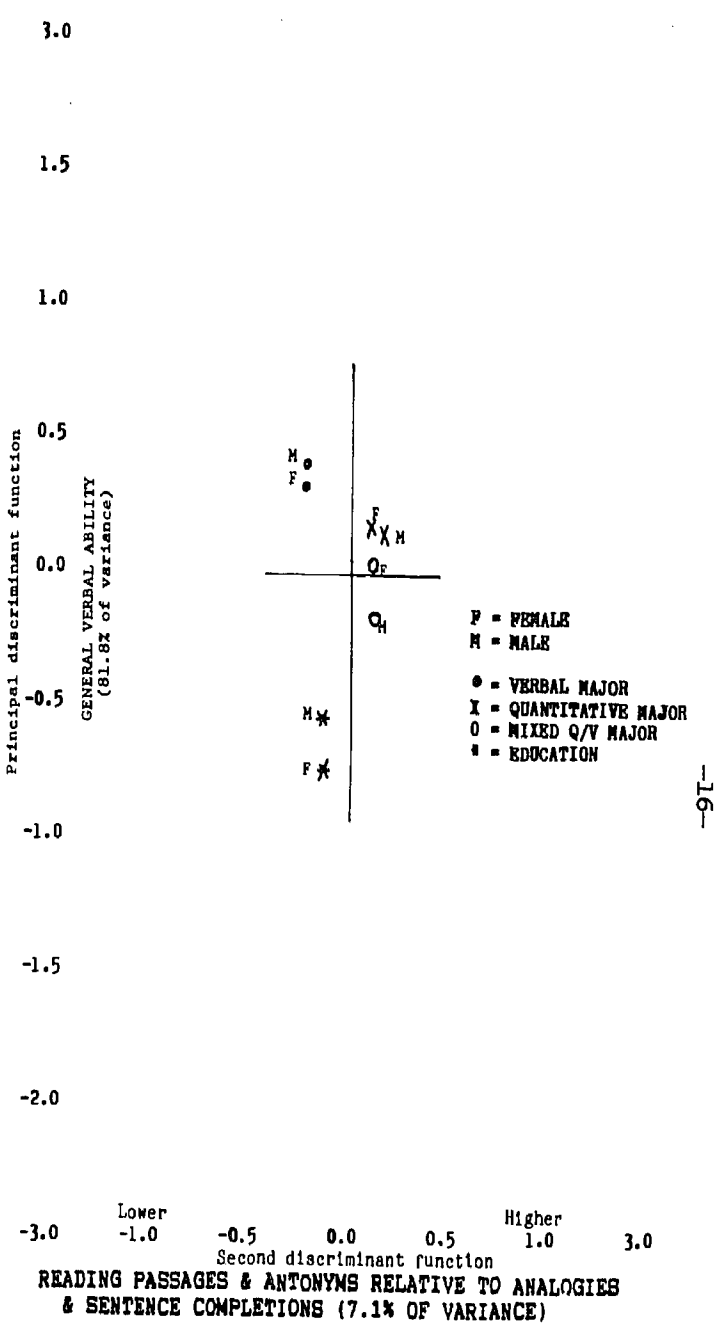


Figure 1. Centroids of sex-by-major subgroups: Discriminant functions of reading comprehension and vocabulary part scores (left) and of analogy, antonym, sentence completion, and reading passage part scores (right)

Figure 1a shows profiles of mean scores for males and females, by undergraduate major area, on vocabulary and reading comprehension (upper frame), and on sentence completion (SC) and reading passage (RD) part scores (lower frame). Means on the two reading comprehension component item types are profiled because they were consistently (positively) weighted in the sex-by-major analysis, but differentially weighted in the analysis involving ethnic groups—RD, high positive, versus SC, high negative.

It should be kept in mind that differences on the principal function may be thought of as reflecting general-ability, total-score differences. Since the second function is uncorrelated with the principal, general ability function, differences on the second function may be thought of as reflecting additional or supplemental information provided by separate treatment of part scores on the item types.

The degree of separation of centroids of groups (sex and/or major area) along the second-function (horizontal) axis (Figure 1), reflects the degree of departure-from-parallelism exhibited by the profiles of part score means for the corresponding groups in Figure 1a. For example, consider data for males and females in verbal areas versus quantitative areas. In Figure 1, centroids for quantitative-area majors (male and female) are located to the right of the major vertical line that corresponds to the grand mean for the second discriminant score.

The second discriminant score, in standardized form, is defined by 1.47 (RC) - 1.31 (VO), as indicated in Table 4. Thus, the location of the centroids for both male and female quantitative-area majors indicates that, on the average, they scored high on reading comprehension relative to vocabulary. In Figure 1a, the quantitative-area mean profiles of vocabulary and reading comprehension are consistent with this interpretation. Centroids for verbal majors (male and female) in Figure 1, on the other hand, are located to the left of the major vertical, indicating low average reading comprehension relative to vocabulary, consistent with the slope of the mean profiles in Figure 1a. The mean profiles in Figure 1a indicate average deviation from the grand mean for all minority ethnic groups in standard deviation units.

A high degree of consistency in the results of both the two-score and the four-score analyses involving sex-by-major subgroups is suggested by comparison of data summarized in Figure 1 with that summarized in Figure 1a. First, the two plots of group centroids in Figure 1 are almost identical. Second, the profiles in Figure 1a of VO and RC means and of SC and RC means, for males and females within each of the four major areas are basically parallel. Third, for both sets of part scores, mean profiles for verbal-area and education majors differ in slope from profiles for majors in the quantitative and mixed quantitative/verbal (Q/V) areas. The mean profiles for majors in verbal fields and in education indicate lower reading comprehension than vocabulary scores (also, lower reading passage than sentence completion scores), while profiles for majors in the other areas indicate the opposite patterns.

For the minority ethnic groups (see Figure 2), although the second function of reading comprehension and vocabulary scores was not significant, means were computed and used to locate centroids of the groups (left frame of

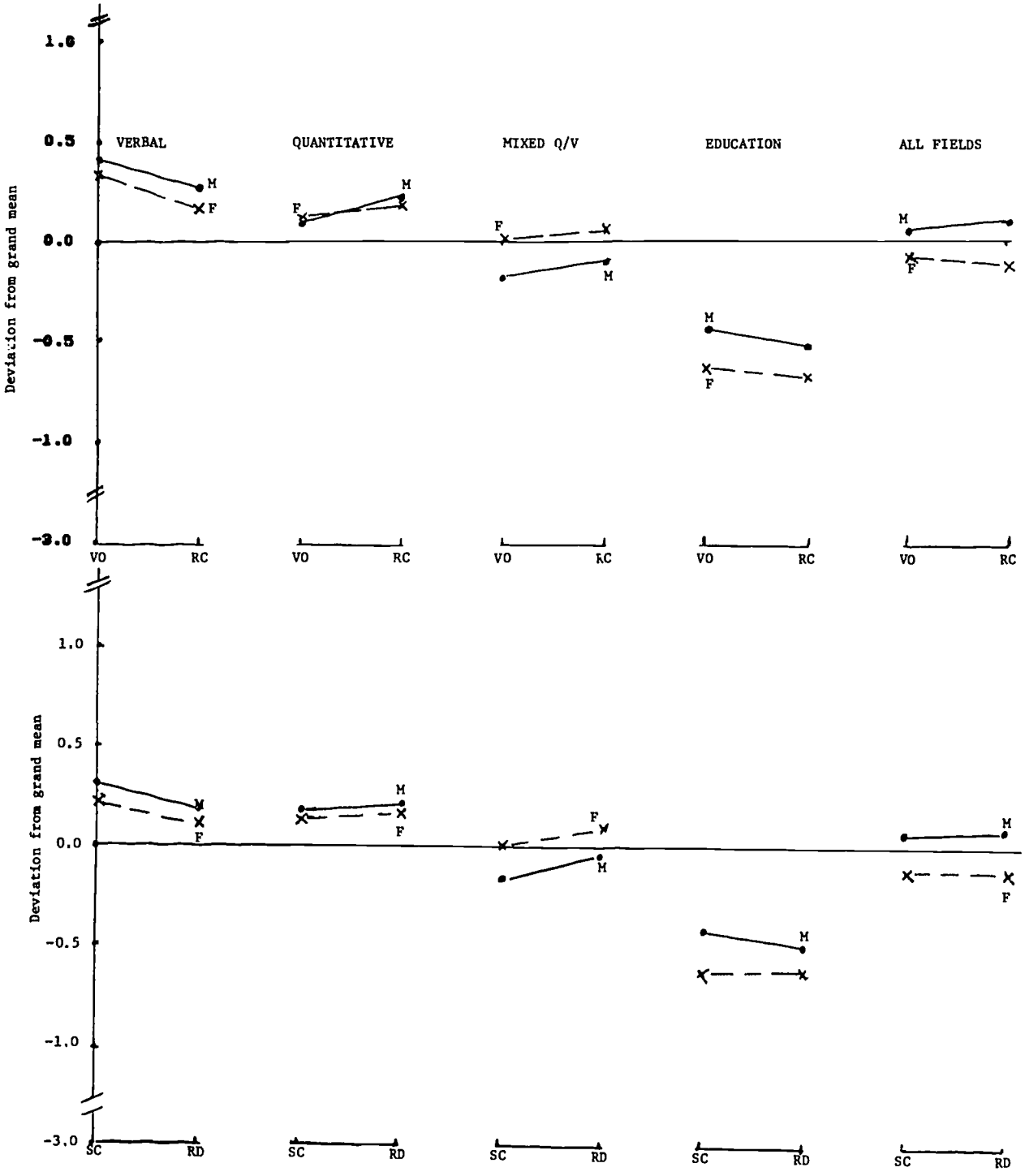


Figure 1a. Profiles of mean vocabulary and reading comprehension part scores (upper), and sentence completion and reading passage part scores (lower), by sex and undergraduate major area

