DOCUMENT RESUME

ED 395 021 TM 025 038

AUTHOR Wilson, Kenneth M.
TITLE Population Differences in Speed Versus Level of GRE Reading Comprehension: An Exploratory Study. GRE Board Professional Report No. 84-09aP.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Graduate Record Examinations Board, Princeton, N.J.
REPORT NO ETS-RR-89-36
PUB DATE Dec 89
NOTE 82p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Correlation; English; Ethnicity; Graduate Study; Higher Education; *Language Proficiency; *Majors (Students); *Reading Comprehension; Reading Tests; *Scoring; Sex Differences; *Test Items; Verbal Tests
IDENTIFIERS *Graduate Record Examinations; Population; *Speed

ABSTRACT

Possible population differences in speed versus level of Graduate Record Examinations (GRE) reading comprehension scores were explored. The study used operational measures computed post hoc from item-level data in GRE files for a pre-October 1977 version of the verbal test in which 40 GRE reading comprehension (RC) items were included as a separately timed section administered under then current formula-scoring instructions. The "level" (formula) score was defined by performance on the first 20 RC items, completed by almost all examinees; and the "speed" (formula) score was defined by performance on the next 20, attempted by fewer examinees, and a formula score based on 28 odd-numbered discrete-verbal items was z-scaled in data for more than 21,000 examinees. Patterns of difference between correlated z-scaled means were analyzed for U.S. and non-U.S. examinees by sex, ethnicity, English as primary language, and four broad graduate major areas. Patterns of relative standing on the scores were generally consistent with the hypothesis that speed/level differences would be present in analyses by graduate major area and English proficiency status, but not in analyses by sex or ethnicity. An appendix contains three tables of correlations of subtests and GRE scores by examinee subgroup. (Contains 9 figures, 10 tables, and 62 references.) (Author/SLD)

GRE®
RESEARCH

Population Differences in Speed Versus Level of GRE Reading Comprehension: An Exploratory Study

Kenneth M. Wilson

December 1989

GRE Board Professional Report No. 84-09aP
ETS Research Report 89-36

Educational Testing Service, Princeton, New Jersey

Population Differences in Speed Versus Level of GRE Reading Comprehension: An Exploratory Study

Kenneth M. Wilson

GRE Board Report No. 84-09aP

December 1989

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, N.J. 08541

Acknowledgements

This study was made possible by the support of the Graduate Record Examinations Board. Foster Schoenthaler provided assistance in extracting data from GRE files. Nancy Petersen programmed and conducted a variety of basic statistical analyses during the course of the study. Critical reviews of various drafts of the report were provided by Thomas Donlon, Ronald Flaugher, Neal Kingston, Maria Pennock-Roman, and Charlotte Solomon. Nancy Parr contributed an editorial review. Joanne Farr provided helpful, patient assistance in preparing the manuscript. These contributions are acknowledged with appreciation.

ERRATA

Wilson, K. M. (1989). Population differences in speed versus level of GRE Reading Comprehension: An exploratory study (GRE Board Professional Report No. 84-09aP and ETS Research Report 89-36). Princeton, NJ: Educational Testing Service.

Page 67 (References).

Wilson, K. M. (1985a). Erroneously identified as ETS Research Report 84-23, should be identified as ETS Research Report 84-38.

Wilson, K. M. (1986b). Erroneously identified as ETS Research Report 84-38, should be identified as ETS Research Report 86-37.

Abstract

The concept of differences in "speed" versus "level" of reading comprehension is well established, and tests of reading ability frequently provide separate measures of "speed of reading with understanding" and "level of ability to read with understanding." This study was undertaken to explore possible population differences in speed versus level of GRE reading comprehension, using operational measures computed post hoc from item-level data in GRE files for a pre-October-1977 edition of the verbal test--that is, a version in which 40 GRE reading comprehension (RC) items were included as a separately timed section, administered under then current formula-scoring instructions.

The "level" (formula) score was defined by performance on the first 20 RC items (RC1), completed by almost all examinees, and the "speed" (formula) score was based on the second 20 items (RC2)--the number of RC2 items "attempted" ranged from 0 - 20. RC1, RC2, and a formula score based on 28 odd-numbered discrete-verbal items (DVodd) were z-scaled in data for more than 21,000 examinees tested in a scheduled administration of the GRE in October 1976. Patterns of differences between correlated z-scaled means (mean RC2 minus mean RC1) were analyzed for (a) U.S. examinees classified by sex, ethnicity, English-language communication status (English primary language, or EPL, versus English second language, or ESL), and four broad graduate major areas, and (b) non-U.S. examinees classified by sex, language status, and academic area.

It was considered plausible that speed/level differences would be present in analyses by graduate major area and by EPL/ESL status, but not in analyses by sex or by ethnicity. Patterns of relative standing on RC2 and RC1 were generally consistent with hypothesis: RC2 > RC1 (higher standing on speed than on level) for humanities and social science majors, and non-U.S EPL examinees; RC2 < RC1 for physical science majors and bioscience majors, and both U.S. and non-U.S. ESL examinees. For all U.S. ethnic groups studied (African American, Asian American, Hispanic American, and White American), the basic pattern was as expected, namely, RC2 = RC1, and this pattern tended to obtain for males and females. Criterion-related validity for RC1 and RC2 was explored, using self-reported undergraduate GPA (SR-UGPA). RC2 tended to be more highly correlated than RC1 with the SR-UGPA criterion, for subgroups of U.S. examinees, except for Hispanic examinees and ESL examinees. For these two partially overlapping subgroups, and for subgroups of non-U.S. examinees, coefficients for RC1 (level) were consistently larger than those for RC2 (speed).

The findings suggest that further exploration of the role of "speed" in measures of GRE reading comprehension, and in other GRE ability measures, is warranted.

Table of Contents

Table of Contents (Continued)

Table of Contents (Concluded)

Introduction

Tests of reading ability almost universally include a section designed to measure reading comprehension, and a section designed to measure vocabulary. In many instances, reading tests also provide a score for "reading rate" or speed of comprehension as well as a score for level or accuracy of comprehension. (For reviews of reading tests see the "Tests and Reviews: Reading" sections in various editions of Buros [e.g., 1965]).

The verbal section of the Graduate Record Examinations (GRE) General Test[a] includes (a) "reading comprehension" (RC) items (reading passages and sets of related questions) that, like comparable items on "reading tests," call for complex, discourse-level analysis, and (b) three "discrete-verbal" (DV) item types (antonym, analogy, and sentence-completion items or questions), so called because they provide limited context and involve word-level to sentence-level analysis only.

The GRE reading items are described as follows by the GRE Program (e.g., ETS, 1988):

> (The reading comprehension sets are intended) . . . to measure the ability to read with understanding, insight, and discrimination. This type of question explores the examinee's ability to analyze a written passage from several perspectives, including the ability to recognize both explicitly stated elements in the passage and assumptions underlying statements or arguments in the passage as well as the implications of those statements or arguments. Because the written passage upon which reading comprehension questions are based presents a sustained discussion of a particular topic, there is ample context for analyzing a variety of relationships . . . . (Examinees) are not expected to rely on outside knowledge, which (they) may or may not have, of a particular topic (pp. 30-31).

GRE antonym, analogy, and sentence-completion questions assess, respectively, (a) ability to identify words that are opposite in meaning from a stimulus word, (b) ability to discern relationships in pairs of stimulus words, and (c) ability to identify among several pairs of words or phrases the set that best completes a sentence from which two such pairs have been deleted. These items appear to be measuring lexical knowledge as well as aspects of ability to reason with words.

For purposes of the present study--an exploratory assessment of speed versus level of GRE reading comprehension--the foregoing descriptions are intended primarily to highlight the fact that the GRE reading comprehension sets are very explicitly designed to measure aspects of "functional ability to read with comprehension."

---

[a] See end of text for numbered notes.

It is considered useful, though not essential, to think of the discrete-verbal items as measuring accumulated word knowledge (as well as more general verbal reasoning abilities), in contrast to the functional reading skills measured by the reading comprehension items.

Implications of Timed Test Administrations

Historically, the GRE has been designed as a measure of "power," or level of ability to perform the tasks represented by various test items, not speed of responding. Time limits have been established for pragmatic administrative reasons, not to evoke a speed-of-response factor. The amount of time per section, based on experience, is judged to be sufficient to permit a majority of examinees in the general test-taking population to attempt (consider and evaluate alternative responses to) a majority of the test questions.

However, under practical administrative conditions, it is not possible to eliminate "speed" as a factor in GRE scores. Accordingly, the pragmatic response has been to standardize the speed component in separately timed sections of sucessive editions of the GRE. According to ETS guidelines, each such section should meet certain test-completion criteria, namely, (a) all examinees should be able to attempt 75% of the test items, and (b) at least 80% should be able to complete all of the test items.[2]

While test sections meeting the criteria outlined above are said to be unspeeded, it is evident that a substantial percentage of examinees are unlikely to be able to finish typical unspeeded versions of the GRE. Thus, even if the criteria are met approximately in every instance, it is reasonable to infer that each separately timed section of the GRE General Test is measuring to some extent both speed and level of ability to perform the tasks represented by the test items.

Rationale for Studying Speed versus Level of GRE Reading Comprehension

Prior to October 1977, in GRE verbal measure the RC and DV sets were presented in separately timed sections. The pre-October 1977 verbal measure consisted of a timed reading comprehension section made up, typically, of six reading passages and accompanying sets of questions (40 in all), and a timed 55-item section made up of a balanced representation of the three DV item types.

For purposes of the present study, it is assumed the scores earned by examinees on the timed GRE reading comprehension sections reflect to some extent differences in speed of reading with compre- hension and differences in level of ability to read with comprehension. There are models for obtaining scores for speed of comprehension and level of comprehension from single, "speeded" administrations of reading comprehension tests--that is, models that do not require differentially speeded subsections with special instructions.

One such model was used in this exploratory study to compute a "speed of comprehension" score and a "level of comprehension" score from item-level data in GRE files for examinees who took one pre-October 1977 form of the GRE General Test--that is, a version of the test with a timed reading comprehension section, a necessary condition for generating speed and level scores.[3] The scores were employed to explore theoretically plausible differences in speed versus level of GRE reading comprehension for selected GRE subpopulations. An exploratory assessment was also made of the relationship of the operational measures of speed and level to an external academic criterion--self-reported undergraduate grade point average (SR-UGPA).

Overview of Principal Assumptions and Hypotheses

Study design and procedures were guided by several working propositions, assumptions, or hypotheses, outlined briefly below.

1. GRE reading comprehension sets and GRE discrete-verbal sets are measuring somewhat different aspects of general verbal ability. Although closely related, RC and DV scores have different properties --for example, different patterns of correlation with external criteria, presence of major-area differences in patterns of performance, and so on. Empirical evidence bearing on the foregoing is provided in the following section.

2. Essentially all achievement or aptitude tests, including tests of reading comprehension, that are administered with time limits are to some extent measuring "speed" as well as "power," or level of knowledge, skill, understanding, and so on, in the domains sampled by the test items (e.g., Gulliksen, 1950; Lord, 1956). It follows that when administered with time limits that do not permit all examinees to complete all the test questions--an accepted administrative condition for GRE tests--a GRE reading comprehension test is to some extent measuring speed as well as level of ability to read with comprehension.

3. Viewed from the perspective of cognitive science, four complexly interrelated processes appear to be involved in reading, namely, word recognition, accessing semantic word information, sentence processing, and discourse analysis (e.g., Curtis & Glaser, 1983; Glaser, Lesgold, & Lajoie, 1985). Differences in speed of carrying out any one or all of these processes may affect total performance in reading. The comments of Curtis, Lesgold, and Lajoie (1985) on this point are cited, illustratively, as follows:

> If, during reading, part of the thinking capacity is given over to word recognition, less capacity may remain for joining con- cepts that need to be interrelated in the reader's mind . . .. That is, when word recognition is slow, comprehension processes become resource limited . . . whereas faster recognition allows more effort to be directed to understanding of what is read. In fact, poorer readers are generally slower at word recognition (pp. 52-54; see article for citations of research).

There is a substantial body of evidence indicating that individual differences in speed of accessing (in memory) and processing verbal material are related to differences in performance on (a) tests of reading comprehension specifically (e.g., Curtis & Glaser, 1983; Glaser, Lesgold, & Lajoie, 1985), and (b) verbal ability tests generally (e.g., Hunt, Lunneborg, & Lewis, 1975; Hunt, 1978, 1987; McClelland, 1979; Benton & Kiewra, 1987).[4]

4. The model used to generate a "speed of comprehension" score and a "level of comprehension" score from a single, timed administration of the Cooperative Reading Comprehension Test (ETS, 1960a, 1960b) appears, conceptually, to be applicable to any timed--that is, speeded--reading comprehension test. In the Cooperative Reading Comprehension Test (CRCT) model, the first half of the test (completed by almost all examinees) is scored for level and the total test (completed by relatively few examinees) is scored for speed. More detail regarding the CRCT model and its application in the present study is provided later.

5. Certain subgroups of GRE examinees may tend to perform differently on measures of speed and level of reading comprehension, such as those defined by the CRCT model or variants of that model, namely, nonnative-English speaking examinees and subgroups defined by major areas differing in verbal-relative-to-quantitative emphasis.

Nonnative-English speaking examinees. The logic of obtaining a rate-free reading comprehension score and a speeded score is apparent in the case of examinees for whom English is a second language (ESL). For ESL examinees, scores obtained under time constraints may underestimate level of ability to read with comprehension in English, the nondominant language.

Research on the verbal test performance of foreign ESL examinees, although not specifically focused on reading comprehension, tends to support this logical proposition. For example, lower test-completion rates (based on last-item-attempted indices) for foreign ESL examinees than for the general examinee population have been reported for the GRE verbal measure (Angelis, Swinton & Cowell, 1979), and for the Graduate Management Admission Test (GMAT) verbal section (e.g., Sinnot, 1980). Results of experimental studies of decoding time for verbal materials in bilingual tasks also support the propostion (e.g., Dornic, 1980). Dornic reported as follows:

> (P)erformance deteriorated (that is, solution time [for verbal items] increased) as a function of difficulty (i.e., with increasing load on attention, short-term memory, and simple forms of reasoning ability), clearly more so for the subordinate language. Moreover, the difference between the languages was enhanced when time-stress was added to the task-stress. (The subjects were repeatedly urged to perform as fast as possible) (p. 27).