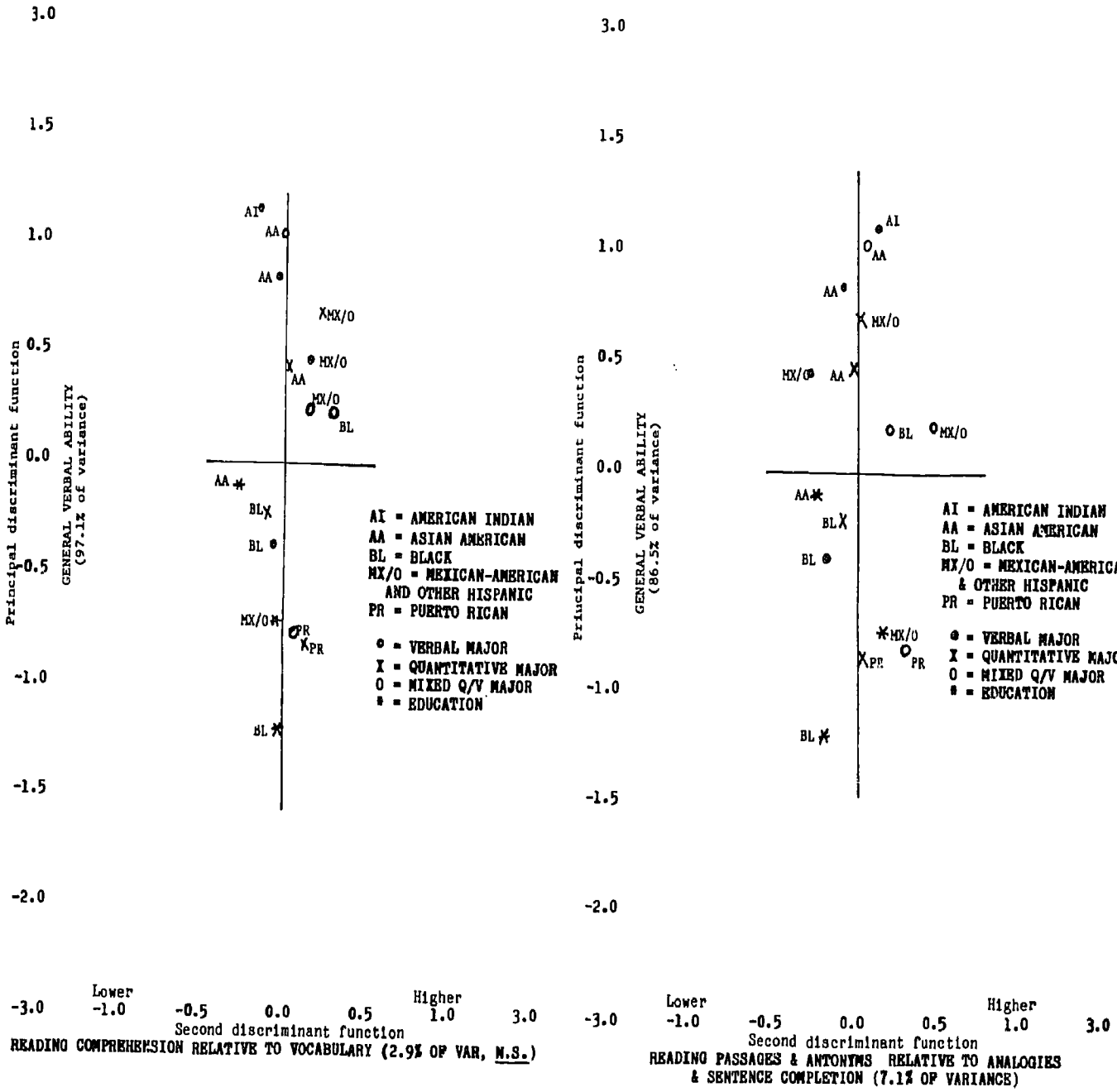


Figure 2. Centroids of ethnic-by-major subgroups: Discriminant functions of reading comprehension and vocabulary part scores (left) and of analogy, antonym, sentence completion, and reading passage part scores (right)

Figure 2). It may be seen that differentiation of majors in education and verbal fields from majors in the quantitative and mixed quantitative and verbal (mixed Q/V) fields was not as pronounced as in the analysis involving sex groups. This outcome is consistent with the fact that the second function was not statistically significant when vocabulary and reading comprehension were the independent variables.

As may be seen in Figure 2a, there were modest departures from parallelism in the profiles of means on vocabulary and reading comprehension for ethnic groups in quantitative fields and in mixed Q/V fields. (Mean profiles for Whites are included in Figure 2a even though they were not involved in the discriminant analysis). For example, Blacks and Asian Americans in these areas tended to have somewhat higher relative standing on RC than on VO, while the opposite was true for Mexican American and other Hispanics, and for Puerto Ricans. When subgroups were defined in terms of sex and major area, profiles for men and women within each major area tended to be parallel. This set of results suggests that the pattern of relative development of reading comprehension versus vocabulary that is characteristic of majors in a given area does not tend to differ by sex, but that the pattern may tend to differ by ethnic group.

Placement of second function centroids for ethnic groups on the basis of the four-score analysis was not consistent with placement for the two-score analysis. This is a reflection of the anomalous pattern of coefficients (weights) for the four verbal item-type part scores on the second discriminant function in the ethnic group analysis. Lack of parallelism for the mean profiles of ethnic groups, by major area, on the two reading comprehension component item types is evident in Figure 2a (lower frame).

Neither the particular pattern of ethnic-group differences reflected in Figures 2 and 2a, nor the "same-sign pairing" of vocabulary and reading comprehension item-type part scores (antonyms with reading passages versus analogies with sentence completions) was anticipated.

MDA Involving Quantitative Test and Analytical Test Part Scores

Results of the MDA involving the three quantitative item-type part scores, and the MDA involving the two analytical item-type part scores, are summarized in Table 5. At least two significant functions were obtained in each analysis. Second-function subgroup differences were more pronounced for analytical item types than for quantitative item types.

MDA involving quantitative part scores. Regardless of the groups involved, the principal discriminant function of the three quantitative item-type part scores was defined primarily by the quantitative comparison item-type part score (based on 30 items) and the regular mathematics part score (based on 20 items). The principal function accounted for a higher proportion (94 to 97 percent) of the among-groups variance than did the principal function in the verbal part-score analysis. And the second function contrasted performance on data interpretation (10 items) with performance on one or both

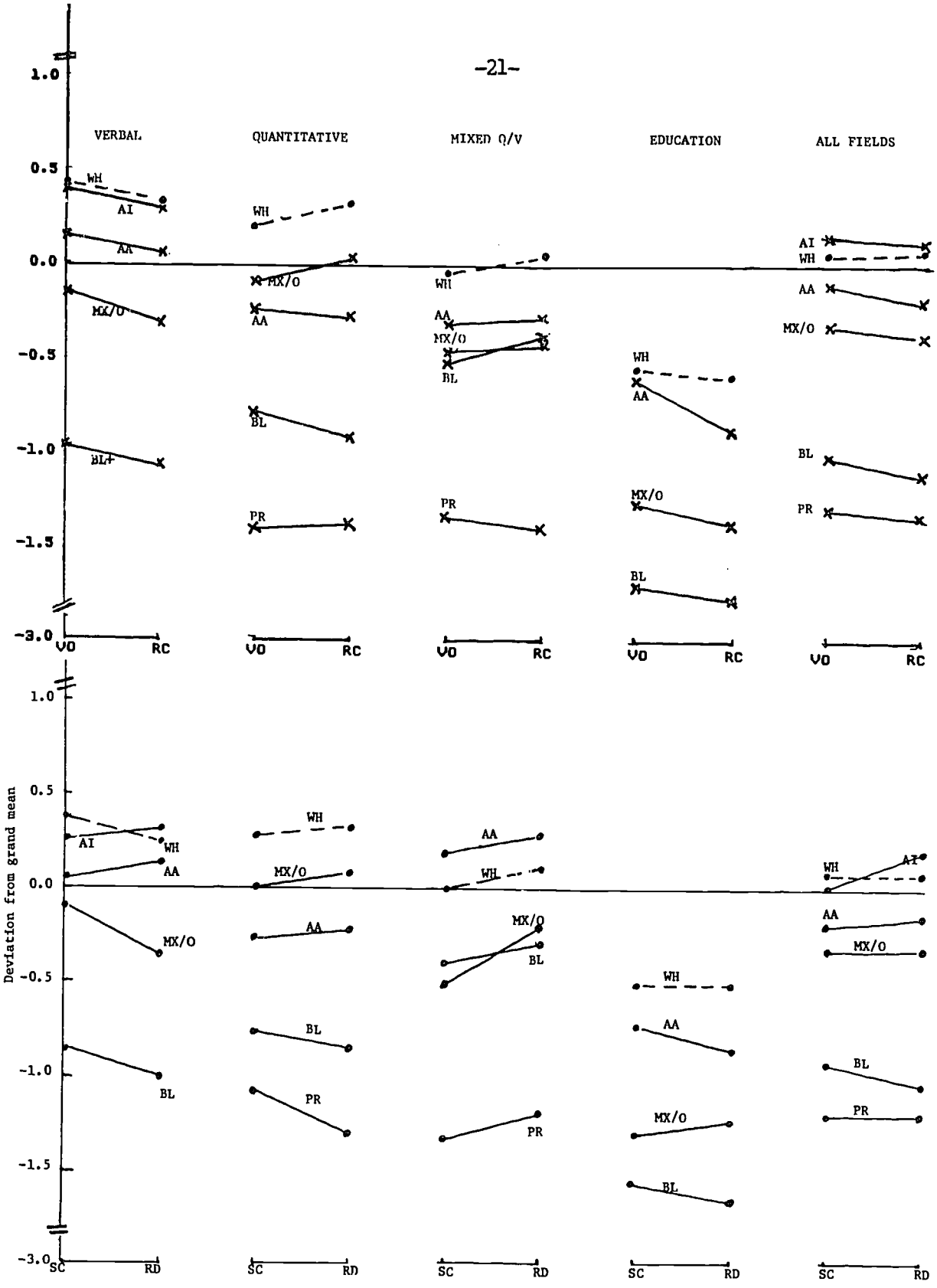


Figure 2a. Profiles of mean vocabulary and reading comprehension scores (upper), and sentence completion and reading passage scores (lower), by ethnic group and broad undergraduate major area

Table 5

Standardized Diacriminant Function Coefficienta for Quantitative Item-Type Part Scorea and Analytical Item-Type Part Scorea: Sex-by-Major-Area and Ethnic-by-Major Subgroup Analyaea

	Criterion groupa for analyaia			
	Ethnic groupa by major area (20 groupa)		Sex by major area (8 groupa)	
Part scorea employed	Principal function	Second function	Principal function	Second function
Quantitative Teat*				
(QC) Quantitative Compariaon	.50	-.56	.52	.21
(RM) Regular Mathematica	.45	-.39	.51	-.96
(DI) Data Interpretation	.23	1.18	.10	1.05
Percent variance	94.2	4.3	97.0	2.6
Significance	p <.001	p <.001	p <.001	p <.001
Analytical Teat				
(AR) Analytical Reaaoning	.80	-.74	.84	-.73
(LR) Logical Reaaoning	.35	1.04	.29	1.11
Percent variance	90.5	9.5	76.0	24.0
Significance	p < .001	p <.001	p <.001	p <.001

*In the quantitative teat analysia, a third diacriminant function waa aignificant ($p < .001$). It accounted for only 0.4 percent of the total variance and aerved primarily to diatinguiah groups with high acorea on RM and DI relative to scorea on QC (primarily malea in education) from groupa with low RM and DI relative to QC (primarily femalea in verbal majors).

of the remaining item types.

However, there were some differences in outcomes involving the second function. For example, for the sex-by-major-area groups, the second function contrasted performance on data interpretation (primarily, weight 1.05) and quantitative comparisons (weight .21) with performance on regular mathematics (-.96). For ethnic groups classified by area, the second function contrasted data intrepretation scores (weight 1.18) with scores on the other two quantitative item types (-.56 for QC and -.39 for RM).

(Group discriminant-score centroids). Centroids of groups on the first and second discriminant functions of the three quantitative item types (QC, RM, and DI) are shown in Figure 3 (sex-by-major in the left portion of the figure, and ethnic-group-by-major in the right portion). Sex differences were primarily in level of quantitative ability (note separation of sex-group centroids within major areas by the principal function). Very modest sex differences in performance on DI (data interpretation) and QC (quantitative comparison) relative to RM (regular mathematics) are suggested by the separation of the second-function means of males and females in three of the four major areas. In all major areas except the quantitative area, males tended to score somewhat higher than females on this function. The nature of the sex differences (which are subordinate to major-area differences) is indicated by the profiles of quantitative part-score means in Figure 3a (upper frame).

For the minority ethnic groups classified by field, the second function differentiated majors in verbal fields from those in quantitative fields. The verbal-area majors were high on data interpretation relative to quantitative comparison and regular mathematics, while the opposite pattern was characteristic of majors in quantitative fields (see mean profiles in Figure 3a, lower section). Puerto Ricans in quantitative and mixed quantitative/verbal (Q/V) fields were lowest on this function. American Indians in verbal fields, and Asian Americans in verbal fields and mixed Q/V fields were highest. Departures from parallelism in the quantitative part-score profiles of students, by major area, were more evident for ethnic groups than for males and females (see Figure 3a).

MDA involving analytical part scores. The pattern of weights for analytical reasoning and logical reasoning item-type part scores was similar on both the principal and the second discriminant functions in both the analysis by sex and major area and the analysis by ethnic group and major area (see Table 5). The principal function, representing level of general analyti0 cal ability, was defined more by the predominant analytical reasoning item type (38 items) than by the logical reasoning item type (12 items). The second function contrasted performance on the analytical reasoning (AR) and logical reasoning (LR) item-type part scores. The second function accounted for almost 25 percent of the total variance in the analysis involving classification by sex and field, and approximately 10 percent in the ethnic-group-by-field analysis. Thus, the amount of additional information about group differences provided by separate treatment of analytical ability item types was greater than that provided by treating verbal or quantitative item types separately.

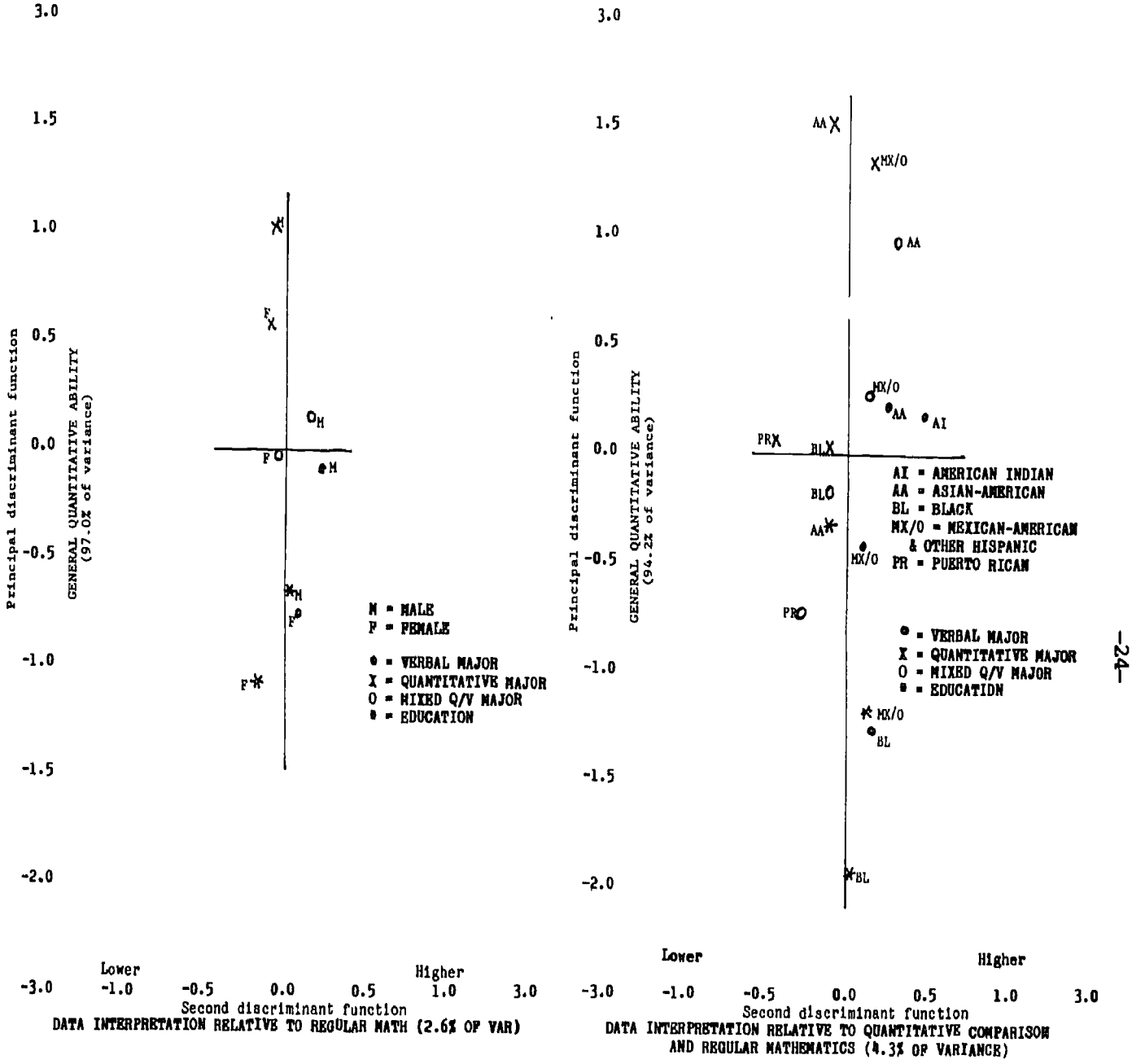


Figure 3. Centroids of sex-by-major subgroups (left), and ethnic-by-major subgroups (right): Discriminant functions of quantitative comparison, regular mathematics, and data interpretation item-type part scores

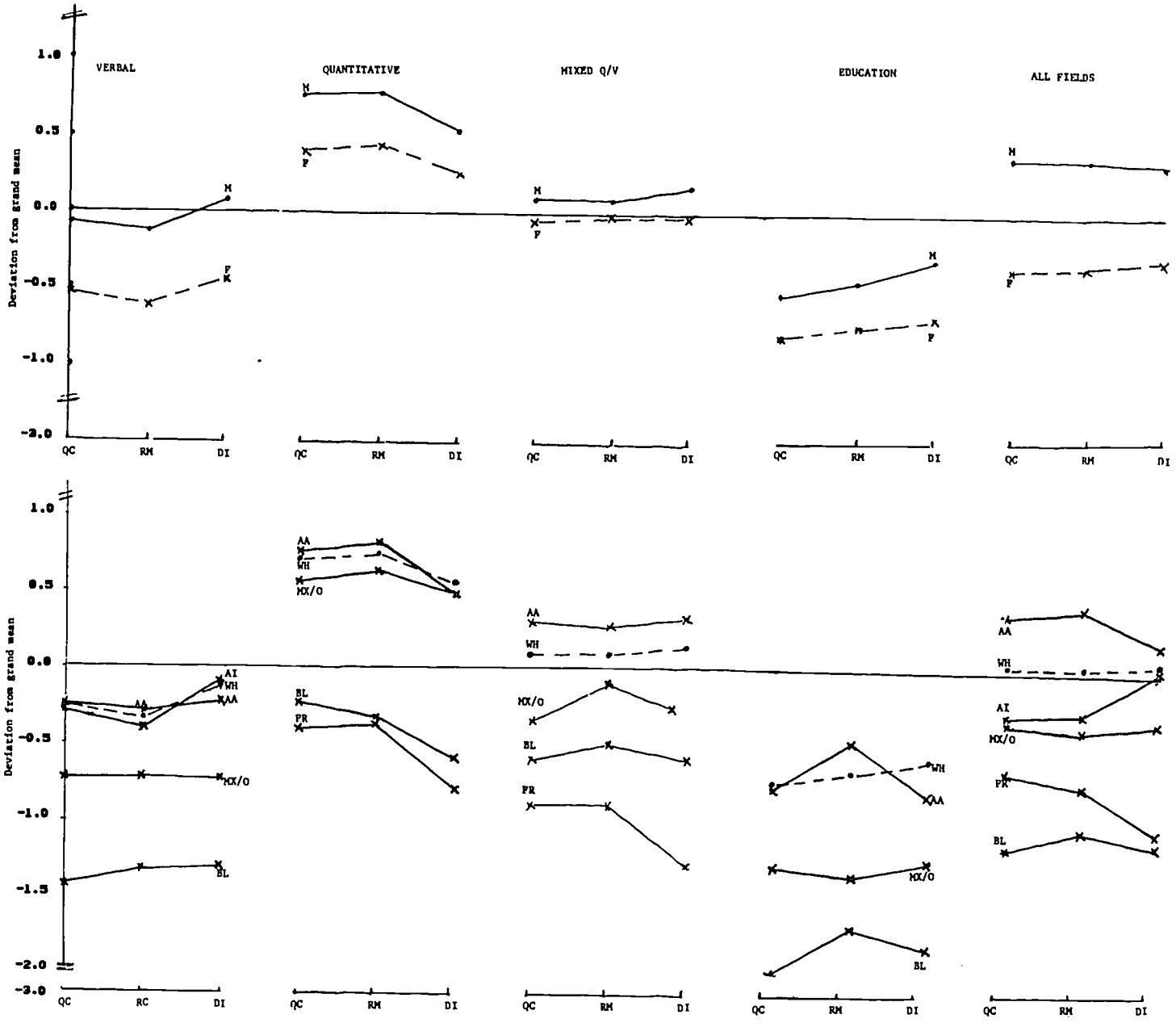


Figure 3a. Profiles of mean scores on quantitative comparison, regular mathematics, and data interpretation part scores, by sex and undergraduate area (upper), and by ethnic group and undergraduate area (lower).

(Group discriminant-score centroids). Figure 4 shows group centroids from the sex-by-field analysis (left portion of the figure) and the ethnic-group-by-field analysis (right portion of the figure). In the sex-by-field analysis, the second function sharply differentiated majors in verbal fields (with higher logical reasoning than analytical reasoning scores) from majors in all other fields who had logical reasoning scores relatively lower than analytical reasoning scores (see Figure 4a). Relatively modest second-function sex differences are also evident—in each major-area, males had slightly higher scores than females on the second function.