Thus, time constraints had greater impact on verbal performance in the subordinate language than on performance in the dominant language of the bilingual subjects.

Subgroups defined by discipline. It is possible that there may be discipline-related differences in speed of processing verbal material. Generally speaking, majors in highly verbal fields (epitomized by the humanities) may tend to perform relatively better on a more speeded reading comprehension test than on a less speeded test, due to factors associated with their extensive involvement in activities involving primarily verbal processing. On the other hand, majors in fields with heavy demands on processing quantitative material (epitomized by the math-sciences and physical sciences) may tend to perform relatively better on level of comprehension than on speed of comprehension.

Viewed from an information processing perspective, for example, there may be major-area-related differences in the extent to which "automatic information processing" (AIP) has become established (e.g., LaBerge & Samuels, 1974)--the greater the extent of AIP, the more time to be devoted to analytical problems. It is plausible (a) that over 16 or more years of increasingly specialized concentration in subjects whose mastery involves extensive general verbal processing, majors in highly verbal fields may have developed a higher degree of automaticity in processing verbal material than have their counterparts in the math-sciences and physical science fields, and (b) that a higher degree of AIP is conducive to greater speed of reading comprehension.

- Evidence presented in the following section suggests that majors in verbal fields tend to have more extensive vocabularies than their counterparts in quantitative fields, a factor theoretically conducive to speedier resolution of elemental decoding and memory search phases of the reading process. More extensive word knowledge also suggests less need to infer meaning of unfamiliar words from context.

6. It is of interest to explore the possibility of differences in speed relative to level of reading comprehension in other major GRE subpopulations, including those defined by sex and ethnic group membership. Experimental changes in time limits have been found not to affect the relative standing of such subgroups on the GRE General Test (e.g, Wild & Durso [1979]) or similar admission tests (such as the Scholastic Aptitude Test [SAT], for example).[5] In the SAT sample, no significant interactions were found between time conditions and subgroup membership in the analysis by ethnic group even though test completion rates were lower for Black examinees than for White examinees under all time conditions. This pattern is also present in data obtained in current administrations of the SAT. Dorans, Schmitt, and Bleistein (1988), for example, report lower section-completion rates for Black and Hispanic examinees than for White and Asian-American examinees.

Such findings suggest, as a working hypothesis, that groups defined by gender and by ethnicity will tend to have the same relative standing on level and speed of reading comprehension.

7. As indicated earlier, GRE RC and DV subscores are expected to have different patterns of relationships with external academic criteria. However, little research appears to have been conducted to assess the comparative validity of differentially speeded but otherwise parallel reading comprehension (or other cognitive) tests for predicting such criteria. And, there is little direct evidence of differential predictive validity for Speed of Comprehension and Level of Comprehension scores on the Cooperative Reading Comprehension Test itself.

There is evidence (e.g., Lord, 1956; Kendall, 1964) suggesting the possibility of higher validities for more speeded than for less speeded cognitive tests.

- Selected findings of a classic study of speed factors in tests and academic grades (Lord, 1956), probably unique of its type, are of greatest immediate interest. Lord administered several short, differentially speeded but otherwise parallel verbal, numerical, and spatial tests to more than 600 U.S. Naval Academy students. One of three short, highly speeded vocabulary (antonyms) tests was the best single predictor of end-of-course grades regardless of subject.[6]

Kendall (1964) cites research in which more highly speeded tests were found to have higher predictive validities than did less highly speeded tests. Based on the foregoing, it was considered useful to explore relationships between various GRE verbal subscores and self-reported undergraduate grade point average in the present study.

Evidence Regarding the Psychometric Distinctiveness of GRE Reading Comprehension

A considerable amount of research has been undertaken to assess the factor structure of items included in the verbal and quantitative sections of the GRE General Test (in both pre-October 1977 and subsequent editions), and in the verbal section of the Scholastic Aptitude Test, which is made up of item types identical to those employed in the GRE verbal measure. Typically, at least two factors, called "reading comprehension" and "vocabulary," have been identified in these studies.

In several studies (e.g., Powers, Swinton, & Carlson, 1977; Rock, Werts, & Grandy, 1979; Powers & Swinton, 1981), a "reading comprehension" factor was defined by both reading comprehension and sentence completion items, and a separate factor defined by loadings from antonym and analogy items was labeled as a vocabulary factor. Kingston and Dorans (1982), however, identified a reading comprehension factor that was defined primarily by the GRE reading comprehension items only,

and a vocabulary factor defined by the three discrete verbal item types. The same factorial identification was found by Dorans and Lawrence (1987) in an analysis involving items in the SAT verbal measure. Dorans and Lawrence commented as follows on the implications of their findings:

> If the format of SAT-Verbal were to be revised, results of this study point out that it appears reasonable to add more reading comprehension items to obtain a more unique reading score than currently reported (pp. 80-81).[7]

Further evidence of the distinctness of GRE reading comprehension sets as measures of reading ability (as opposed to verbal ability) is provided in a study by Lord and Wild (1985), who examined the efficiency of DV and RC items as measures of reading ability (defined by "number-right true score on the GRE reading items") and verbal ability (defined by total verbal performance). Among other things, they concluded as follows:

> If one wishes to measure 'reading ability,' item-for-item, this is best done with reading items. . .. The fact that the [findings for reading and for other item types] are so different indicates that the two abilities differ substantially (p. 14).

Major-Field Differences in Performance on Verbal Items

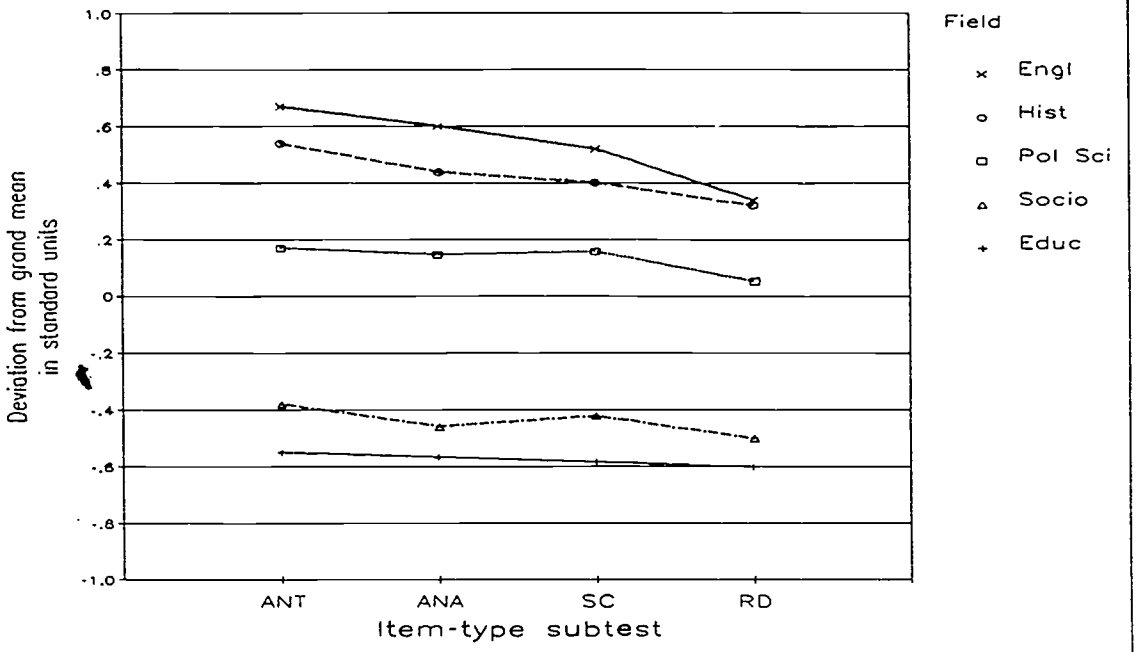
Systematic major-field differences in patterns of average performance on GRE verbal items have been reported (e.g., Wilson, 1985a, 1986b). As shown in Figure 1a, majors in verbal fields tend to perform relatively better on each of the DV (vocabulary) items (antonym, analogy, and sentence-completion questions) than on the reading comprehension items, but the opposite pattern obtains for majors in quantitative fields (Figure 1b). This pattern was found to be consistent for subgroups defined by gender and ethnic group membership.

In these studies it was also found that a GRE reading comprehension subtest (based on sentence completion and reading comprehension items) correlated more highly with an external academic performance criterion (self-reported undergraduate grade point average) than did a GRE "vocabulary" score (based on antonymn and analogy items).[8] Validity coefficients for the subscore based on reading comprehension items only typically were higher than those for subscores based on either the antonym or the analogy questions.

Rationale for Developing Level and Speed of Comprehension Scores from a Single Administration

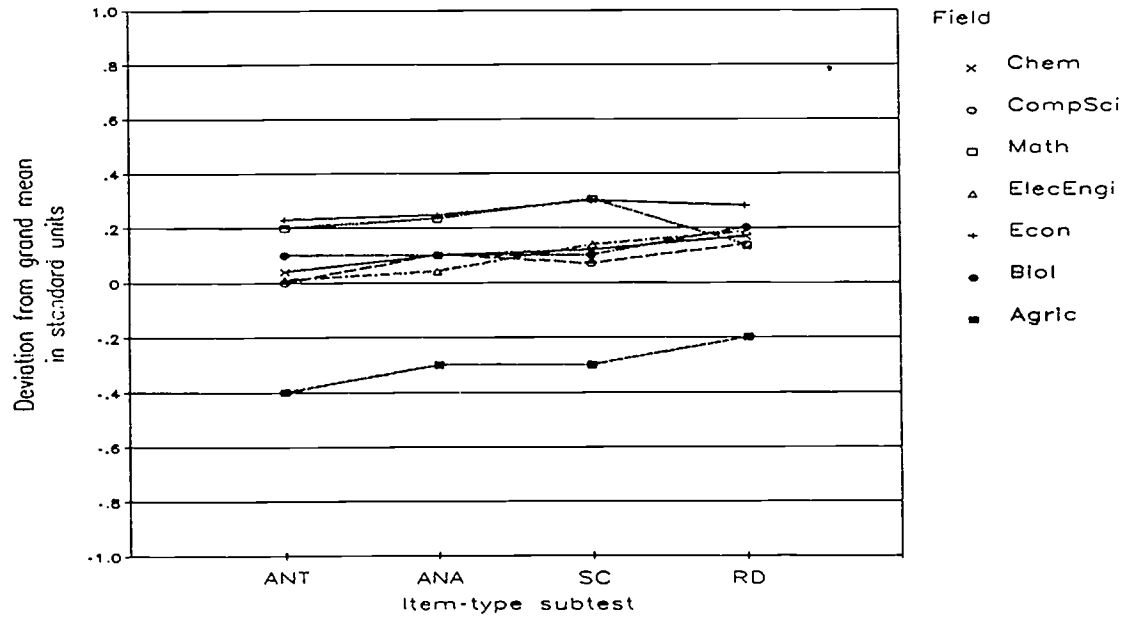
The model for generating the speed and level of comprehension scores employed in this study was suggested by, and conforms closely to, the model employed for developing such scores for the Cooperative

Fig. 1a. Means on GRE verbal-item-type subtests for examinees in five primarily verbal fields



Note. Adapted from Wilson (1985a: 13).

Fig. 1b. Means on GRE verbal-item-type subtests for examinees in five primarily quantitative fields and two biological science fields



Note. Adapted from Wilson (1985a: 13).

Reading Comprehension Test (CRCT)--a basic component in the Cooperative English Tests series (ETS, 1960a, 1960b).[9]

The CRCT Speed/Level Rationale

The CRCT is made up of a 60-item vocabulary (VO) test (with a 15-minute time limit), and a 60-item reading comprehension (RC) test (with a 25-minute time limit). From the 60-item RC section, two scores are obtained, labeled Level of Comprehension (level) and Speed of Comprehension. The Level of Comprehension score is based on the first 30 items; the Speed of Comprehension score is based on all 60 items. Scores on the Vocabulary (VO) Test and the Speed of Comprehension Test (that is, the total score based on all 60 RC items) are averaged to obtain a total Reading score.

The CRCT rationale for obtaining a level and a speed score from a single test administration is explained in the manual as follows:[10]

> The first Reading score is based on the number of items the student answers correctly out of the first 30 items . . .. Since experimental tests have shown that most students have time to try all of these items, this is primarily a power score representing Level of Comprehension.
>
> The second Reading score is based on the number of items the student answers correctly out of all 60 items . . .. This score has been shown to be dependent on how fast students can read the passages with understanding . . . and is aptly labelled Speed of Comprehension.

The technical manual (ETS, 1960b) also reports the results of a number of studies relating one or more of the reading scores to grades in samples of high school, college, and graduate school students. All of the scores (vocabulary, speed, level, and total reading) were not included in each of the studies, the educational levels of the samples differed, and so on. The findings do not provide a clear basis for evaluating the comparative validity of the various scores for predicting academic performance criteria.[11]

Limitations of the CRCT Model

The technical manual does not offer a rationale for the differentiation of speed and level scores. It does not suggest types of circumstances, if any, in which these two scores might be expected to have differential predictive validity--that is, types of criteria for which the speed score might be expected to be more, or less, valid than the level score.[12]

A recognized limitation of the CRCT model for developing separate measures of level and speed of reading comprehension from a single test administration is that the two measures are not experimentally independent, and thus are relatively highly related (due to part-whole correlation). This problem would appear to be present in all single-

administration approaches to assessing speed and level (see Rindler, 1979).[13]

The relatively high (spurious) correlation involved in the half-test/whole-test CRCT model complicates interpretation of observed differences in performance on the measures of speed and level. The problem of experimental dependence involved in treating the whole test score as the "speed of comprehension" score, as suggested above, is a generic one. However, if essentially all examinees complete the first half of a reading comprehension test, CRCT-type level scores clearly resemble "power scores" (essentially free of items-not-reached variance).

Adaptation of the CRCT model in the present study. In the present study, an effort is made to avoid this problem by treating scores on the second half of a GRE reading comprehension test as the principal "speed of comprehension" measure. This approach provides two reading scores, one thought of as reflecting performance under "power" conditions and the other as reflecting performance under "speed" conditions. It cannot be assumed that the two halves are otherwise parallel--for example, the reading passages necessarily will involve different content, the difficulty levels of the sets of questions will tend to vary across test halves, and so on.

Characteristics of the GRE Reading Test Employed in the Study

The study employed data for Form YGR2 of the GRE General (Aptitude) Test, administered in October 1976.[14] As indicated at the outset, the verbal section in pre-October 1977 forms of the GRE General Test included 55 discrete-verbal (DV) items (with a 25-minute time limit), and a reading comprehension (RC) section including a total of 40 questions based on 6 reading passages (with a 50-minute time limit). The test was administered under formula-scoring instructions.

The "GRE Reading Comprehension Test" (Form YGR2)

The reading comprehension section was composed of six reading passages and accompanying sets of questions. The topics addressed by the several reading passages and the number of questions (and question numbers) associated with each passage were as follows:

1. On the role of women in the union movement (7 questions: numbered 1-7)
2. On the distinction between drama and narrative literature (6: 8-13)
3. On the nature of ecological systems (7: 14-20)
4. On the contribution of one Black woman to the development of educational opportunity in the South (7: 21-27)
5. On the nature of aggression (7: 28-34)

6. On evidence of changes in climatic zoning of the earth (6: 35-40).

The several sets of questions were of approximately the same average level of difficulty, judging from median difficulty indices (equated delta [ED], based on percentage of correct responses): median EDs were 11.6, 10.5, 11.9, 11.4, 12.2, and 11.5.[15] However, the range of difficulty within the respective sets of questions varied somewhat. For example, all the questions associated with passages 5 and 6 had EDs of at least 10.0; the other sets included one or more items with lower EDs.

Table 1 shows distributions of EDs for RC1 (items 1-20) and RC2 (items 21-40) and additional information regarding the properties of these two sets of items.

Degree of speededness. According to internal ETS criteria, tests are judged to have an acceptable degree of speededness if 100% of test takers reach 75% of the test items (item 30 for RC and item 41 for DV) and 80% reach the last item. Plots of these percentage indices are shown for the reading comprehension test (in Figure 2a) and for the discrete verbal section (in Figure 2b), for all examinees and for examinees in the upper and lower quintiles on total verbal score; vertical bars indicate the respective criterion percentages.

- o For the total sample, the percentage completing the RC section (about 84%) marginally exceeded the 80% target, while the percentage completing three-fourths of the items (98%) slightly failed to meet the 100% target. The DV section was substantially more speeded: only 66% of the analysis sample completed the section, and 96% reached item 41.

- o Most of the upper-quintile examinees completed the RC section but less than two-thirds of lower-quintile examinees did so. The DV section was even more speeded for the lower-quintile examinees; less than one half of the subgroup completed the DV section.

- o However, essentially all the examinees in the analysis sample completed the first half of the RC section (RC1): 1,957 of the 1,960 examinees (99.8%) responded to item 20.

Based on this sample analysis, it is reasonable to expect almost all GRE examinees to be able to complete the first 20 RC items (RC1), and that RC1 thus may appropriately be thought of as reflecting differences in level of ability to read with comprehension. On the other hand, scores based solely on the second 20 RC items (RC2), or all the RC items, are directly affected to some extent by differences in speed of responding to the test items--RC2 variance clearly includes a speed-of-reading component associated with the number of not-reached items.

On the basis of the test analysis results, it seems reasonable to think of RC1 and RC2 as reading comprehension tests with the following

Table 1

Distribution of Difficulty Indices (Equated Delta) for Items 1-20 (RC1) and Items 21-40 (RC2) of the GRE Reading Comprehension Test, and Other Basic Characteristics of the RC1 and RC2 Subsections

Items 1-20 (RC1) EqD		Items 21-40 (RC2) EqD	
16.		16.	3
15.		15.	
14.		14.	36
13.	378	13.	238
12.	0238	12.	0024
11.	269	11.	11499
10.	2899	10.	117
9.	00678	9.	
8.		8.	56
7.		7.	
6.	1	6.	
Mdn	11.0		11.5
Number of passages	3		3
Total lines	159		161

Classification of passages:

Social Studies	(11.6)*	Narrative	(11.4)
Humanities	(10.5)	Argumentative	(12.2)
Biosciences	(11.9)	Physical Sciences	(11.5)

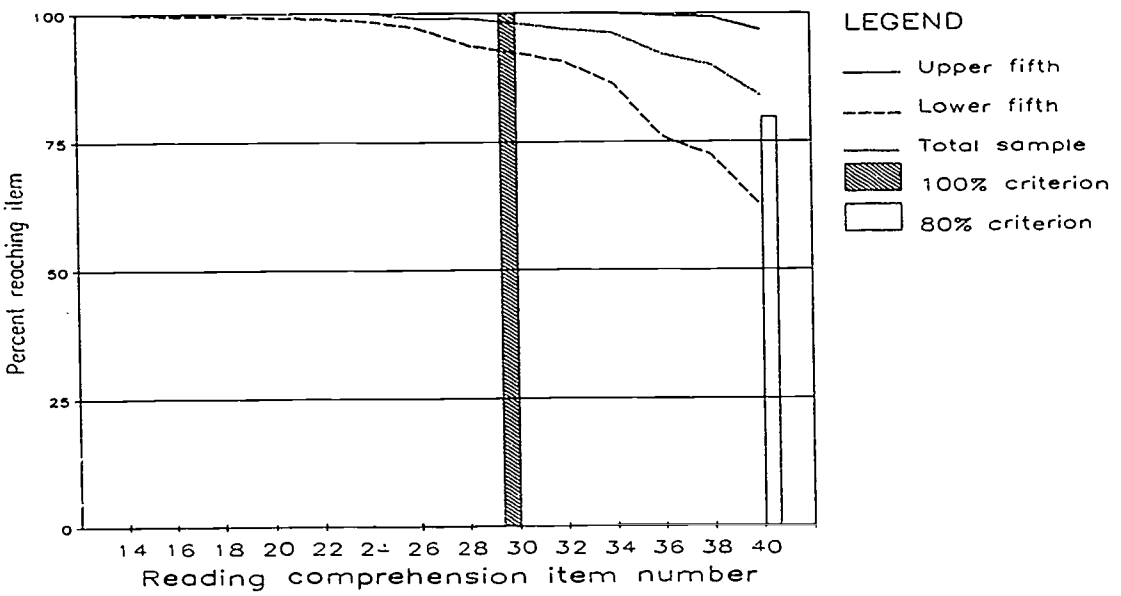
Note. The individual delta items may be read by combining whole and successive decimal digits in each row. For example, deltas for items in RC2 are 16.3, 14.3, 14.6, 13.2, and so on.

* Median ED for items associated with this passage.

Source of data: Routine test analysis data in ETS files for Form YGR2, administered in October, 1976.

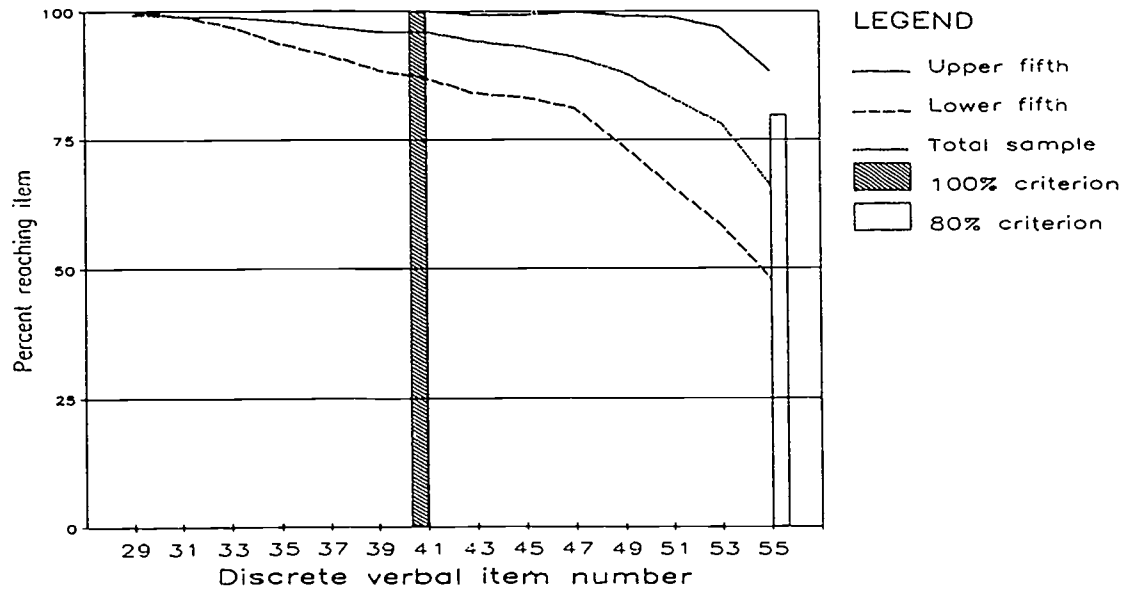
Delta.tbl(x)

Fig. 2a. Percentage of examinees reaching designated GRE Reading Comprehension (RC) items, by level of performance on the total GRE Verbal Test (upper and lower quintiles)



Note: Adapted from standard test analysis for this administration.

Fig. 2b. Percentage of examinees reaching designated GRE Discrete Verbal (DV) items, by level of performance on the total GRE Verbal Test (upper and lower quintiles)



Note: Adapted from standard test analysis for this administration.

characteristics:

(a) of the same length, as indexed by the number of items,

(b) balanced as to total amount of verbal processing required, as indexed by total number of lines,

(c) about equally difficult, as indexed by average ED, but

(d) differentially speeded--RC2 has a speed component, indexed by items not completed, that is minimally present in RC1.

Content and stylistic differences. The sets of reading passages in RC1 and RC2 differ with respect to types of subject matter and stylistic emphasis. A goal of test development is to provide in each reading passage all the context or information needed to answer the associated questions--specific prior knowledge is not required in order to understand the passages. Stylistic differences in the passages may also have differential effects on performance for some individuals or subgroups. The possibility of effects due to factors other than speed of processing the material needs to be kept in mind in evaluating subgroup differences in performance on the first versus the second 20 RC items.[16]

Sample, Data, and Study Procedures

The sample was composed of 22,175 examinees who took Form YGR2 of the GRE General (Aptitude) Test in October 1976, for whom item-level test data and responses to GRE registration form or background questions were available in GRE files: questions as to (a) sex, (b) citizenship status [U.S. vs. non-U.S.], (c) better language of communication [English vs. other], (d) ethnic group membership, (e) graduate major field, and (h) self-reported undergraduate grade point average in the major field. Table 2 shows the numerical distribution of the sample, by graduate major area, for subgroups defined by citizenship status (U.S. vs. non-U.S.), sex, reported English-language communication status (EPL vs. ESL)[17], and (for U.S. citizens only) reported ethnic group membership (Asian American or Asian, Black or Bl, Hispanic or Hsp [Mexican American, Puerto Rican, and Other Hispanic combined], and White). Figure 3 shows the percentage distributions by major area for these same subgroups.

o In both citizenship categories, proportionately fewer females than males were in physical science fields (including mathematical sciences); ESL examinees tended to be concentrated more heavily in quantitative fields than in the social sciences or humanities; this trend was not apparent, however, for U.S ESL-examinees.

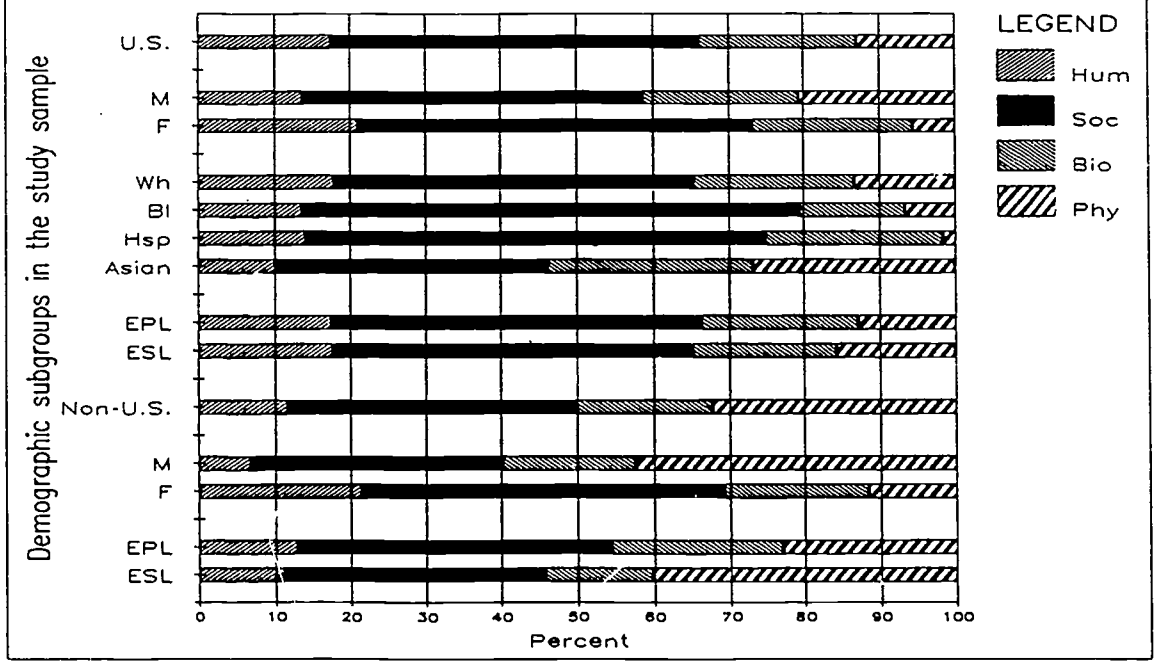
o Among U.S. examinees, proportionately fewer Black examinees and Hispanic examinees were in quantitatively oriented disciplines; both groups were more highly concentrated in the social sciences.