Ethnic group second-function differences, as well as field differences, are suggested by the location of centroids for ethnic group members classifed by major areas (right portion of Figure 4). The second-function mean for Asian Americans in verbal majors, for example, was considerably lower than that for American Indian and Mexican American/other Hispanic verbal majors, reflecting the fact that Asian Americans, alone among the ethnic groups, had lower relative standing on logical reasoning than on analytical reasoning, as shown in Figure 4a. Other instances of lack of consistency in ethnic-group profiles within major areas may be seen in Figure 4a. For example, Blacks in all major areas tended be higher in logical reasoning than in analytical reasoning, inconsistent with major-area trends in some instances.

The Relationship of GRE Item-Type Part Scores to SR-UGPA

This section presents findings bearing on the relative predictive role of item-type part scores within each of the three GRE General Test ability measures for various subgroups. All analyses were based on data that were pooled by undergraduate major area, after within-department z-scaling, for selected groups: males, females, Whites, Blacks, Asian Americans, three Hispanic groups collectively (Mexican Americans, Puerto Ricans, and other Hispanics), and for all ethnic minorities including American Indians and individuals reporting an unclassified "other" ethnic-group membership (MIN-T).

For each of these classifications, pooled within-department intercorrelations were computed for item-type part scores, total scores, and self-reported undergraduate GPA (SR-UGPA), the academic performance criterion. Multiple correlation coefficients for various combinations of part scores and total scores were computed. Coefficients for the part-score/total-score composites were compared with corresponding coefficients for the total score composite (V*Q*A*).

Verbal item-type part scores. Table 6 shows pooled within-department correlations (r) between SR-UGPA and (a) V*-verbal total score and (b) verbal item-type part scores (VC, RC, ANT, ANA, SC, and RD). The number of items making up each score is also shown. To reiterate a point made earlier, generally speaking, shorter item-type subtests are less reliable than longer subtests. If two tests of different length are measuring the same ability, the longer subtest will tend to have greater predictive validity than the shorter subtest. If the shorter subtest has equal or greater validity, factors other than differences in reliability must be considered in explaining the outcome.

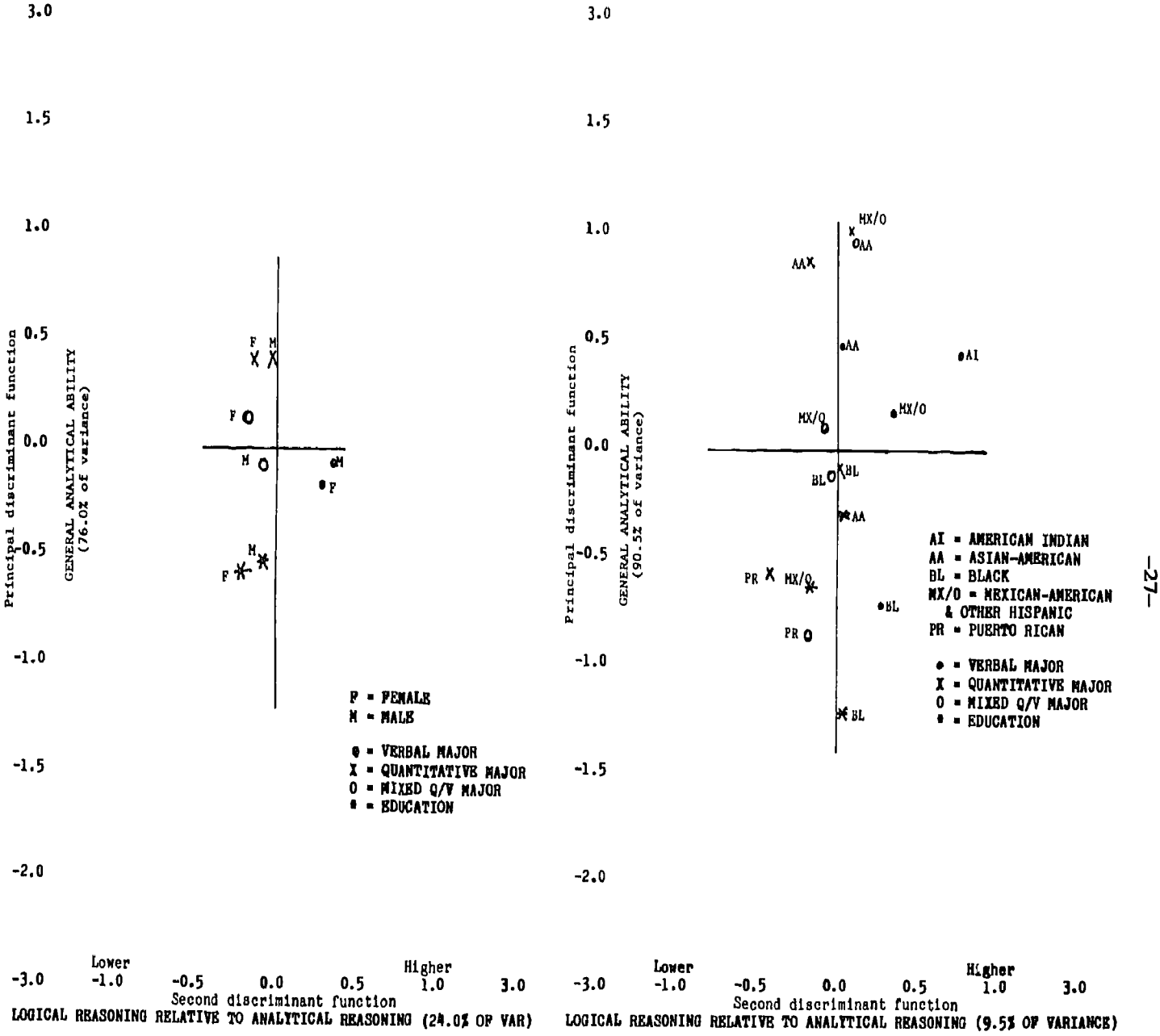


Figure 4. Centroids of sex-by-major subgroups (left), and ethnic-by-major subgroups (right): Discriminant functions of analytical reasoning and logical reasoning item-type part scores

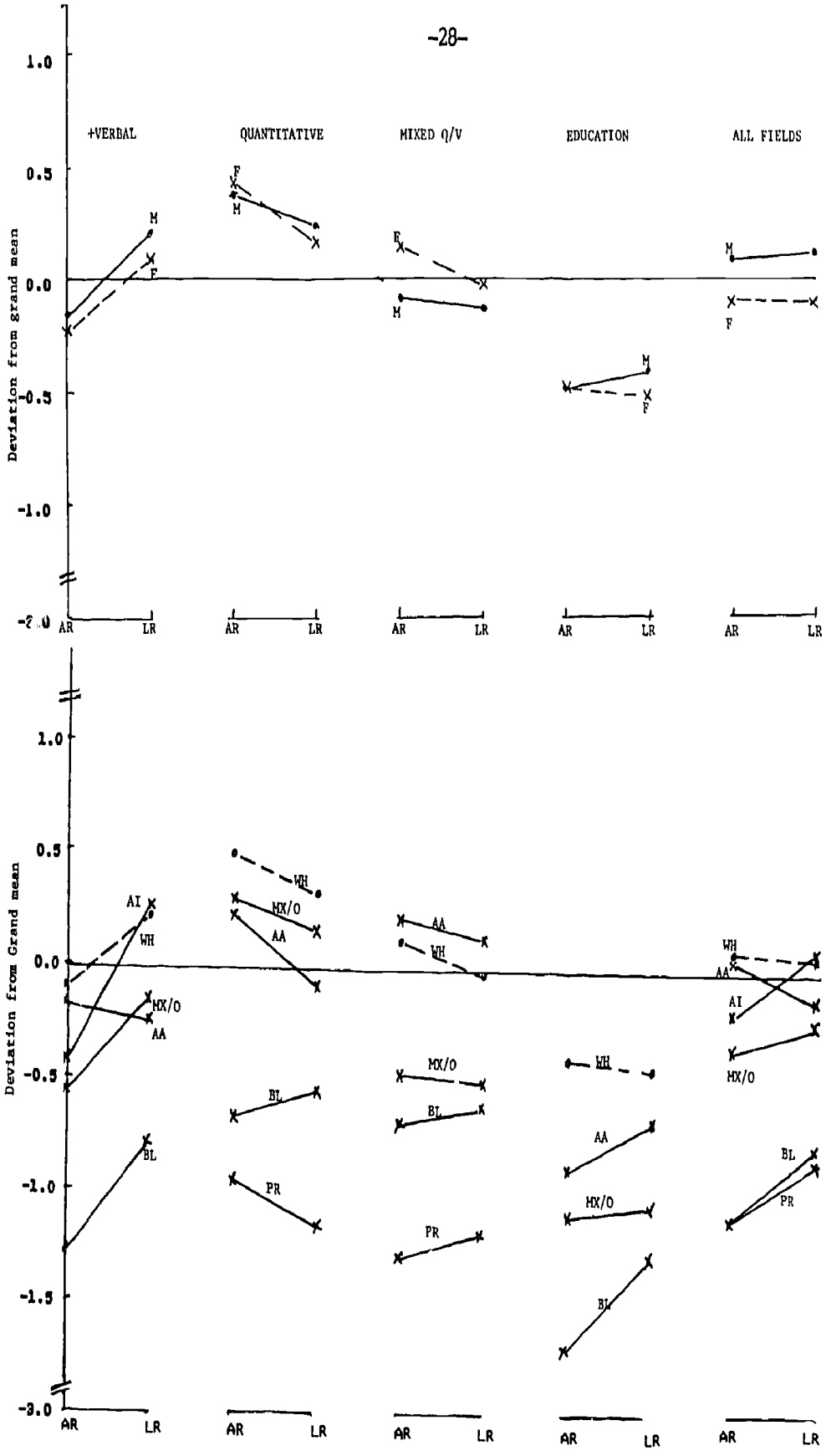


Figure 4a. Profiles of means on analytical reasoning and logical reasoning part scores, by sex and undergraduate area (upper), and by ethnig group and undergraduate area (lower)

Table 6

Simple Correlations with SR-UGPA of V* Total Score and Verbal Test Item-Type Part Scores, and Multiple Correlations for Designated Part-Score/Total-Score Sets, for Selected Demographic Subgroups, by Undergraduate Major Area