Table 2

Distribution of the Sample by Graduate Major Area, within Subgroups by Citizenship Status (U.S. vs Non-U.S.)

Citizenship/Group	Humanities	Social Science	Bioscience	Physical Science	Total
	Number of examinees				
U.S.	3,692	20,258	4,429	2,718	21,097
Male	1,398	4,599	2,103	2,121	10,221
Female	2,294	5,659	2,326	597	10,876
EPL	3,616	10,055	4,346	2,651	20,668
ESL	76	203	83	67	429
White	3,383	8,963	4,064	2,480	18,890
Black	147	711	151	71	1080
Hispanic	61	263	64	46	434
Asian	27	98	73	72	270
Non-U.S.	125	410	195	348	1,078
Male	49	242	127	307	725
Female	76	168	68	41	353
EPL	63	205	112	114	494
ESL	62	205	83	234	584

Figure 3. Major-area distribution by citizenship, sex, ethnicity, and self-reported English-language communication status



Computation of Formula Scores for GRE Variables

Using item-level data for the total sample (N = 22,175), formula scores (R-W/4) were computed for the GRE variables designated below. Formula scoring was used because the test was administered under formula-scoring instructions (which do not encourage purely random marking of items not reached):[18]

(Level of Comprehension)

RC1 = formula score on the first half of the GRE RC items (items 1-20), thought of as a measure of level of reading comprehension, free of speed-related variance associated with not-reached items.[19]

(Scores with a Speed Component)

RC2 = formula score on the second half of the reading items (items 21-40), a composite of speed and level of reading ability, thought of as a "speed of comprehension" indicator.

RCodd = formula score on 20 odd-numbered reading items (1, 3, . . ., 39), thought of as a 20-item surrogate for the total 40-item RC score (a composite of level and speed).[20]

RCeven = formula score on 20 even-numbered reading items (2, 4, . . ., 40), thought of as generally comparable to RCodd, computed primarily for comparative purposes.

DVodd = formula score on 28 odd-numbered discrete verbal items (analogy, antonym, and sentence completion item types), thought of as a 28-item surrogate for the total score on the 45 DV items (a vocabulary score with both a speed and a power component).

Vform or Vf = total formula score on 95 GRE verbal items (with a speed as well as a power component).

Qform or Qf = total formula score on 55 GRE quantitative items (speededness not assessed).

For exploratory purposes, a residual variable, RC2res, reflecting performance on RC2 relative to expectation based on RC1 score was also computed: RC2res = RC2 - RC2', where RC2' is RC2 predicted from RC1. By virtue of the derivation process, RC2res is expected to be uncorrelated with RC1 but relatively highly correlated with RC2.

Means, Standard Deviations, and Intercorrelations of the Variables

Means and standard deviations for the formula-scored GRE variables are shown in Table 3. Table 4 shows intercorrelations of these variables for 21,079 U.S. citizens (above the diagonal), and 1,078 non-U.S. citizens (below the diagonal).

Table 3

Performance of the Study Sample (N = 22,175) on Derived GRE Variables

Variable	No. items	Mean	S.D.
RC1	20	12.9	4.1
RC2	20	11.0	4.8
RCodd	20	11.6	4.2
RCeven	20	12.3	4.5
DVodd	28	12.0	5.8
Vform	95	47.4	17.7
Qform	55	29.7	11.1

Table 4

Intercorrelations of the Variables in the Study Sample: By Citizenship Status

Variable	RC1	RC2	RC-odd	RC-even	DV-odd	V-form	Q-form	RC2-res
RC1	--	.66	(.83)	(.82)	.62	(.81)	.56	[.00]
RC2	.70	--	(.84)	(.87)	.67	(.85)	.54	[.75]
RCodd	(.88)	(.88)	--	.72	.66	(.84)	.55	[.39]
RCeven	(.87)	(.89)	(.82)	--	.67	(.86)	.58	[.44]
DVodd	.68	.73	.74	.72	--	(.91)	.47	[.35]
Vform	(.65)	(.88)	.90	.92	.91	--	.59	[.42]
Qform	.41	.34	.36	.40	.21	.33	--	[.24]
RC2res	[.01]	[.70]	[.36]	[.39]	[.35]	[.40]	[.07]	--

Note. Coefficients above the diagonal are for U.S. citizens (N = 21,079); coefficients below the diagonal are for Non-U.S. citizens (N = 1,078). Parentheses indicate part-whole coefficients; coefficients in brackets are for the residual variable.

As expected from the item analysis results, the sample mean for RC1 (which included one very easy item) is somewhat higher than that for RC2. The mean for RCeven is slightly higher than that for RCodd.[21] Other noteworthy trends include the following:

1. The correlations between scores on level and speed subtests, RC1 and RC2, respectively, are lower than those for scores on the comparably speeded RCodd and RCeven subtests in both samples (.66 vs. .72 for U.S. examinees, and .70 vs. .82 for foreign examinees).

2. RC2 is related more closely to DVodd (with a speed component) than is RC1 in both samples (.67 vs. .62 for U.S. examinees; .73 vs. .68 for non-U.S. examinees). Similarly, RC2 contributes more to the total verbal score (somewhat speeded) than does RC1 (part-whole coefficients of .85 vs. .81 for U.S. examinees; .88 vs. .85 for foreign examinees).

3. In contrast, scores on the RCodd and RCeven subtests--tests that by inference are roughly comparable with respect to speededness-make a comparable contribution to total verbal performance (part-whole coefficients are .84 and .86 [U.S.]; .90 and .89 [non-U.S]).[22]

4. Coefficients for RC1 and RCodd (analogous to the CRCT-defined "level of comprehension" and "speed of comprehension" scores) are quite high (.834 for U.S. citizens and .876 for non-U.S. citizens), reflecting spurious effects due to lack of experimental independence (that is, the items scored for level are included in the items scored for speed).

5. Generally speaking, coefficents involving various pairs of verbal subtests are higher for foreign examinees than for U.S. examinees, suggesting that the verbal skills of foreign examinees are less sharply differentiated than those of the predominantly native-English-speaking U.S. examinees.

Preliminary Operations on the Variables

To facilitate exploratory evaluation of differences in relative standing on targeted GRE variables for designated examinee subgroups, the formula score distributions were standardized, through a "z-scale" transformation--that is, formula scores were expressed as deviations from the total-sample means, in standard-deviation units.

- o Following the z-scale transformation, each test variable had a mean of 0 and a standard deviation of 1.0 in the total study sample (N = 22,175). Thus, in the total sample, mean z(a) = mean z(b), . . ., = mean z(z) = 0. However, for any given subgroup, inequalities may be observed in average standing on any pair of variables (e.g., mean z(a) < mean z(b), or vice versa).

Intercorrelations of the variables, of course, were not affected by the z-scale transformation.

Analytical Rationale

Based on evidence and lines of reasoning presented in earlier sections, speed versus level differences (marked by differences in correlated z-scaled means on RC1 and RC2) were expected in comparisons involving (a) subgroups differing in English-language background, and (b) subgroups defined by graduate major area,[23] but not in comparisons involving subgroups defined by (c) sex or (d) ethnicity. The pattern of expected outcomes is outlined below:

Expected outcome (Z-scaled means)	Group
Level = Speed (RC1 = RC2)	Male, Female; Asian American Black, Hispanic American, White
Speed > Level (RC2 > RC1)	EPL; Humanities, Social Sciences; all U.S.(because predominantly EPL)
Speed < Level (RC2 < RC1)	ESL; Physical Sciences, Biosciences; all non-U.S.(due to ESL-effects).

Related analyses were undertaken to evaluate expected major-area differences in GRE reading comprehension relative to vocabulary--DVodd, somewhat speeded, representing word knowledge and other aspects of general verbal ability as distinguished from reading ability.

Expected outcome (Z-scaled means)	Examinees classified by graduate major area
RCodd > DVodd	Physical Sciences, Biosciences
RCodd < DVodd	Humanities, Social Sciences

Questions regarding possible differences in criterion-related validity for Speed (RC2) and Level (RC1), with respect to academic criteria, were explored for U.S. examinees only, using self-reported undergraduate GPA (SR-UGPA) in the major field as the external criterion. Simple correlations between the GRE subtests and the SR-UGPA criterion were computed by major area and by sex in order to assess consistency or lack of consistency in direction of the differences between subtest coefficients--questions regarding possible differential validity for subgroups were not at issue.

Expected correlational outcome with respect to size of coefficients:

RC1 = RC2 > DVodd (Level = Speed > DVodd).

Procedure

To assess the extent to which observed outcomes were consistent with expectation, descriptive statistics (means, standard deviations,

and intercorrelations) were computed for the z-scaled GRE variables defined for the study, for each of the basic demographic and academic subgroups specified--that is, separately, by citizenship status, for examinees classified by sex, ethnic group (U.S. only), EPL/ESL status, and graduate major area.

For each subgroup (males, females, EPL, ESL, and so on), the speed/level difference of primary interest was:

Speed vs. level = (mean RC2 minus mean RC1).

Findings described in detail later suggested that the most distinctive and pervasive pattern of RC2/RC1 inequality, plausibly interpretable as reflecting speed versus level effects, was that associated with graduate major area.

Based on the foregoing, further exploration of speed/level differences was undertaken to assess the extent to which the pattern of RC1/RC2 inequality associated with graduate major area was consistent for various subgroups.

- o The difference value, mean RC2 minus mean RC1, was computed by graduate major area for subgroups of U.S. and non-U.S. examinees classified by (a) sex, (b) ethnicity (U.S. examinees only), and (c) EPL/ESL status.

- o The difference value, mean DVodd minus mean RCodd, was computed for these same subgroups.

To assess consistency of major-area-related speed/level differences with control for "general verbal ability," examinees were classified by level of total GRE verbal score--upper 27%, middle 46%, and lower 27%, respectively.

- o Mean RC2 minus mean RC1 was evaluated by graduate major area for subgroups within the general verbal ability categories.

- o In a related analysis, mean RC2res was similarly evaluated.

Finally, an exploratory evaluation was made of the relationships of RC1, RC2, and DVodd to the z-scaled SR-UGPA criterion, in subgroups of U.S. examinees by major area and sex, to assess consistency of findings regarding the criterion-related validity of these variables.

Findings

Subgroup Performance: Patterns of Speed/Level Differences

Table 5 shows means and standard deviations of z-scaled scores on RC1, RC2, and DVodd for designated subgroups of examinees classified by citizenship status. The means are plotted in Figure 4a (U.S. exam-

Table 5

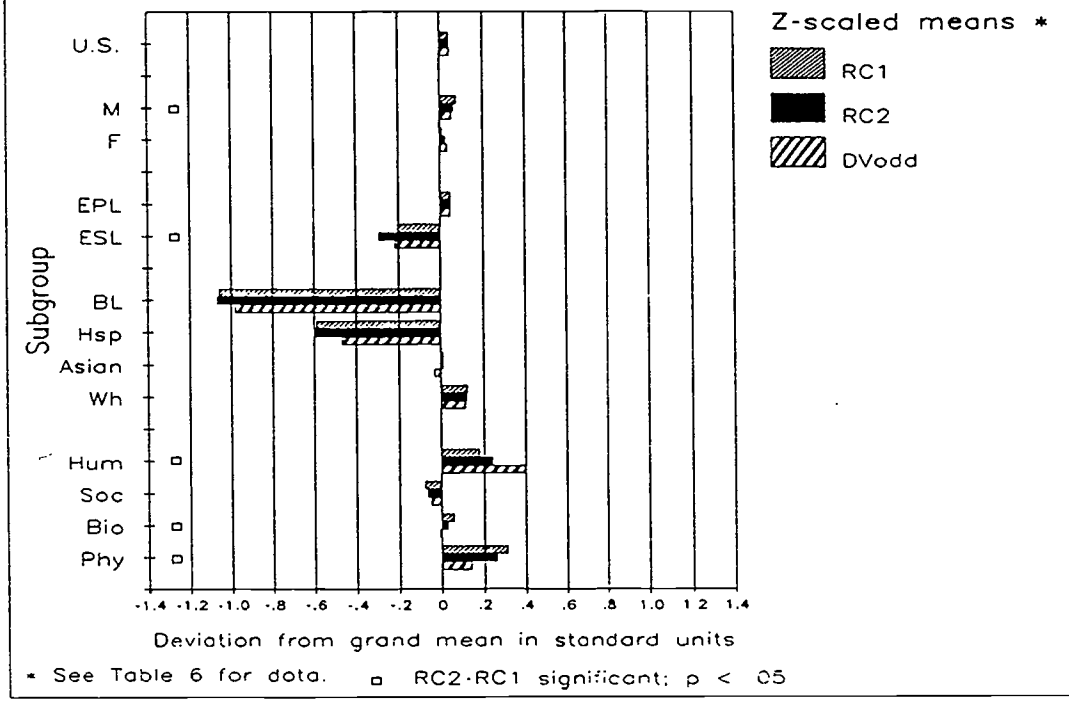
Performance of Demographic and Academic Subgroups on RC1, RC2, and DVodd, by Citizenship Status*

Group	N	Z-scaled mean			Z-scaled standard deviation		
		Level	Speed	DVodd	Level	Speed	DVodd
U.S.	21,097	.04	.04	.05	.97	.97	.98
EPL	20,668	.05	.05	.05	.97	.97	.98
ESL	429	-.21	-.29	-.22	1.08	1.08	1.05
Hum	3,692	.18	.24	.40	.93	.93	1.03
Soc	10,258	-.08	-.07	-.05	1.00	1.00	.98
Bio	4,429	.06	.03	-.01	.91	.93	.86
Phys	2,718	.31	.26	.14	.90	.93	.93
Male	10,221	.08	.06	.06	.97	.97	.97
Female	10,876	.01	.02	.04	.97	.98	1.00
White	18,890	.12	.12	.12	.92	.93	.94
Asian	270	.01	.01	-.03	.97	1.05	1.04
Hispanic	434	-.59	-.60	-.47	1.09	1.02	.95
Black	1,080	-1.06	-1.07	-.98	1.04	.91	.86
Non-U.S.	1,078	-.87	- .88	-.91	1.19	1.11	1.02
EPL	494	-.52	- .49	-.50	1.20	1.12	1.08
ESL	584	-1.16	-1.21	-1.26	1.10	.99	.82
Hum	125	-.82	-.54	-.54	1.20	1.16	1.16
Soc	410	-.87	-.85	-.84	1.22	1.17	1.09
Bio	195	-.84	-.90	-.93	1.17	1.07	.88
Phy	348	-.89	-1.02	-1.12	1.17	1.02	.91
Male	725	-.95	-1.02	-1.05	1.15	1.13	1.03
Female	353	-.69	-.57	-.63	1.18	1.09	1.01
No. of items	20			20			28

Note: Underscoring indicates that the difference between the pair of correlated RC1 and RC2 means is significant ($p < .05$). In the non-U.S. sample, as in the U.S. sample, ESL examinees had relatively lower standing on RC2 (speed) than on RC1 (level), and the opposite was true for EPL examinees, but the difference did not reach the $p < .05$ level in either instance.

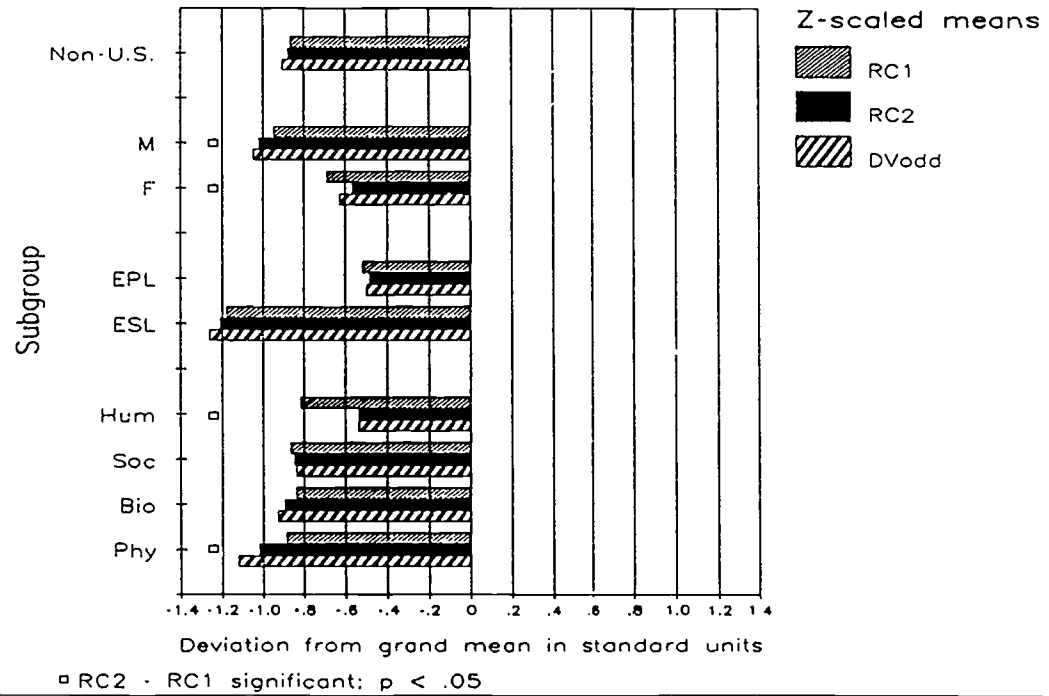
*RC1 represents level of reading comprehension; RC2 represents "speed" of comprehension; DVodd represents vocabulary.

Fig. 4a. Z-Scaled means of U.S. demographic and academic subgroups on RC1 (20 items), RC2 (20 items), and DVodd (28 items)



* See Table 6 for data. ▫ RC2-RC1 significant; $p < .05$

Fig. 4b. Z-Scaled means of non-U.S. academic and demo-grphic subgroups on RC1 (20 items), RC2 (20 items), and DVodd (28 items)



▫ RC2 - RC1 significant; $p < .05$

inees) and Figure 4b (non-U.S. examinees). In evaluating these means it is useful to recall that the variables were z-scaled in the combined sample (U.S. and non-U.S. examinees). The patterns of subgroup means on Speed and Level (RC2 and RC1), respectively, were generally consistent with the "speed relative to level" patterns hypothesized. And, in both citizenship groups, the pattern of major-area differences in relative standing on the GRE reading comprehension and discrete verbal ("vocabulary" subtests) was as expected.

EPL/ESL status. In analyses by EPL/ESL status, the expected outcome was RC2 < RC1 for ESL examinees, and RC2 > RC1 for EPL examinees); a similar pattern was considered plausible in analyses involving U.S./non-U.S. examinees--due to the disproportionate number of ESL examinees in the non-U.S. population).

o The "RC2 minus RC1" discrepancies for groups classified by EPL/ESL status in both the U.S. and the non-U.S. samples are generally consistent with expectation. However, it is evident that the EPL/ESL-parallel pattern, considered plausible for U.S. versus non-U.S. examinees, is not present: for U.S. and non-U.S. examinees generally the observed outcome was RC1 = RC2.

Graduate major area. For humanities and social science majors the expected speed/level outcome was RC2 > RC1; the opposite was expected for bioscience and physical science majors. For reading relative to discrete verbal performance, the expected pattern was DV > Reading (higher performance on discrete verbal than reading items) for the former pair of groups, and DVodd < Reading for the latter pair.

o In both the U.S. and the non-U.S. samples, the pattern of mean differences (for RC2 minus RC1, and for DVodd relative to Reading) for the four graduate major-area subgroups was consistent with expectation.

o It is noteworthy that in the primary U.S. sample, (a) humanities majors and majors in physical sciences had relatively high means on all three verbal measures, (b) physical science majors had higher RC1 and RC2 means than did majors in the humanities, (c) humanities majors had substantially higher means on DVodd than did physical science majors, and (d) these two subgroups differed, as expected, in performance on RC2 relative to performance on RC1.

o Note that in the non-U.S. sample, means on all verbal subscores (especially RC2 and DVodd) tended to be lower for majors in the physical sciences than for majors in the other areas. The verbal subtest means tended to decrease along the humanities-social sciences to biosciences-physical sciences (verbal-relative-to-quantitative emphasis) continuum. This may simply reflect differential development of general English proficiency across major-area subgroups of foreign examinees--that is, the more verbal the major, the greater the need to develop English proficiency in order to pursue that field of concentration in an English-speaking environ-

ment. However, despite this difference, it is apparent that the speed/level (RC2 relative to RC1) pattern by major area is as consistent for non-U.S. examinees as it is for U.S. examinees.

Sex and ethnic group. In analyses by sex and ethnic group membership, the expected outcome was RC1 = RC2.

o The observed outcomes for U.S. ethnic groups were very consistent with expectation--the mean difference (RC2 - RC1) varied between .00 and -.01.

o The "RC2 minus RC1" discrepancies for groups classified by EPL/ESL status in both the U.S. and the non-U.S. samples are generally consistent with expectation. However, it is evident that the EPL/ESL-parallel pattern, considered plausible for U.S. versus non-U.S. examinees, is not present: for U.S. and non-U.S. examinees generally, the observed outcome was RC1 = RC2.

o In both the U.S. and the non-U.S. samples, outcomes by sex were not RC1 = RC2, as anticipated: in both samples, without regard to statistical significance, the outcomes were RC2 < RC1 for males, and RC2 > RC1 for females. In the large U.S sample, gender-related differences were very slight (-.02 for males [statistically significant, $p < .05$], and .01 for females).[24]

o Differences were larger, and statistically significant, for both males and females in the non-U.S. sample.

o The unanticipated gender-related differences plausibly reflect "major-area" effects: for example, males (RC2 < RC1) are enrolled disproportionately in quantitative fields (also RC2 < RC1), while the opposite holds for females. In any event, there is no a priori basis for expecting sex-related speed/level differences.

Consistency of Major-Area-Related Patterns

To evaluate the consistency of the major-area findings for subgroups, "mean RC2 minus mean RC1" and "mean DVodd minus mean RCodd" (RCodd representing total score on reading comprehension) were computed by major area for each subgroup, using data provided in Table 6. In addition to providing means for RC2, RC1, and DVodd for subgroups by graduate major area, Table 6 provides z-scaled means for total verbal score and (for perspective) total quantitative.

Pertinent mean differences are plotted in Figure 5 (for RC2 minus RC1), and Figure 6 (for DVodd minus RCodd)--note that both figures are plotted to the same scale. Certain trends are noteworthy.

o The major-area-related "RC2 minus RC1" and "DVodd minus RCodd" patterns tend to be (a) generally parallel and (b) consistent across subgroups.

Table 6

Means of Subgroups on GRE Reading Comprehension, Vocabulary, Total Verbal, and Total Quantitative by Graduate Major Area and Citizenship

Subgroup	N	Reading			Vocabulary	Vform (total)	Qform (total)
		RC1 "Level"	RC2* "Speed"	RCodd "Total"	DVodd**		
U.S. Hum	3,692	.18	.24 +	.24	.40 +	.33	-.19
Male	1,398	.20	.26 +	.23	.45 +	.39	.05
Female	2,294	.16	.23 +	.24	.37 +	.30	-.33
EPL	3,616	.19	.25 +	.25	.41 +	.35	-.18
ESL	76	-.38	-.27 +	-.35	-.23 +	-.15	-.64
White	3,383	.23	.30 +	.29	.45 +	.40	-.13
Black	147	-.90	-.89 +	-.82	-.70 +	-.90	-1.28
Hispanic	61	-.59	-.48 +	-.60	-.44 +	-.15	-.87
Asian	27	.14	.24 +	.24	.37 +	.30	.25
Soc Sci	10,258	-.08	-.07 +	-.07	-.05 +	-.09	-.26
Male	4,599	-.04	-.05 -	-.06	-.01 +	-.02	-.04
Female	5,659	-.11	-.09 +	-.08	-.08 =	-.13	-.44
EPL	10,055	-.08	-.07 +	-.06	-.05 +	-.08	-.26
ESL	203	-.19	-.32 -	-.24	-.19 +	-.26	-.54
White	8,963	.02	.02 =	.04	.04 +	.02	-.16
Black	711	-1.14	-1.12 +	-1.13	-1.06 +	-1.23	-1.38
Hispanic	263	-.71	-.71 =	-.71	-.57 +	-.73	-.96
Asian	98	.04	.05 +	.05	.06 +	.06	.04
Biosci	4,429	.06	.03 -	.01	-.08 -	.01	.17
Male	2,103	.05	.01 -	-.03	-.12 -	-.01	.10
Female	2,326	.06	.04 -	.05	-.05 -	.00	-.05
EPL	4,346	.06	.03 -	.02	-.08 -	.00	.17
ESL	83	-.15	-.17 -	-.17	-.31 -	-.27	-.25
White	4,064	.11	.08 -	.06	-.04 -	.04	.22
Black	151	-.97	-1.06 -	-.98	-.98 =	-1.11	-1.15
Hispanic	64	-.33	-.40 -	-.36	-.40 -	-.39	-.39
Asian	73	-.02	-.04 -	.00	-.10 -	-.01	.41

Page 1 of 3 pages

Table 6, continued

Subgroup	N	Reading			Vocabulary	Vform (total)	Qform (total)
		RC1 "Level"	RC2* "Speed"	RCodd "Total"	DVodd**		
Phys Sci	2,718	.31	.25 -	.28	.14 -	.27	1.00
Male	2,121	.29	.23 -	.24	.12 -	.26	1.05
Female	597	.40	.33 -	.41	.23 -	.32	.82
EPL	2,651	.32	.27 -	.30	.15 -	.29	1.01
ESL	67	-.12	-.41 -	-30	-34 -	-.30	.62
White	2,480	.36	.31 -	.33	.19 -	.33	1.05
Black	71	-.79	-1.04 -	-.90	-.85 +	-.97	-.33
Hispanic	46	-.32	-.44 -	-.42	-.41 +	-.43	.31
Asian	72	-.05	-.15 -	-.09	-.25 -	-.14	1.22
Non-U.S.							
Hum	125	-.82	-.54 +	-.73	-.54 +	-.69	-.62
Male	49	-.88	-.84 +	-.98	-.72 +	-.89	-.52
Female	76	-.79	-.34 +	-.57	-.48 +	-.55	-.70
EPL	63	-.33	-.01 +	-.19	-.03 +	-.09	.58
ESL	62	-1.32	-1.07 +	-1.28	-1.06 +	-1.29	-.67
Soc Sci	410	-.87	-.85 +	-.90	-.84 +	-.95	-.48
Male	242	-1.02	-.99 +	-1.04	-1.00 +	-1.10	-.34
Female	168	-.65	-.64 +	-.69	-.61 +	-.72	-.68
EPL	205	-.56	-.48 +	-.49	-.45 +	-.54	-.60
ESL	205	-1.18	-1.22 -	-1.30	-1.23 +	-1.35	-.36
Biosci	195	-.84	-.90 -	-.94	-.93 +	-.99	-.14
Male	127	-.89	-1.06 -	-1.08	-1.01 +	-1.09	-.10
Female	68	-.74	-.60 +	-.67	-.79 -	-.79	-.22
EPL	112	-.62	-.63 -	-.65	-.66 -	-.71	-.15
ESL	83	-1.14	-1.26 -	-1.32	-1.31 +	-1.36	-.13