Area/Group	(N)	V* r (76)#	VO r (40)	RC r (36)	ANT r (22)	ANA r (18)	SC r (14)	RD r (22)	VO,RC Q*A* R(a)	ANT,ANA SC,RD,Q*A* R(b)	V*Q*A* R(c)	Diff b - a	Diff b - c	Diff a - c
Total sample	9303	31	26	30	23	24	26	27	37	37	36	00	01	00
Female (tot)	4715	34	30	33	25	28	27	30	38	39	38	01	01	00
Male (tot)	4588	27	23	27	20	20	24	24	36	36	35	00	01	01
WH (tot)	7900	29	24	27	21	22	23	26	34	34	34	00	00	00
MIN-T	1133	32	27	32	22	27	29	28	42	42	41	00	01	01
BL (tot)	361	40	35	37	23	40	35	32	45	48	45	03	03	00
HISP-T	374	24	19	24	16	17	21	22	31	31	31	00	00	00
AA (tot)	346	28	21	31	19	19	28	27	44	45	43	01	02	01
Verbal (tot)	2354	38	32	37	28	30	33	33	40	40	39	00	01	01
Female	1337	40	35	38	31	33	32	34	42	42	42	00	00	00
Male	1017	35	28	36	24	27	24	32	38	39	37	01	02	01
WH	1954	35	29	34	24	27	28	31	36	36	36	00	00	00
MIN-T	308	41	37	39	31	37	42	32	44	47	43	03	04	01
BL	139	36	34	33	25	38	38	26	38	45	37	07	08	01
HISP-T	75	40	34	37	27	35	34	31	41	42	41	01	01	00
AA	70	35	27	39	23	27	44	30	47	52	44	05	08	03
Quant (tot)	3037	27	23	27	21	20	22	26	39	39	39	00	00	00
Female	842	31	26	31	24	23	22	31	40	40	39	00	01	01
Male	2195	26	22	26	20	18	21	24	39	39	39	00	00	00
WH	2477	24	21	24	19	17	19	22	35	35	35	00	00	00
MIN-T	457	30	22	32	17	22	23	32	47	47	46	00	01	01
BL	113	48	44	43	31	44	37	40	58	58	58	00	00	00
HISP-T	130	22	14	27	12	13	17	28	34	35	31	01	04	03
AA	204	22	12	29	11	12	20	30	49	49	45	00	04	04
Q/V (tot)	2276	27	22	26	19	19	21	24	33	33	33	00	00	00
Female	1097	31	26	29	23	23	22	29	36	36	36	00	00	00
Male	1179	23	18	21	15	15	19	19	32	32	31	00	01	01
WH	1986	27	22	26	20	19	20	25	33	33	33	00	00	00
MIN-T	234	21	18	18	15	17	20	14	35	36	35	01	01	00
BL	33	34	30	27	19	41	31	21	37	48	37	11	11	00
HISP-T	141	15	12	15	11	08	15	13	37	37	37	00	00	00
AA	47	33	32	27	29	29	33	14	45	47	45	02	02	00
Educ (tot)	1636	33	30	31	23	29	29	28	37	37	37	00	00	00
Female	1439	34	30	32	23	29	29	28	38	39	38	01	01	00
Male	197	35	35	29	29	29	30	23	38	39	37	01	02	01
WH	1483	32	28	30	22	27	28	25	35	35	35	00	00	00
MIN-T	134	41	36	37	24	37	34	31	45	47	45	02	02	00
BL	76	28	19	31	03	30	19	31	41	48	39	07	09	02
HISP-T	28	34	35	23	29	30	43	07	37	37	37	00	00	00
AA	25	70	68	62	62	61	56	55	72	72	71	00	01	01

Note. Entries are pooled within-department correlation coefficients or differences between coefficients without leading decimals. Verbal = English, history, sociology and political science; quantitative = chemistry, computer science, mathematics, electrical engineering, and economics; Q/V = biology and agriculture. WH = White; MIN-T = total minority (including American Indians); BL = Black; HISP-T = Hispanic total; AA = Asian-American. V* = raw verbal total score, z-scaled by test form; VO = vocabulary (ANT + ANA), RC = reading comprehension (SC + RD); ANT = antonyms, ANA = analogies, SC = sentence completion, and RD = reading passages. Q* and A* are raw quantitative and analytical ability total scores, z-scaled by test form. R(a), R(b), and R(c) are multiple correlation coefficients for designated combinations of part scores and/or total scores.

Number of items included in the designated score.

Multiple correlation coefficients (designated by R) are shown for several score composites: (a) = VO and RC with Q* and A*, (b) = ANT, ANA, SC, and RD with Q* and A*, and (c) = the V*Q*A* total score composite. The entries in the last three columns of Table 6 are differences between designated pairs of multiple correlation coefficients: R(b) and R(a), R(b) and R(c), and R(a) and R(c). Generally speaking, R(b), in which the four verbal item-type part scores are substituted for the V* total score, should be equal to or greater than either R(a), with vocabulary and reading comprehension substituted for V*, or R(c); R(a), in turn, should be equal to or greater than R(c), for the V*Q*A* total-score composite. One should keep in mind in evaluating the increments in multiple correlation that some incremental validity may be due to "overfitting" the data when item-type part scores are "best-weighted" to predict a given criterion. Corrections for shrinkage have not been made.

Based on the simple correlations of verbal scores with SR-UGPA, the relative validity of the vocabulary and reading comprehension item-type part scores tended to be relatively consistent across subgroups.

o In the bulk of subgroups, the simple correlation of reading comprehension (RC) with SR-UGPA was higher than that for vocabulary. In several groups, the RC part-score coefficient was actually higher than that for the V* total score.

o RC was a particularly strong predictor for majors in quantitative fields. The coefficient for RC was equal to the coefficient for the V* total score for male, female, and White quantitative majors.

o The RC coefficient was higher than that for the total verbal score (V*) in the all-minority sample (MIN-T), and in the samples of Hispanics (HISP-T) and Asian Americans. In these latter three subgroups, the coefficient for a part score based on only one of the reading comprehension item types (the 22-item reading passage part score) was greater than that for V*.

o For Blacks in quantitative fields, and for Black students without regard to field, the vocabulary score appeared to have a comparatively strong predictive role. The analogy item-type score appeared to be more predictive than the score based on the antonymns items. For the minority samples (other than Hispanics) in biology and agriculture (fields of mixed quantitative and verbal emphasis), and in education (except for Blacks), coefficients for vocabulary were as high as, or higher than, those for reading comprehension.

The pattern of simple correlations suggests that the vocabulary and reading comprehension item types tend to have differential validity, a finding consistent with expectations based on studies cited earlier. In several subgroups, higher multiple correlations were obtained when these two part scores were substituted for the V* total score in composites with the Q* and A* total scores. The increases in multiple correlation associated with use of the vocabulary and reading comprehension part-scores were quite modest. However, increments in validity occurred in several comparatively large samples.

For example, the coefficient for the VO,RC,Q*,A* composite was slightly higher than that for the V*Q*A* total score composite in several relatively large subgroups involving males: (a) all males, (b) males in verbal fields, (c) males in fields of mixed quantitative and verbal emphasis, as well as (d) the small sample of males in education. Slight incremental validity for the part scores was also observed for the total minority sample (N = 1,133) and smaller MIN-T samples in verbal and quantitative fields.

Increments in multiple correlation when the four verbal item-type part scores were used, instead of VO and RC, are shown in the table. If we take into account the greater potential shrinkage in multiple correlation coefficients involving six independent variables (as compared to three or four), there is little suggestion that the use of four verbal part scores rather than only vocabulary and reading comprehension part scores in predictive composites might result in more meaningful increments in validity.

Perspective regarding the relative predictive roles of vocabulary and reading comprehension part scores when treated in combination with Q* and A* total scores is provided in Table 7, which shows standard partial regression coefficients (beta weights) for VO, RC, Q*, and A*. Validity coefficients for VO and RC, and multiple correlation coefficients for the part-score/total score composite and for V*Q*A*, are included for easy reference. In evaluating the findings, primary emphasis should be placed on identifying patterns in the data—for example, consistencies and inconsistences in relative weights for vocabulary and reading comprehension across certain subgroups.

An atypically strong role for reading comprehension relative to vocabulary for majors in quantitative fields is indicated by the consistent, comparatively smaller size of the beta weight for the vocabulary subscore. In seven of the eight quantitative subgroups, this trend was present. And vocabulary was negatively weighted (indicating suppression effects) in three of the four minority samples of majors in quantitative fields.

Table 7 also provides evidence bearing on the relative contribution of total scores to prediction. The analytical ability total score (A*), for example, contributed relatively more to prediction in samples of minority students than in samples made up wholly or predominately of White students. Note, for example, trends in the beta weights for A*, by subgroup, and for majors in verbal fields and in fields of mixed quantitative and verbal emphasis (agriculture and biology).

Quantitative item-type part scores. Table 8 shows validity coefficients for quantitative comparison (QC), regular mathematics (RM), data interpretation (DI) item-type part scores; two anomalous negative validity coefficients are included in the table. Beta weights for the three quantitative part scores and V* and A* total scores in a predictive composite are shown. The amount by which the part-score/total-score multiple correlation differed from that for the V*Q*A* total-score multiple correlation is also shown for each subgroup.

The findings in Table 8 suggest that the criterion-related variance

Table 7

Correlation of GRE Vocabulary and Reading Comprehension Scores with SR-UGPA, and Their Relative Contribution in Predictive Composites with Q* and A* Total Scores, for Selected Groups, by Undergraduate Major Area

AREA/GROUP	N	Simple correlation			Beta weights				(R)	V*Q*A*	Diff
		V*	VO-40	RC-36	VO	RC	Q*	A*	a	b	(a - b)
Total sample	9303	31	26	30	08	15	19	05	37	36	01
Female	4715	34	30	33	12	16	17	03	38	38	00
Male	4588	27	23	27	04	13	22	05	36	36	00
WH	7900	29	24	28	08	14	17	03	34	34	00
MIN-T	1133	32	27	32	05	12	22	11	42	41	01
BL	361	40	35	37	13	12	17	12	45	45	00
HISP-T	374	24	19	24	02	12	19	06	31	31	00
AA	346	28	21	31	-07	16	26	16	44	43	01
Verbal (Tot)	2354	38	32	37	13	24	10	00	40	39	01
Female	1337	40	35	38	17	23	10	01	42	42	00
Male	1017	35	28	36	06	28	12	-02	38	37	01
WH	1954	35	29	34	11	24	09	-02	36	36	00
MIN-T	308	41	37	39	15	20	05	12	44	43	01
BL	139	36	34	33	19	14	01	13	38	37	00
HISP-T	75	40	34	37	18	22	-11	12	41	41	00
AA	70	35	27	39	-06	28	23	10	47	44	03
Quant (Tot)	3037	27	23	27	02	12	26	07	39	39	00
Female	842	31	26	31	05	16	22	08	40	39	01
Male	2195	26	22	26	01	10	29	06	39	39	01
WH	2477	24	21	24	03	10	24	06	35	35	00
MIN-T	457	30	22	32	-05	15	34	09	47	46	01
BL	113	48	44	43	23	03	24	21	58	58	00
HISP-T	130	22	14	27	-07	20	28	-06	34	31	03
AA	204	22	12	29	-20	19	35	15	49	45	04
Mixed (Tot)	2276	27	22	26	06	11	18	06	33	33	00
Female	1097	31	26	29	09	14	19	02	36	36	00
Male	1179	23	18	22	03	09	20	07	32	31	01
WH	1986	27	22	26	07	12	17	05	33	33	00
MIN-T	234	21	18	18	02	-02	24	16	35	35	00
BL	33	34	30	27	36	-39	28	24	37	37	00
HISP-T	141	15	12	15	-04	-02	31	12	37	37	00
AA	47	33	32	27	20	-01	11	28	45	45	00
Educ (Tot)	1636	33	30	31	12	13	15	05	37	37	00
Female	1439	34	30	32	12	14	19	02	38	38	00
Male	197	35	35	29	24	03	07	12	38	37	01
WH	1483	32	28	30	11	13	15	04	35	35	00
MIN-T	134	41	36	37	15	15	16	08	45	45	00
BL	76	28	19	31	-03	16	30	03	41	39	01
HISP-T	28	34	35	23	30	02	-12	22	39	37	02
AA	25	70	68	62	48	21	18	-05	72	71	01

Note. Entries are coefficients without leading decimals. Negative beta weights indicate suppressor effects. Verbal analyses were based on majors in English, history, sociology, and political science; quantitative analyses were based on majors in chemistry, computer science, mathematics, electrical engineering, and economics; mixed emphasis analyses were based on agriculture and biology majors. Group designations are WH = White; BL = Black;, HISP-T = Mexican-American, Puerto Ricans, and Other Hispanics; AA = Asian Americans; and MIN-T = Minority total (all ethnic minority groups, including American Indians). VO = Vocabulary (40 items) and RC = Reading Comprehension (36 items); V*, Q* and A* are raw total verbal, analytical, and quantitative scores, z-scaled by test form.