Subgroup	N	Reading RC1 "Level"	Reading RC2* "Speed"	Reading RCodd "Total"	Vocabulary DVodd**	Vforml (total)	Qform (total)
Phys.Sci.	348	-.89	-1.02 -	-1.05	-1.12 -	-1.15	.63
Male	307	-.93	-1.07 -	-1.10	-1.16 -	-1.19	.64
Female	41	-.60	-.68 -	-.69	-.79 -	-.81	.53
EPL	114	-.47	-.62 -	-.55	-.67 -	-.66	.60
ESL	234	-1.10	-1.21 -	-1.29	-1.33 -	-1.39	.64

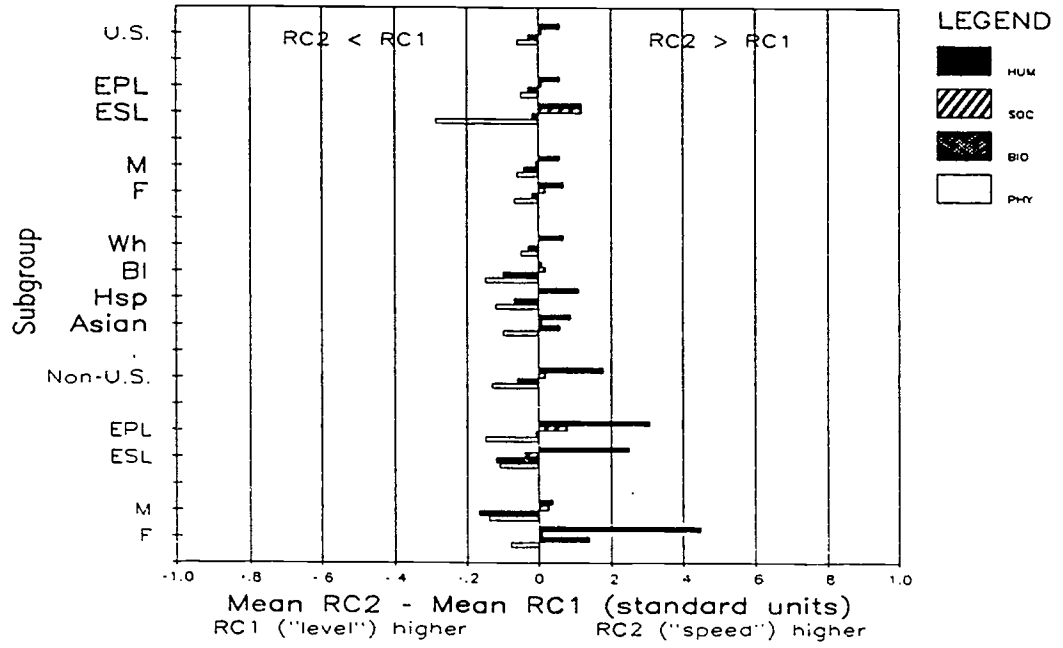
Note: These are means of z-scaled scores (that is, formula scores expressed as deviations from the grand mean for U.S. and Non-U.S. examinees) for:

RC1 = score on first 20 RC items (level);
RC2 = score on second 20 RC items (speed);
RCodd = score on 20 odd-numbered RC items (total reading);
DVodd = score on 28 odd-numbered discrete-verbal items (vocabulary);
Vform = verbal total score on all RC and DV items;
Qform = quantitative total score.

* Signs following entries in the RC2 column are intended to indicate the direction of observed differences: "+" = RC2 > RC1 ("speed" > "level"); "-" = the opposite, and "=" indicates no difference.

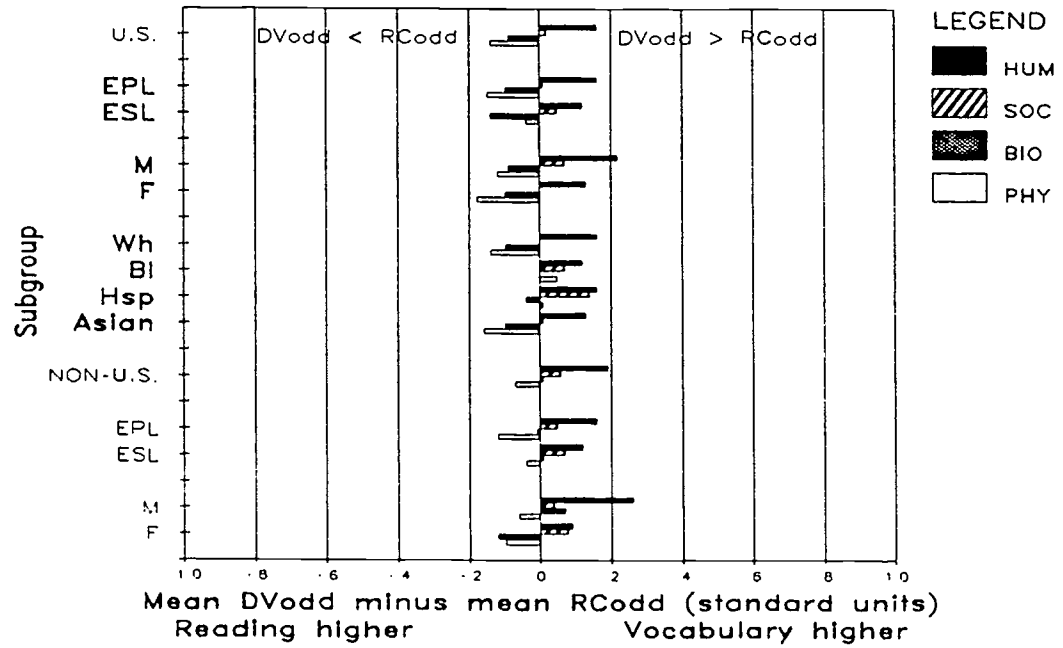
** Signs following entries in the DVodd column are intended to indicate the direction of observed differences: "+" indicates DVodd mean (vocabulary) higher than the reading mean (RCodd) for the subgroup; a "-" indicates the opposite; "=" indicates no difference.

Fig. 5. Potterns of major-area speed/level differences (mean RC2 minus mean RC1) for designated subgroups of examinees: By citizenship status



Note For means of subgroups, see Table 6.

Fig. 6. Major-area differences in standing on DVodd ("vocabulory") and RCodd ("total reading") for subgroups, by citizenship status: (Mean Z[DVodd]) - (Mean Z[RCodd])



Note For means of subgroups, see Table 6.

o At the same time, speed/level discrepancies by major area, while paralleling the pattern of discrete-verbal/reading differences, appear to be somewhat sharper.

o Major-area effects were especially pronounced in analyses involving non-U.S. examinees[25] and U.S. examinees classified by English language background. This suggests the possibility of heightened speed/level differentiation due to interaction of (a) effects associated with generic discipline-related differences in emphasis on verbal processing, and (b) effects associated with differential development of proficiency in English as a second language.

Consistency of major-area speed/level effects for ability-level subgroups. "Mean RC2 minus mean RC1" was computed for major-area subgroups classified by level of total GRE verbal score--z-scaled verbal formula scores selected so as to correspond to the upper 27%, middle 46%, and lower 27% of the U.S.-examinee distribution (assuming normality). The results are plotted in Figure 7a (based on detail provided in Table 7).

o Major-area effects were very consistent for the large ability-level subgroups in the U.S. sample; major-area differences for non-U.S. examinees were somewhat less regular and considerably sharper than those for U.S. examinees--a phenomenon alluded to earlier.

o Examinees with higher verbal ability tended to perform better on RC2 than on RC1, while the opposite was true for examinees in the lower ability subgroup.

Additional perspective on the foregoing is provided in Figure 7b (also based on detail provided in Table 7), which shows the means of major-area subgroups, by verbal score level, on RC2res--a residual variable reflecting the extent to which scores on RC2 differed from prediction based on RC1.

It is clear that the major-area pattern for RC2res (in Figure 7b) generally parallels that for RC2 minus RC1 (in Figure 7a). Major-area differences in mean RC2res are reduced somewhat in the U.S sample, presumably due to the introduction of control for differences in RC1.

Exploratory Assessment of Criterion-Related Validity

An exploratory analysis was made of the criterion-related validity of RC1 (20 items), RC2 (20 items), DVodd (28 items), and the total verbal score (based on 95 RC and DV items). The analysis was concerned primarily with obtaining evidence of the possibility of systematic differences in the level of correlation of these verbal subtests, especially RC1 (level) versus RC2 (speed), with the SR-UGPA criterion. The analysis was also concerned with evaluating the working hypothesis, based on evidence from previous research, of greater criterion-related

Fig. 7a. RC "speed" relative to RC "level" (mean RC2 minus mean RC1) for major-area subgroups, by level of GRE verbal ability (Vform): U.S. vs Non-U.S. examinees

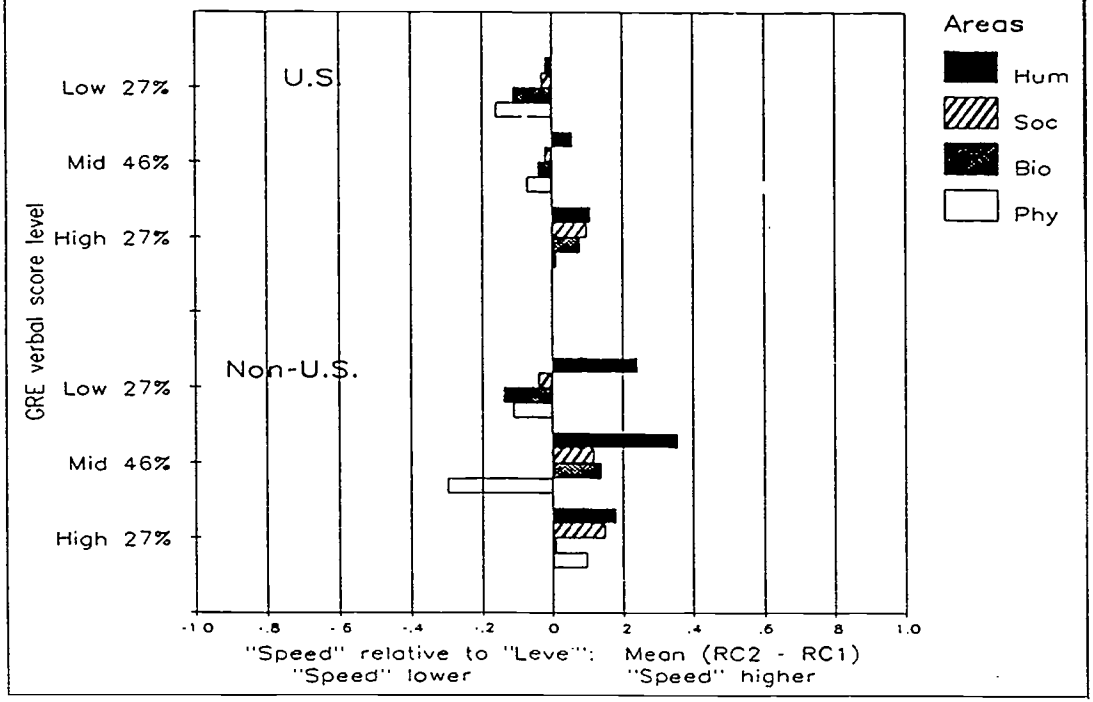


Fig. 7b. Mean RC2res for major-area subgroups by level of general GRE verbal ability (Vform): U.S. vs Non-U.S. examinees

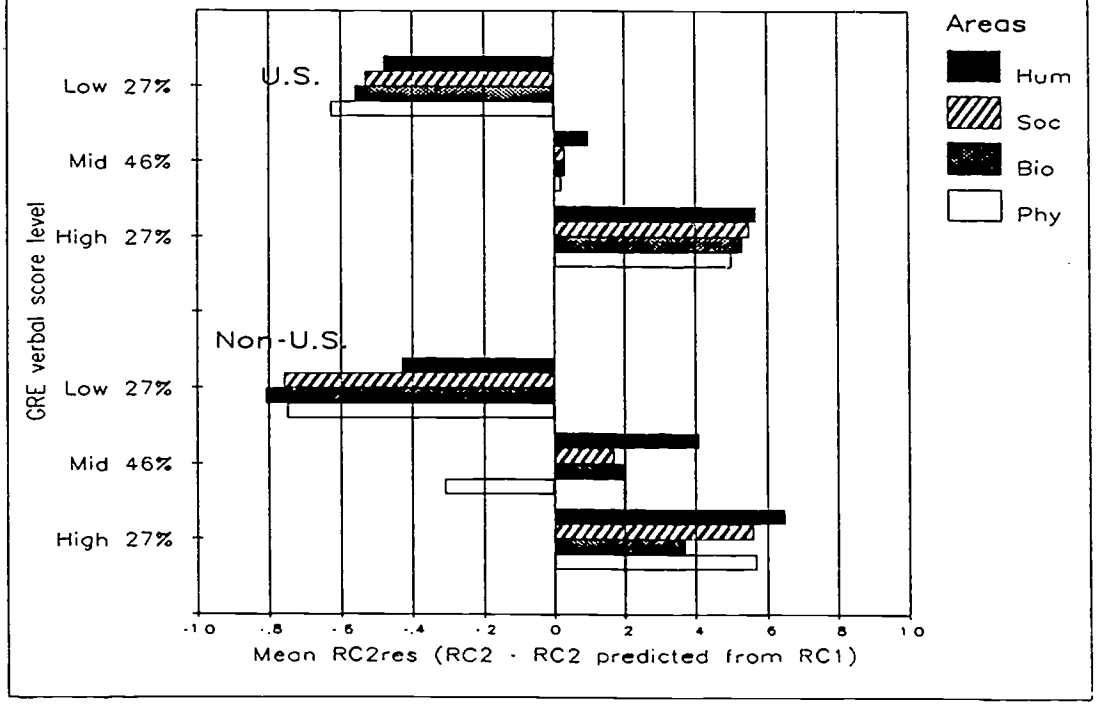


Table 7

Means on RC1, RC2, DVodd, and RC2res for Major-Area Subgroups by Total GRE Verbal Score Level

	N	RC1	RC2	DVodd	RC2res
U.S. (High V)					
Hum	1,382	.90	1.01	1.42	(.57)
Soc	2,329	.91	1.01	1.24	(.55)
Bio	957	.94	1.02	1.05	(.53)
Phy	892	1.05	1.06	1.13	(.50)
U.S. (Mid V)					
Hum	1,614	.06	.12	.09	(.11)
Soc	4,627	.13	.11	-.02	(.03)
Bio	2,317	.18	.14	-.11	(.03)
Phy	1,356	.26	.19	-.12	(.02)
U.S. (Low V)					
Hum	696	-.99	-1.01	-.91	(-.48)
Soc	3,302	-1.07	-1.10	-1.01	(-.53)
Bio	1,155	-.91	-1.02	-.96	(-.56)
Phy	470	-.93	-1.09	-.98	(-.63)
Non-U.S. (High)					
Hum	16	.88	1.06	1.44	(.65)
Soc	48	.78	.93	1.15	(.56)
Bio	14	.84	.83	.94	(.37)
Phy	25	.95	1.05	.83	(.57)
Non-U.S. (Mid)					
Hum	45	-.21	.15	.00	(.41)
Soc	105	.02	.14	-.14	(.17)
Bio	51	.03	.17	-.23	(.20)
Phy	60	.20	-.10	-.24	(-.31)
Non-U.S. (Low)					
Hum	64	-1.68	-1.42	-1.42	(-.43)
Soc	257	-1.54	-1.58	-1.50	(-.76)
Bio	130	-1.36	-1.50	-1.41	(-.81)
Phy	263	-1.32	-1.43	-1.50	(-.75)

Note: Examinees were classified according to level of total GRE verbal score: high 27%, middle 46%, and low 27% in the total sample. Disproportionately large numbers of non-U.S. examinees are in the low 27% of the verbal-score distribution.

validity for RC than for DV subtests. Generally speaking, expectation for the relative size of coefficients was:

Expectation for GRE/SR-UGPA correlations: RC1 = RC2 > DVodd.

GRE subtest/SR-UGPA correlations were computed for the principal subgroups of U.S. examinees and for selected subgroups of non-U.S. examinees.[26] Coefficients were computed separately for (a) U.S. and non-U.S. examinees classified by graduate major area, (b) for U.S. ethnic and gender groups by graduate major area, (c) for non-U.S. EPL and ESL examinees, and (d) for U.S. ESL examinees (only 429 of 21,097 U.S. examinees reported this status).

Coefficients obtained in large subsamples of U.S. examinees classified by graduate major area and by gender are shown in Table 8; Table 9 shows pooled within-major-area coefficients for non-U.S. examinees and for U.S. ethnic minority groups.[27] For the large sample of U.S. Whites, coefficients are shown separately by graduate major area. In addition, coefficients are shown for non-U.S. EPL and ESL examinees and for U.S. ESL examinees. The coefficients for Vform (total verbal score, 95 items) are shown primarily for perspective.

The last two columns of each table provide the evidence that is most pertinent for purposes of this study, namely, differences between coefficients for speed of comprehension (RC2) versus level of comprehension (RC1), in the "(b-a)" column, and for RC2 vs DVodd, in the "(b-c)" column. These differences are plotted in Figure 8.

Speed/level differences. Coefficients for RC1 and RC2 are of particular interest. For U.S. examinee subgroups--except ESL examinees and Hispanic examinees--coefficients for RC2 (speed), by and large, were higher than those for RC1 (level). For U.S. ESL and Hispanic examinees, as well as for non-U.S. examinees generally, and for EPL and ESL subgroups within the Non-U.S. sample, the opposite pattern prevailed--that is, coefficients for RC1 (level) were higher than those for RC2 (speed).

The emergence of systematic differences in criterion-related validity coefficients for RC1 and RC2 represents an unanticipated outcome.

- o The RC1 > RC2 pattern for non-U.S. examinees and for U.S. Hispanics suggests that RC scores obtained under speeded conditions may tend to to be less valid predictors of criterion performance than are scores obtained under unspeeded conditions in samples with large proportions of nonnative-English speakers.[28]

RC/DV differences. Scores on the GRE reading comprehension subtests were more highly correlated with the criterion than were scores on the discrete-verbal subtest (DVodd) in all comparisons except one (the coefficient for DVodd was slightly higher for U.S. Black examinees). This finding was expected for U.S. examinees; it proved to be true as well for non-U.S. examinees.

Table 8

Simple Correlations of RC1, RC2, DVodd, and Vform with Self-Reported Undergraduate Grade Point Average (SR-UGPA), by Graduate Major Area and Sex: U.S. Examinees Only

	N	Correlation with SR-UGPA				Difference	
		RC1 (a)	RC2 (b)	DVodd (c)	Vform	b-a	b-c
Hum.	3,692	.31	.32	.28	.34	.01	.04
Male	1,398	.38	.39	.33	.40	.01	.05
Fem	2,294	.27	.28	.26	.30	.01	.02
Soc	10,258	.35	.36	.32	.38	.01	.04
Male	4,599	.35	.37	.33	.39	.02	.04
Fem	5,659	.36	.36	.32	.39	.00	.04
Bio	4,429	.25	.27	.21	.28	.02	.06
Male	2,103	.22	.28	.20	.28	.05	.07
Fem	2,326	.27	.27	.22	.29	.00	.05
Phys	2,718	.28	.29	.20	.28	.02	.09
Male	2,121	.28	.31	.21	.30	.03	.10
Fem	597	.24	.22	.18	.24	-.02	.04
Number of items		20	20	28	95		

Note. Coefficients for RC1, RC2, and DVodd are for subtests of comparable length (and reliability, by inference). The coefficients for the total GRE verbal score are shown for perspective. It is noteworthy that in several instances the coefficient for RC2 or RC1 (20 items) is equal to or higher than that for the total verbal score (95 items).

Table 9

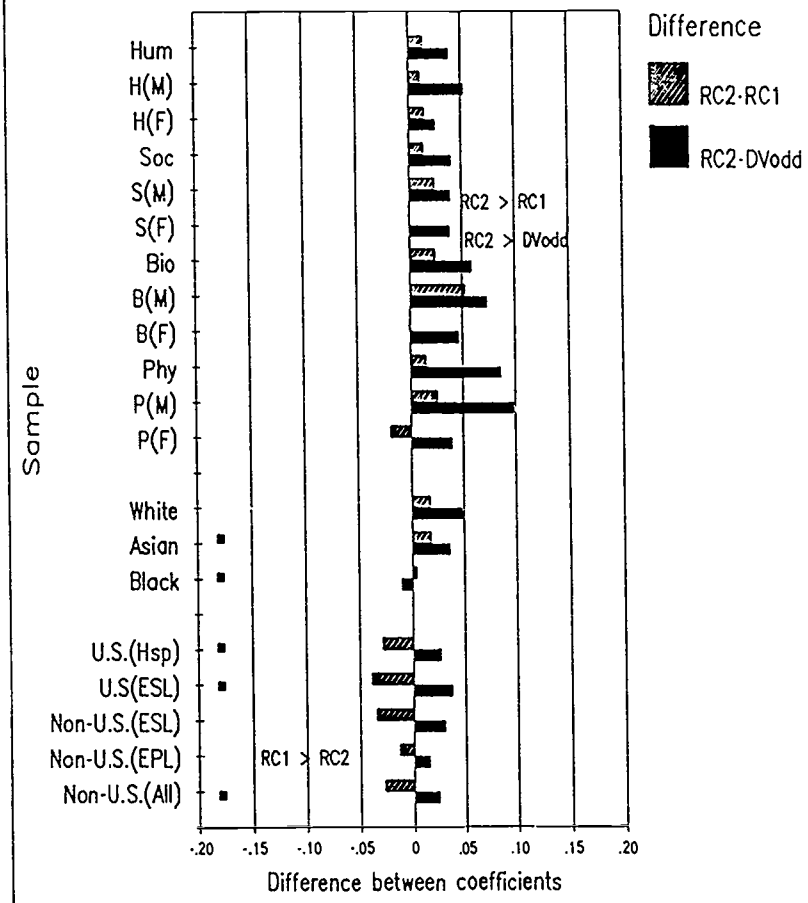
Simple Correlations of RC1, RC2, DVodd, and Vform with SR-UGPA in the Sample of Foreign Examinees and in Samples of U.S. Examinees, Classified by EPL/ESL Status and by Ethnic Group*

Group	N	Correlation with SR-UGPA				Difference	
		RC1 (a)	RC2 (b)	DVodd (c)	Vform	(b-a)	(b-c)
Non-U.S.	1,078	.33	.30	.27	.33	-.03	.03
ESL	584	.32	.28	.25	.32	-.04	.03
EPL	494	.34	.33	.31	.37	-.01	.02
ESL (U.S.)	429	.31	.28	.24	.30	-.04	.04
Hsp (U.S.)	434	.36	.33	.30	.38	-.03	.03
Bl**(U.S.)	929	.24	.25	.26	.28	.01	-.01
Asian (U.S.)	270	.29	.31	.26	.31	.02	.05
White (Hum)	3,383	.28	.30	.27	.32	.02	.03
(Soc)	8,963	.31	.33	.28	.35	.02	.05
(Bio)	4,064	.23	.25	.19	.26	.02	.06
(Phy)	2,480	.27	.28	.20	.26	.01	.08
(Total)	18,890	.29	.31	.26	.32	.02	.05

* Coefficients for non-U.S. examinees, and for U.S. Hispanic, Black, and Asian examinees, are pooled within-major-area coefficients (that is, they are size-adjusted averages of coefficients computed in subsamples classified by graduate major area); within-area coefficients are shown separately for the large sample of U.S. White examinees. Coefficients for EPL and ESL examinees, both U.S. and non-U.S., are based on samples that were not differentiated with respect to graduate major area.

** This coefficient does not include data for Black majors in biosciences (N = 151). All coefficients in this subsample were anomalously low or negative: -.061, .039, .011, and -.002, for RC1, RC2, DVodd, and Vform, respectively--coefficients for the total sample (N = 1,080) were .205, .218, .230, and .247. Thus, inferences regarding the direction of differences between coefficients are the same in both cases.

Fig. 8. Relative criterion-related validity of RC2 and RC1, and of RC2 and DVodd



■ Size-adjusted averages of coefficients by major area.

These findings indicate the presence of necessary conditions for inferring differential patterns of criterion-related validity for measures of speed and level of GRE reading comprehension. They do not appear to be due to statistical artifacts--differences in variability, for example (see descriptive statistics in Table 5).

Review and Evaluation of Findings

On the basis of evidence and lines of reasoning developed in detail at the outset, it was considered plausible that, due to the reasonable possibility of underlying differences in "speed of verbal processing," certain GRE population subgroups might tend to perform better on a measure of level of reading comprehension (administered under pure power conditions) than on an otherwise parallel measure administered under speeded conditions, and that the opposite pattern might obtain for other subgroups.

Operational Measures of Speed and Level

It was not feasible to develop parallel versions of a GRE reading comprehension test and administer them under untimed and timed conditions to representative samples of GRE examinees. Instead, based generally on the Cooperative Reading Comprehension Test precedent (see ETS, 1960a, 1960b), operational level and speed of comprehension scores were developed on a post hoc basis, using item-level data available in GRE files from a single, timed administration of a 40-item GRE reading comprehension section. The GRE level score (RC1) was based on the first 20 items, which most examinees were able to attempt within the RC-section time limit; the speed score (RC2) was based on items included in the second half of the test--a score that clearly included a speed-related, "items-not-reached" variance-component.

Hypotheses

Speed/level differences in performance. It was hypothesized that differences in average standing on speed relative to average standing on level would be found for (a) subgroups differing in English-language background (level score > speed score for nonnative-English speaking examinees) and for (b) major-area subgroups (speed score > level score for majors in primarily verbal fields and level score > speed score for majors in primarily quantitative fields).

Speed/level differences were not expected for groups defined by sex or by ethnicity: an a priori rationale for sex-linked or ethnic-group-linked differences in speed of processing verbal material is not readily apparent; and there is empirical evidence indicating that liberalization of time per question on experimental sections of verbal and quantitative measures did not differentially affect the average performance of either GRE subgroups defined by sex and by ethnicity (Wild & Durso, 1979), or similarly defined subgroups of SAT examinees (e.g., Evans, 1980).[29]

Reading comprehension versus discrete-verbal differences. On the basis of previous research involving GRE verbal-item-type part scores (e.g., Wilson, 1985a, 1986a), a systematic pattern of differences in standing on the GRE discrete-verbal (DVodd) subtest relative to standing on the GRE reading comprehension subtests (RC1, RC2, RCodd) was expected for major-area subgroups, as follows: RC1 > RC2 > DVodd for physical science majors and bioscience majors, and RC1 < RC2 < DVodd for humanities majors and social science majors.

Criterion-related validity differences. As to differences in criterion-related validity, on the basis of the studies cited above, it was expected that RC subtests would have higher correlations with SR-UGPA than would the DVodd subtest. There was no clear basis for expecting a particular pattern of differences in criterion-related validity for the RC1 and RC2 subtests. Thus, the expected outcome in terms of predictor-criterion coefficients was as follows: RC1 = RC2 > DVodd.

Findings Regarding Subgroup Performance

Speed/level. The patterns of average relative standing on RC1 and RC2 for the subgroups of GRE examinees involved in this study conformed very closely to the patterns hypothesized.

Major-area-related differences were systematic and pervasive.

- o The outcome "mean RC2 > mean RC1" (speed > level) was present for humanities majors generally, and for humanities majors classified by sex, ethnicity, EPL/ESL status, and general verbal ability level. The "mean RC2 < mean RC1" (speed < level) pattern was equally pervasive for physical science majors.

- o For social science majors, the "humanities pattern" tended to obtain, and the "physical sciences pattern" tended to obtain for bioscience majors; as expected, however, the observed RC2/RC1 inequalities were most clearly defined for the two major-area subgroups that are most clearly differentiated with respect to degree of emphasis on verbal processing, namely, humanities and physical sciences.