Table 8

Correlation of Quantitative Test Item-Type Part Scores with SR-UGPA, and Their Relative Contribution to Prediction in Composites with Total V* and A* Scores, for Selected Groups, by Undergraduate Major Area

Area/Group	N	Q*	Simple correlations QC-30	RM-20	DI-10	Beta weights QC	RM	DI	V*	A*	(R) (a)	V*Q*A* (b)	Diff (a) -
Total Sample	9303	31	27	26	20	11	09	03	20	05	36	36	00
Female	4715	31	27	25	20	09	08	03	25	03	38	38	00
Male	4588	32	29	27	21	13	10	04	14	06	36	36	00
WH	7900	28	25	24	18	10	09	02	20	04	34	34	00
MIN-T	1133	37	35	29	28	16	06	06	13	12	42	41	01
BL	361	38	35	29	30	08	05	09	22	12	45	45	00
HISP-T	374	28	29	21	14	23	03	-05	12	07	33	31	02
AA	346	39	34	30	31	15	06	12	07	17	43	43	00
Verbal (Tot)	2354	26	24	22	18	06	06	01	33	02	39	39	00
Female	1337	27	24	23	20	03	07	03	35	02	42	42	00
Male	1017	26	25	22	17	09	05	00	29	-00	37	37	00
WH	1954	28	25	24	18	05	07	00	31	-01	36	36	00
MIN-T	308	30	30	20	26	07	-05	06	30	12	43	43	00
BL	139	25	22	15	29	-06	-04	16	26	11	39	37	02
HISP-T	75	09	22	02	-15	Not reported: negative simple correlation						41	--
AA	70	38	33	23	42	13	-17	38	19	08	51	44	07
Quant (Tot)	3037	36	32	30	22	15	13	04	12	08	39	39	00
Female	242	33	28	29	21	11	12	05	18	09	39	39	00
Male	2195	38	34	31	23	18	13	05	10	07	39	39	00
WH	2477	32	28	27	19	14	12	03	12	07	35	35	00
MIN-T	457	44	38	36	32	17	13	11	07	11	46	46	00
BL	113	50	43	41	36	06	12	11	26	19	58	58	00
HISP-T	130	30	30	21	16	27	-00	-01	13	-02	32	31	01
AA	204	43	36	35	33	16	13	15	-02	16	45	45	00
Mixed emph (Tot)	2276	29	25	24	20	09	09	04	15	06	33	33	00
Female	1097	30	26	24	21	10	09	04	21	03	36	36	00
Male	1179	29	24	25	20	09	10	06	10	07	31	31	00
WH	1986	28	23	24	19	07	10	05	17	05	33	33	00
MIN-T	234	33	35	24	19	27	03	-02	-02	16	38	35	03
BL	33	29	42	09	10	47	-19	-12	01	19	48	37	11
HISP-T	141	35	36	26	22	27	08	03	-05	12	38	37	01
AA	47	31	30	24	16	10	03	-01	17	29	45	45	00
Educ (Total)	1636	30	28	24	20	12	05	01	23	05	37	37	00
Female	1439	32	30	25	21	14	06	02	23	03	38	38	00
Male	197	29	25	26	23	02	05	03	25	09	37	37	00
WH	1483	29	28	23	18	12	05	-01	22	05	35	35	00
MIN-T	134	37	33	30	29	11	04	09	25	08	45	45	00
BL	76	38	38	29	23	27	09	04	10	02	41	39	02
HISP-T	28	16	-03	25	27	Not reported: negative simple correlation						37	--
AA	25	48	54	34	42	35	-15	-01	59	-05	72	71	01

Note. Entries are correlation coefficients and standard partial regression (beta) weights without leading decimals. Negative beta weights indicate suppression effects. Verbal analyses are based on majors in English, history, sociology, and political science; quantitative analyses involve majors in chemistry, computer science, mathematics, electrical engineering and economics; mixed emphasis fields are agriculture and biology. WH = White; MIN-T = Minority total; BL = Black; HISP-T = Mexican-American, Puerto Rican, and other Hispanic; AA = Asian-American; Q*, V* AND A* = quantitative, verbal, and analytical ability total raw scores, Z-scaled by test form; QC, RM, AND DI are quantitative item-type part scores.

contained in the several quantitative item types tends to be almost completely reflected in their contribution to the total quantitative score. In most instances, for example, the validity coefficient for Q* total score was higher than that for any of the three quantitative component-item-types. In addition, the differences in item-type validity coefficients were consistent with the number of items involved in each. For example, the validity coefficient for data interpretation, a very short subtest, typically was lower than that for the longer, quantitative comparison and regular mathematics subtests.

Moreover, the relative contribution to prediction of the three quantitative item types, as reflected by the relative size of their validity coefficients and the associated beta weights, tended to be quite similar for most of the subgroups. Substituting the three quantitative part scores for the Q* total score did not provide higher multiple correlations (see last three columns) than those provided by the three total scores in any of the larger samples; increments occurred only in the smaller, minority samples.

This pattern of findings differs from that for vocabulary and reading comprehension, in which increments in multiple correlation occurred for several of the larger samples as well as for some of the smaller samples. Since the tendency to "overfit" the data is greater in analyses involving the three quantitative part scores than in those involving only two part scores, if "overfitting" alone accounted for increments in validity, more incremental validity would be expected in the quantitative part analysis than in the verbal part score analysis.

For certain of the detailed findings, there is no a priori interpretive rationale. For example, for the comparatively small minority samples in fields of mixed quantitative and verbal (Q/V) emphasis, the validity of the quantitative comparison item-type part score was about equal to or greater than that for the Q* total score. The validity coefficient for DI was anomalously negative (-.15) for Hispanics in verbal fields. DI variance was suppressed in three analyses involving Hispanics, and so on.

Analytical item-type part scores. Judging from the findings shown in Table 9, the analytical reasoning and the logical reasoning item types exhibited systematically different patterns of relationships with the SR-UGPA criterion.

(1) The relative validity of the AR and LR item types was not consistent across subgroups. LR items were typically more valid than AR items for females, while the opposite was true for males.

(2) For minority samples in verbal fields, the 12-item LR part score was more valid than the 50-item A* total score, and multiple correlations obtained when AR and LR item types were scored separately were higher than those for the V*Q*A* total-score composite.

(3) For all females, and for the large sample of females in education, the multiple correlation for the AR,LR,V*,Q* composite was higher than that

Table 9

Correlation of Analytical Item-Type Part Scores with SR-UGPA, and Their Relative Contribution in Predictive Composites with V* and Q* Total Scores, For Selected Groups, by Undergraduate Major Area

Area/Group	(N)	Simple correlation			Beta weights				(R)	V*Q*A*	Diff
		A*	AR-38	LR-12	AR	LR	V*	Q*	a	b	(a)-(b)
Total sample	9303	26	23	22	02	06	19	19	37	36	01
Female	4715	27	22	26	09	09	23	18	39	38	01
Male	4508	26	24	19	04	03	14	22	35	35	00
WH	7900	23	20	21	01	06	18	18	34	34	00
MIN-T	1133	34	32	27	09	08	12	23	42	41	01
BL	361	37	33	30	09	09	21	17	45	45	00
HISP-T	374	24	22	18	05	05	10	19	31	31	00
AA	346	36	34	29	11	13	04	26	44	43	01
Verbal (Total)	2354	24	19	24	-02	-06	31	11	40	39	01
Female	1337	26	21	26	-01	-06	33	11	42	42	00
Male	1017	22	18	22	-04	-05	28	13	37	37	00
WH	1954	19	15	19	-03	03	30	10	36	36	00
MIN-T	308	35	28	37	02	20	25	07	45	43	02
BL	139	31	26	33	02	18	25	02	39	37	02
HISP-T	75	24	15	35	05	17	30	-09	42	41	01
AA	70	37	29	40	-01	27	25	10	48	44	04
Quantitative (Total)	3037	29	23	26	05	06	11	26	39	39	00
Female	842	29	23	31	01	16	14	22	40	39	00
Male	2195	29	27	20	05	02	09	29	39	39	00
WH	2477	25	22	21	03	06	11	24	35	35	00
MIN-T	457	36	34	25	10	05	06	33	46	46	00
BL	113	48	46	27	20	01	27	22	58	58	00
HISP-T	130	21	18	18	-03	01	07	28	31	31	00
AA	204	36	34	28	12	14	-06	34	46	45	01
Mixed emphasis (Total)	2276	24	22	18	05	04	15	19	33	33	00
Female	1097	24	21	21	-02	07	19	20	36	36	00
Male	1179	24	23	14	08	01	10	20	32	31	01
WH	1986	23	21	18	03	05	16	18	33	33	00
MIN-T	234	29	29	14	17	-01	01	24	35	35	00
BL	33	32	26	31	07	19	20	03	39	37	02
HISP-T	141	26	26	12	13	00	-05	31	37	37	00
AA	47	40	42	12	34	-07	26	05	48	45	03
Education (Total)	1636	28	23	26	00	08	21	16	37	37	00
Female	1439	28	23	26	-02	08	21	19	39	38	01
Male	197	30	36	26	05	09	23	08	37	37	00
WH	1483	26	22	26	-00	09	19	16	36	35	01
MIN-T	134	35	31	28	07	06	25	17	45	45	00
BL	76	27	21	23	01	06	08	31	40	39	01
HISP-T	28	30	31	06	24	-12	37	-16	41	37	04
AA	25	55	50	56	-19	17	59	21	72	71	00

Note. Entries are correlation coefficients and standard partial regression (beta) weights without leading decimals. Negative beta weights indicate suppression effects. Verbal analyses are based on majors in English, history, sociology, and political science; quantitative analyses involve majors in chemistry, computer science, mathematics, electrical engineering, and economics; mixed emphasis fields are agriculture and biology. WH = White; MIN-T = Minority total; BL = Black; HISP-T = Mexican-American, Puerto Rican, and other Hispanic; AA = Asian-American. Q*, V* AND A* = quantitative, verbal, and analytical ability total raw scores, Z-scaled by test form; AR and LR are analytical ability item-type part scores.

for V*,Q*,A*, the total-score composite.

(4) For males and for minorities in quantitative fields, validity coefficients tended to be higher for the AR than for the LR part score.

It appears that the contribution of the two analytical ability item types to prediction was not consistent with their representation in the total analytical ability score. Witness, for example, the predictive strength of the the shorter 12-item LR subtest relative to that of the 38-item AR subtest.

Observed versus Expected Performance for Subgroups

Table 10 shows the observed mean within-department standing of subgroups on the SR-UGPA criterion variable, by undergraduate major area. By design, the expected within-department z-scaled SR-UGPA mean for all students, without regard to subgroup membership, was zero. The means for subgroups indicate the average deviation of their SR-UGPAs from departmental averages in departmental standard deviation units. For example, females in quantitative majors had a mean of 0.02, indicating SR-UGPAs averaging .02 standard deviations above the departmental mean for all students. For males in education, mean SR-UGPA was -0.22 (0.22 standard deviations below departmental means, on the average), and so on.

Also shown in Table 10 are two "observed mean minus expected mean" residual values. The first was obtained using a "total-score estimate" as the expected mean—that is, using general regression equations, by undergraduate major area, with GRE V*, Q*, and A* as the predictors. The second residual value was obtained using a "part-score estimate"—the expected mean was based on comparable general regression equations using the best-weighted set of GRE item-type part scores. Negative entries in the residual columns indicate that the observed mean SR-UGPA was lower than expected, while positive entries indicate the opposite.

The four sets of general major-area regression equations that were used to estimate SR-UGPA, in z-scaled form, were as follows:

Verbal area:
$$.33V^* + .11Q^* + .02A^* = 0.00$$
$$\underline{.12VO + .23RC} + \underline{.05QC + .06RM + .01DI} - \underline{.03AR + .05LR} = 0.00$$

Quantitative area:
$$.12V^* + .26Q^* + .08A^* = 0.00$$
$$\underline{.02VO + .11RC} + \underline{.15QC + .13RM + .04DI} + \underline{.05AR + .06LR} = 0.00$$

Mixed Q/V area:
$$.16V^* + .18Q^* + .07A^* = 0.00$$
$$\underline{.06VO + .11RC} + \underline{.09QC + .09RM + .04DI} + \underline{.05AR + .04LR} = 0.00$$

Education:
$$.23V^* + .15Q^* + .05A^* = 0.00$$
$$\underline{.11VO + .12RC} + \underline{.12QC + .05RM + .00DI} + \underline{.01AR + .08LR} = 0.00$$

For every subgroup, the discrepancy between observed standing and expected standing, when the seven item-type part scores were used as predict-

Table 10

Mean Residuals (Observed minus Expected SR-UGPA) for Subgroups, When Expected SR-UGPA was Based on General Regression Equations Using (a) GRE Total Scores and (b) GRE Item-Type Part Scores: By Undergraduate Major Area

Group/Major Area	N	Observed mean z-scaled SR-UGPA	Observed minus expected mean	
			Total-score estimate	Part-score estimate
Female				
Verbal	1337	0.000	0.022	0.024
Quantitative	842	0.020	0.070	0.076
Mixed Q/V	1079	0.020	0.039	0.039
Education	1439	0.030	0.042	0.042
Male				
Verbal	1017	0.000	-0.034	-0.035
Quantitative	2195	-0.000	-0.021	-0.023
Mixed Q/V	1179	-0.020	-0.037	-0.036
Education	197	-0.220	-0.314	-0.305
White				
Verbal	1954	0.080	0.040	0.042
Quantitative	2477	0.050	0.021	0.020
Mixed Q/V	1986	0.040	0.029	0.028
Education	1483	0.030	0.018	0.018
Black				
Verbal	139	-0.850	-0.434	-0.459
Quantitative	113	-0.260	-0.165	-0.158
Mixed Q/V	33	-0.680	-0.321	-0.326
Education	76	-0.360	-0.176	-0.178
HISP-T				
Verbal	75	-0.190	0.022	0.040
Quantitative	130	-0.150	-0.049	-0.045
Mixed Q/V	141	-0.190	-0.117	-0.115
Education	28	-0.530	-0.243	-0.250
AA				
Verbal	70	-0.230	-0.198	-0.194
Quantitative	204	-0.130	0.010	0.018
Mixed Q/V	47	-0.460	-0.422	-0.407
Education	25	-0.100	-0.097	-0.092

Note. Means of subgroups on departmentally standardized (z-scaled) SR-UGPA indicate indicate average deviation of subgroups from departmental means in departmental SR-UGPA standard deviation units. Expected z-scaled values were generated using general regression equations (for students within each of the four major areas) using (a) GRE V*, Q*, and A* total scores, and (b) the "best set" of GRE item-type part scores. Differences between observed and expected means based on "total-score" and "part-score" equations are tabled.

ors was essentially the same as the discrepancy when the V*, Q*, and A* total scores were used as predictors. For example, when the expected SR-UGPA mean was based on the total scores, for females in verbal majors the observed SR-UGPA mean was better than that expected by .022 z-scaled units. The comparable residual value, when expected SR-UGPA mean was based on the part scores, was .024 units. Mean SR-UGPA for Blacks in verbal areas was lower than expected by .434 z-scaled units using total scores and by .459 units using item-type part scores; for Asian Americans in quantitative fields, performance was higher than expected by .010 z-scaled units using total scores and by .018 units using item-type part scores; and so on.

Summary

Dimensionality of Group Differences

For each GRE ability measure, the subgroups under consideration were found to be significantly differentiated along both a general ability dimension (defined by the principal discriminant function of item-type part scores, all positively weighted) and a secondary, bipolar dimension (defined by a second, significant discriminant function of positively and negatively weighted part scores).

The amount of information about group differences provided by the separate treatment of item-type part scores (that is, the percentage of variance accounted for by the second discriminant function—variance uncorrelated with that of the principal general ability function corresponding to the total ability score) was greater for analytical and verbal item-types, respectively, than for quantitative item-types.

(1) For sex-by-major criterion groups, percentages of variance accounted for by the second discriminant function were (a) 24 percent, by a function contrasting logical reasoning items with analytical reasoning items, and (b) 14 percent and 16 percent for functions involving two sets of verbal item-type part scores, both contrasting reading comprehension item-types with vocabulary item types, as compared to (c) about 3 percent for a second function of the three quantitative item types, in which data interpretation items were contrasted with regular mathematics items. Sex differences were primarily with respect to level of quantitative ability, but modest sex differences were also present in level of performance on data interpretation relative to performance on the other quantitative item types.

(2) The pattern of second-function findings for the minority ethnic-by-major subgroups was generally similar to the foregoing, but there were some differences. For example, the second function typically accounted for a somewhat smaller percentage of variance—about 10 percent in the case of analytical part scores, 3 percent and 7 percent for the sets of verbal part scores, and 4 percent for quantitative part scores. In analyses involving verbal item types, the second function was not significant when vocabulary and reading comprehension scores were used. However, when antonym, analogy, sentence completion, and reading passage part scores were used, the second function was significant, but the coefficients for the two vocabulary and the

two reading comprehension component item types were not consistent in sign—reading passage and analogy items (one a vocabulary item type and the other a reading comprehension item type) were positively weighted, while antonym items (a VO itemtype) and sentence completion items (an RC item type) were negatively weighted.

Part-Score/SR-UGPA Correlations

Analyses Involving Verbal Item Types

In the majority of subgroups, the simple correlation of reading comprehension scores with SR-UGPA was higher than the simple correlation of vocabulary scores with SR-UGPA; in several groups, the RC part-score coefficient was actually higher than the coefficient for the V* total score. The reading comprehension part score was a particularly strong predictor in quantitative fields. For example, for male, female, and White quantitative majors, the coefficient for RC was equal to the coefficient for V*; the RC coefficient was actually higher than that for V* in the total minority sample, and in samples of Hispanics and Asian Americans. For Blacks in quantitative fields, and for Blacks generally, the vocabulary score had a comparatively stronger predictive role. In several subgroups (some comparatively large), very modest increments in multiple correlation were obtained when vocabulary and reading comprehension part scores were substituted for the V* total score in composites with Q* and A* total scores.

Analyses Involving Quantitative Item Types

Correlational findings with respect to quantitative part scores suggest that most of the criterion-related variance in the quantitative item types tended to be reflected through their contribution to the total quantitative ability score. For example, in most subgroups, the validity coefficient for Q* total score was higher than that for any part score, and the correlations of part scores with SR-UGPA tended to vary directly with the number of items included in the respective scores (that is, with their reliability). When quantitative part scores were substituted for the Q* total score in composites with V* and A*, only limited increments in validity were observed, primarily in several minority samples with relatively small Ns.

Analyses Involving Analytical Item Types

The analytical reasoning and logical reasoning item types exhibited differential patterns of validity. Validity coefficients for LR items typically were higher than those for AR items for females, while the opposite was true for males. For minority samples in verbal fields, the 12-item LR part score was more highly correlated with SR-UGPA than the 50-item A* total score. However, for males and for minorities in quantitative fields, coefficients for the AR score tended to be higher than those for the LR score. Substitution of AR and LR part scores for the A* total score resulted in modest increments in multiple correlation in several of the larger subgroups, and in a number of the smaller minority samples as well.

For the minority samples in verbal fields, but not for all Whites, or the predominantly White samples of males and females in verbal fields, the LR item-type part score was more valid than the AR score. Moreover, the LR score made a much stronger contribution to prediction than did the AR score when both were included in a battery with V* and Q* total scores. The less valid AR score served as a suppressor in analyses involving data for the total sample, the sample of Whites, and the predominantly White male and female samples, and in the sample of Asian Americans. The AR score was negligibly weighted in the other verbal-area samples.

Observed Versus Predicted SR-UGPA Means for Subgroups

Predicted SR-UGPA means for subgroups, based on general major-area regression equations using item-type part scores as predictors, were essentially the same as the means predicted by general major-area regression equations using V*, Q*, and A* total scores as predictors.

The observed SR-UGPA means for the samples of minority students were lower than expected, with the following exceptions: Hispanics in verbal undergraduate majors and Asian Americans in quantitative majors, on the average, performed [illegible] better than expected.

Discussion

Scores based on item types included in the current GRE verbal, quantitative, and analytical ability measures clearly appear to be providing some information that is not provided in the respective total scores.

(1) For each ability measure, subgroups differed significantly in relative performance on item type part scores (represented by a bipolar discriminant function, uncorrelated with the principal general ability function of item-type subtests).

(2) Subtests based on item types included in the general ability measures, especially those included in the verbal and analytical ability measures, exhibited systematically different patterns of correlations with undergraduate grades. These differences appear to be independent of statistical artifacts, such as reliability differences associated with subtest length. For example, the criterion-related validity of the 10-item logical reasoning subtest tended to equal or exceed that of the 38-item analytical reasoning subtest; the validity of the 36-item reading comprehension subtest tended to exceed that of the 40-item vocabulary subtest and in some instances was greater than that of the V* total score, and so on.

However, the findings do reflect, to some extent, the effects of statistical artifacts imposed by the use of GRE total and part scores not equated across the six different test forms taken by the examinees in this study. In the original study (Wilson, 1984), the criterion-related validity of the unequated GRE total scores was shown to be consistently lower than that of the GRE converted (equated) scores, in general samples by undergraduate field and

major area. Such a pattern, due by inference to error introduced by using unequated scores, was assumed to be present for the respective part scores. The use of unequated scores may also have affected findings regarding group differences in performance on GRE part scores. However, it was not feasible to assess the nature or the extent of any "form effects" that might be present in the findings.

It is also important to keep in mind that the study employed a self-reported undergraduate GPA criterion rather than a graduate-level performance criterion. The conclusions reached, therefore, should be thought of primarily as working hypotheses for test development and research.

With these limitations in mind, the study findings indicated that major-area differences were stronger and more systematic than ethnic-group or sex differences with respect to both (a) patterns of within-test performance on item-type part scores and (b) patterns of within-test part-score/SR-UGPA correlation.

(1) Major-area (and subgroup) differences were more pronounced for part scores based on analytical and verbal item types, than for part scores based on quantitative item types.

(2) Systematic major-area differences in patterns of correlations with SR-UGPA were more clearly evident for the vocabulary and reading comprehension part scores, and the analytical reasoning and logical reasoning part scores, than for the quantitative comparison, regular mathematics, and data interpretation part scores.

The study findings indicated that the use of item-type part scores, especially those involving verbal and analytical ability item types, resulted in some increase in criterion-related validity for individuals. SR-UGPA means for subgroups based on general major-area regression equations using item-type part scores as predictors were essentially the same as the predicted means based on comparable regression equations using the corresponding total scores. Thus, findings suggest (a) that using part scores as predictors may not result in better inferences regarding the probable relative standing of various subgroups on the SR-UGPA criterion than are provided by the total ability scores; but (b) that the part scores appear to hold out the possibility of modestly improved accuracy of prediction for individuals.

Conclusions and Implications

Continued interest in the potential contribution of subtests based on GRE item types, especially those included in the verbal ability and analytical ability measures, would seem to be warranted.

GRE Verbal Test

o With respect to the GRE vocabulary and reading comprehension subtests used for this study, comparatively strong major-field differences in relative performance were present.

o The reading comprehension subtest was the more valid subtest in the greater number of instances. This pattern has been found in undergraduate-level validity studies involving comparable reading comprehension and vocabulary subscores routinely reported for the Scholastic Aptitude Test. However, the results of this study, and of the earlier study, suggest a potentially useful measurement role for both vocabulary and reading comprehension.

o In several subgroups, some comparatively large, slightly higher multiple correlations were obtained when the GRE vocabulary and reading comprehension part scores were substituted for the V* total score in composites with Q* and A* total scores. The coefficient for the VO,RC,Q*,A* composite was higher than that for the V*Q*A* total-score composite for (a) all males, (b) males in verbal fields, (c) males in fields of mixed quantitative and verbal emphasis, (d) the small sample of males in education, (e) the total minority sample, and (e) smaller all-minority samples in verbal and quantitative fields.

The availability of GRE vocabulary and reading comprehension subtests based on combinations of item types such as those employed in this study, would permit the independent assessment of two dimensions of verbal ability that are judgmentally distinguishable. Moreover, these two components of the GRE verbal ability measure have been found to be factorially distinguishable in several studies (Powers & Swinton, 1981; Powers, Swinton, & Carlson, 1977; Rock, Werts, & Grandy, 1982). A contrary finding was reported in a more recent GRE factor analysis (Stricker & Rock, 1985). "[The conflicting findings] may be traceable to important differences between these [the earlier] investigations" in the samples involved, test content examined, or statistical methodology (p. 24).

Consideration might also be given to the potential value of a "reading comprehension" subtest defined exclusively by reading passage items, and a vocabulary subtest defined by the three "discrete" item-types. A GRE measure of reading comprehension, so defined, would have a high degree of face validity. Study findings suggest that it would also have quite useful predictive validity.

GRE Analytical Test

With respect to the analytical ability measure, in the earlier study of GRE part scores (Wilson 1984) it was concluded (p. 34) that "analytical reasoning items behave more like quantitative ability items, while logical reasoning items behave more like verbal ability items." Among other things, the AR item-type part score was found to be more highly correlated with quantitative item-type part scores than with verbal part scores, while the opposite was true for the LR item-type part score. And AR/LR correlations were lower than either AR/quantitative-part-score or LR/verbal-part-score correlations.

The findings of the present study extend those of the earlier study. They reinforce the general conclusion that the AR and LR item-types are

distinguishable along a verbal versus quantitative dimension.

o Majors in verbal fields were high in logical reasoning relative to analytical reasoning ability, and quantitative-area majors tended to be high on analytical reasoning relative to logical reasoning. The LR part score had higher criterion-related validity for verbal-area majors, while the AR score was more valid for quantitative-area majors. In addition:

o LR items were typically more valid than AR items for females, while the opposite was true for males; for all females, and for the large sample of females in education, the AR,LR,Q*,V* composite had a higher correlation with SR-UGPA than did the total-score composite.

o For minority samples in verbal fields, the 12-item LR part score was more valid than the 50-item A* total score; for males and minorities in quantitative fields, the AR part score was more valid than the LR part score. And study results suggest that the logical reasoning item-type part score may be a relatively stronger predictor for verbal-area minority students than for the majority of verbal-area majors.

Given such differences in findings for the two analytical ability item-type part scores, it seems reasonable to conclude that subtests based on the analytical and logical reasoning item types would provide more information—information potentially useful for prediction and for diagnosis—than is provided by the summarization of performance on these item types in a total analytical ability score. The argument for separate treatment of these two item types, based on the findings of this study, is reinforced by results of a factor analysis (Stricker & Rock, 1985) from which the investigators concluded, among other things, that the analytical ability item types under consideration have relatively little in common with each other.

GRE Quantitative Test

Study findings do not permit strong conclusions regarding the potential contribution of part scores based on quantitative item types. There were significant major-field differences in performance on data interpretation item types relative to performance on the other quantitative item types. Verbal-area majors, for example, performed better on DI items than on the other item types, while the opposite tended to be true for quantitative-area majors. There were also modest sex differences in performance on data interpretation relative to the other quantitative item types.

However, substitution of the three quantitative item-type part scores for the Q* total quantitative score in a battery which included V* and A* total scores did not lead to increments in multiple correlation in the larger subsamples. This pattern of findings differs from that for vocabulary and reading comprehension, and analytical reasoning and logical reasoning, in which, under comparable conditions, increments in multiple correlation occurred for several of the larger samples as well as for some of the smaller ones. The tendency to "overfit" the data is greater in analyses involving the three quantitative part scores than in those involving only the two verbal

ability or analytical part scores. Thus, if "overfitting" alone accounted for observed increments in validity, then more (not less) incremental validity would be expected in the quantitative part-score analysis than in the verbal part-score or the analytical part-score analysis. The lack of incremental validity for quantitative part scores, in the circumstances, lends weight to the observed incremental validity for verbal ability and analytical ability part scores.

Implications

The findings that have been reviewed indicate a potentially useful role for GRE General Test item-type part scores. Based on the available evidence, item-type part scores hold out a promise of very modestly increased predictive validity over that provided by the total scores. Further research involving item types would contribute to better understanding of the constructs being measured by the general ability tests.

Given clearly identified subtests based on item types, test takers and test users would be afforded a more precise delineation of the functional skills that contribute to the respective general abilities. Subscores based on item types included in the current verbal and analytical measures (and possibly the quantitative measure as well) would seem to offer one potentially useful basis basis for broadening the scope of measurement of the GRE General Test, within if not beyond the strongly established verbal and quantitative domains.

However, available evidence regarding the predictive value of part scores is based solely on studies that have employed undergraduate-level GPA criteria. Confirmatory studies involving graduate-level performance criteria are needed. Moreover, the amount of added predictive validity to be expected by the use of item-type part scores is likely to be very modest. This expectation is consistent with general experience in prediction research. After two or three major predictors of given criteria are identified, it is difficult to find additional predictors that contribute more than marginally to the original battery. Experience with the analytical ability measure represents a case in point.

Adding any new score to the GRE General Test would inevitably pose problems for test development. Given the complexities of graduate-school admission settings, problems would also be involved in efforts to establish and maintain an adequate empirical basis for interpreting the new scores.

Consideration of the potential benefits associated with item-type part scores should be balanced by a thorough assessment of the potential costs of dealing with the problems that would be posed by their introduction.

References

Educational Testing Service (1984). GRE 1984-85 Information Bulletin. Princeton, NJ: Author.

Klecka, W. R. (1975). Discriminant analysis. In Nie, Norman H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H (eds.), Statistical package for the social sciences. NY: McGraw-Hill Book Co.

Powers, D. E., & Swinton, S. S. (1981). Extending the measurement of graduate admission abilities beyond the verbal and quantitative domain. Applied Psychological Measurement, 5, 141-158.

Powers, D. E., Swinton, S. S., & Carlson, A. B. (1977). A factor analytical study of the GRE Aptitude Test (GRE Board Professional Report 75-11P). Princeton, NJ: Educational Testing Service.

Ramist, L. (1981a). Validity of the SAT verbal subscores (Internal memorandum). Princeton, NJ: Educational Testing Service.

Ramist, L. (1981b). Further investigation of the validity of SAT verbal subscores (Internal memorandum). Princeton, NJ: Educational Testing Service.

Rock, D. A., Werts, C., & Grandy, J. (1982). Construct validity of the GRE Aptitude Test--An empirical confirmatory study (GRE Board Research Report GREB No. 78-1P & ETS RR-81-75). Princeton, NJ: Educational Testing Service.

Stricker, L. J., & Rock, D. R. (1985). Factor structure of the GRE General Test for older examinees: Implications for construct validity (GRE Board Research Report GREB No. 83-10R & ETS RR-85-9). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1984). The relationship of GRE General Test item-type part scores to undergraduate grades (GRE Board Professional Report GREB No. 81-22P & ETS RR-84-38). Princeton, NJ: Educational Testing Service.