U.S. ESL examinees (representing a very small percentage of the total U.S. sample) had higher z-scaled means on Level (RC1) than on Speed (RC2); the hypothesized major-area-related Speed/Level inequalities were more sharply defined for ESL examinees than for examinees generally, and for subgroups of non-U.S. examinees than for the corresponding subgroups of U.S. examinees.

Speed/Level differences were not present for any U.S. ethnic group; slight Speed/Level discrepancies for U.S. examinees classified by sex plausibly reflect gender-related major-area effects: that is, proportionately more males than females were physical science majors, in both the U.S. and the non-U.S. populations.

Major-area-related Speed (RC2) versus Level (RC1) patterns were found to be consistent in analyses controlled for level of total verbal score. As expected, examinees in the "higher verbal" subgroup (upper 27%) had higher means on both RC2 and RC1 than did those in the "lower verbal" subgroup (lower 27%).

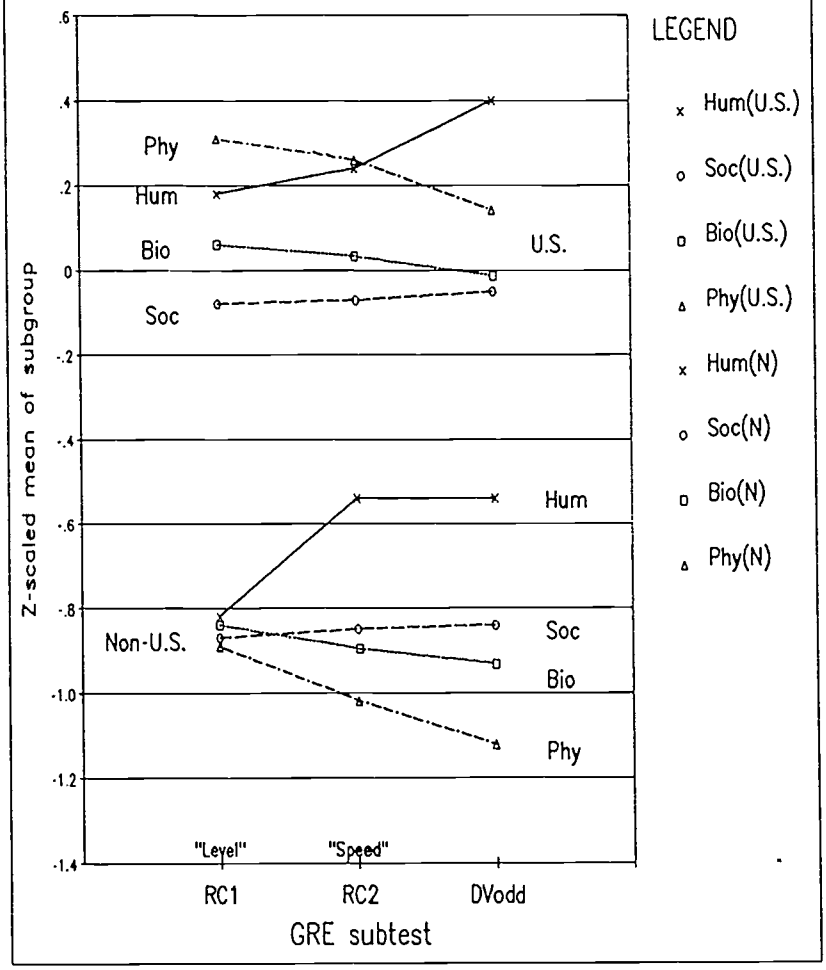
RC/DV performance differences. There were systematic major-area-related differences in relative standing on the GRE reading and discrete-verbal subtests, consistent with expectation. The pattern of "DV minus RC" outcomes paralleled the pattern of "RC2 minus RC1" outcomes consistently, across subgroups defined by sex and EPL versus ESL status (for U.S. and for non-U.S. examinees), and by ethnicity (for U.S. examinees).

Trends illustrated. The basic major-area patterns that have been alluded to are illustrated in Figure 9, which shows profiles of z-scaled means on Level (RC1), Speed (RC2), and DVodd for U.S. and non-U.S. examinees classified by major area. For both U.S. and non-U.S. examinees, it is evident that for humanities and social science majors, the "RC1 < RC2 < DVodd" pattern obtains, while for physical sciences and sciences and biosciences, the pattern is "RC1 > RC2 > DVodd." The patterns, of course, are most pronounced for humanities and physical science majors. Among non-U.S., but not U.S. examinees, major-area differences on RC1 are much less pronounced than are differences on RC2 and DVodd.[30]

In evaluating the superior performance of humanities majors on DVodd, note (from Figures 1a and 1b) the discrete-verbal subtest was more highly speeded than the reading comprehension subtest. From an "information processing" perspective, if the higher discrete-verbal scores of majors in verbal fields are thought of as indicating, among other things, that the examinees involved have more extensive vocabularies than their counterparts in the physical sciences, it is plausible that a more extensive lexicon may contribute to greater speed of reading with comprehension--for example, by facilitating speedier resolution of memory search phases of the reading process, by reducing the need to infer meaning of words from context, and so on.[31]

Interpretive perspective. The differences in average (z-scaled) standing on RC1 relative to standing on RC2 that are illustrated in Figure 9 constitute necessary conditions for inferring differences in level versus speed of reading comprehension in the various subgroups. However, it is important to recognize that such an interpretive inference, albeit plausible, involves an assumption that the RC1 and RC2 scores developed on a post hoc basis reasonably approximate scores obtained (a) under untimed and timed conditions, respectively, on (b) otherwise parallel versions of a GRE reading comprehension test. RC1 and RC2 clearly were not designed to be parallel tests of reading comprehension. A review of the properties of the two operational measures points up departures from strict parallelism that need to be taken into account.

Fig. 9. Means on RC1, RC2, and DVodd, for U.S. examinees and non-U.S. examinees, by major area



As to the first element in the basic assumption--that scores on RC1 and RC2 approximate scores obtained under untimed (power) and speeded conditions--on the basis of test analysis results, it is reasonable to infer only (a) that the RC1 scores were substantially free of speed-related, items not reached (INR) variance, and (b) that scores on RC2 included a significant INR-variance component--by inference from test analysis results, individual differences in INR "scores" ranged from 0 to 20, while almost all examinees completed RC1. At the same time, the test behavior was evoked under general time contraints (some time pressure was inherent in the test situation); also the test conditions permitted faster-working examinees to review their work on RC1 as well as on RC2.

As to the assumption that RC1 and RC2 are "otherwise parallel," it has been established (see Table 2 and related discussion) that the RC1 and RC2 subtests employed in the study were (a) of equal length, (b) balanced as to total amount of verbal processing required (as measured by total number of lines), and (c) about equally difficult (the mean formula score for RC2 was slightly higher than that for RC1). With respect to these important properties, RC1 and RC2 appear to be roughly parallel.

However, the reading passages in RC1 and RC2 were not parallel as with respect to either (a) subject matter or (b) style of writing. GRE RC sets are written in such a way as to assure that they conform to the basic assumption that the questions are answerable based solely on information provided in the reading passages (for example, ". . . questions are to be answered on the basis of information provided in the passage" [ETS, 1988: p. 32]).

Acceptance of the validity of this assumption does not, of course, rule out the plausible influence of differences in "prior knowledge structure" on the speed with which examinees from different disciplines are able to process passages with subject matter from their respective disciplines.[32] Thus, we cannot rule out the possibility that the observed RC2 versus RC1 outcomes reflect to some extent interactions between passage characteristics and both major area and EPL/ESL status.[33]

Direct empirical evidence bearing on these possibilities does not appear to be available for GRE test takers.[34] Given the subject matter of the passages in RC1 and RC2, however, the observed patterns of performance by major area do not suggest the presence of interactions between subject matter and major area--for example, humanities and social science majors performed relatively less well on RC1 (which included passages from the humanities and social sciences) than on RC2.

Of course, interactions between stylistic emphasis in reading passages and examinees' major fields or their linguistic backgrounds cannot be ruled out. It is conceivable, for example, that passages written in narrative or argumentative style (used for two of the three passages in RC2) may tend to be relatively more difficult for majors

in the physical sciences, or for foreign-ESL examinees (RC2 < RC1) than for either majors in the humanities (RC2 > RC1), or foreign-EPL examinees (RC2 < RC1).[35]

Thus, ambiguities due to lack of parallelism in the operational measures employed in this study complicate what appears to be a generally plausible interpretation of the findings as reflecting speed versus level differences in GRE reading comprehension for subgroups defined by major area, and by EPL versus ESL status--and, of course, the absence of such differences in the case of subgroups defined by sex or ethnicity.

Findings Regarding Criterion-Related Validity

In the exploratory assessment of criterion-related validity in variously defined subsamples of U.S. and non-U.S. examinees, it was found that, in all but one of the subsamples, GRE RC subtests (20 items) were more highly correlated with SR-UGPA than was the GRE discrete-verbal (DVodd) subtest (28 items). This was consistent with expectation based on previous research (Wilson, 1985a, 1986b; also Wild, McPeek, & Koffler, 1988) on the relationship of GRE verbal item-type part scores to self-reported UGPA. In a number of instances, the coefficient for a 20-item RC subtest was approximately equal to or slightly higher than the coefficient for the total 95-item GRE verbal score. However, contrary to expectation, two distinct patterns of differences in criterion-related validity were observed for the operational measures of level (RC1) and speed (RC2). On the one hand, in subgroups of U.S. examinees (except Hispanics and ESL-examinees), RC2 was more closely related to the criterion than was RC1; on the other hand, for subgroups of foreign examinees, and for U.S. Hispanics and ESL examinees, the opposite validity pattern was observed.

For the findings indicating a higher degree of criterion-related validity for GRE reading comprehension items than for GRE discrete-verbal items, there is both (a) relatively clear empirical precedent and (b) a generally straightforward and plausible explanatory rationale. For the unexpected, systematic patterns of differential criterion-related validity for RC1 and RC2 neither of the foregoing is present. It seems useful, therefore, to evaluate the "expected" pattern of findings first and then turn attention to the more complex, "unexpected" pattern.

General interpretive rationale for RC versus DV differences. Higher correlations for reading comprehension subtests than for the discrete-verbal subtest appear to be understandable on the basis of differential degrees of direct overlap between the types of tasks represented by test items and the types of tasks that students perform in carrying out their academic assignments.[36] Generally speaking, predictive validity should tend to increase as the resemblance between the test situation and the criterion situation increases, and vice versa.[37]

The GRE reading comprehension sets appear to represent a "standardized work sample" of a complex functional ability (involving numerous component elements) that examinees exercise naturally in completing their academic assignments. The discrete-verbal items are measuring verbal skills that contribute to the general functional ability (reading), as well as to the performance of related verbal reasoning tasks.

It thus seems logical that a GRE reading comprehension subtest, as an essentially direct measure of a complex functional ability that is used in the criterion context, should tend to be more closely related than a GRE discrete-verbal subtest that provides indirect measures of important component abilities. Thus, there is a plausible explanation for the differences in correlations for RC and DV substests. However, "explaining" the unexpected patterns of differences in level of correlation for RC1 and RC2 with SR-UGPA is not so straightforward. Two patterns of differences require explanation.

1. Coefficients for RC2 were relatively consistently larger than those for RC1 in U.S. subgroups, except for (a) ESL examinees (individuals who report that they communicate better in a language other than English), and (b) Hispanic-American examinees, more than one fifth of whom reported ESL status.

2. Coefficients for RC1 were consistently larger than those for RC2 (a) for the partially overlapping subgroups of ESL and Hispanic American examinees and (b) for subgroups of non-U.S. examinees generally, and in classifications according to EPL and ESL status.

Interpretive perspective based on previous research findings is limited. For example, results of the most directly pertinent validity studies--that is, studies of the criterion-related validity of conceptually and, in a sense, operationally, comparable Speed and Level scores on the Cooperative Reading Comprehension Test in college-level and secondary-level samples (ETS, 1960b)--do not indicate any systematic pattern of differences in coefficients for Level and Speed. And, generally speaking, questions regarding the comparative validity of differentially speeded, but otherwise parallel cognitive tests for predicting academic criteria have not received much attention, and few empirical studies have been designed to provide answers to such questions. In fact, during the course of this study, no studies were located that dealt with the comparative criterion-related validity of unspeeded and speeded, or differentially speeded, reading (or other) tests in samples differentiated in terms of EPL versus ESL status.

Literally interpreted, the findings indicate that in samples made up predominantly of native-English-speaking U.S. examinees, the speeded RC2 subtest was more closely related to the external academic criterion than was the unspeeded RC1 subtest, but that the opposite was true in samples that included a significant proportion of nonnative-English-speaking examinees. In other words, it appears that "speed of response variance" in a GRE reading comprehension measure may contribute to its

criterion-related validity in samples of native-English speakers, but diminish its criterion-related validity in samples of nonnative-English speakers.[38]

In evaluating this working hypothesis, it is useful to examine first some empirical evidence indicating clearly that it is plausible to posit positive (validity-enhancing) effects for a speed component in GRE reading comprehension scores. We can then attempt to rationalize the negative (validity-diminishing) effects posited for speed as a component in the reading scores of nonnative-English speaking examinees.

Evidence of positive aspects of speed in cognitive tests. In evaluating the proposition that a speed component in reading or other verbal ability tests may tend to enhance validity (in samples in which developed native-language verbal skills are being assessed) it is pertinent (a) to re-examine and elaborate somewhat on the findings of Lord's (1956) study of speed factors in tests and academic grades in a sample of undergraduate level students and (b) to consider, in some detail, evidence regarding the comparative predictive validity of a purposely speeded reading comprehension measure, namely, the Reading Comprehension section of the Secondary School Admission Test (SSAT) (ETS, 1987) and less speeded verbal and quantitative sections of that test.[39]

(Lord, 1956). As noted at the outset, Lord analyzed scores on short, differentially speeded, but otherwise parallel, verbal, spatial, and arithmetic reasoning tests, and end-of-course grades in several subject areas, for a large (N = 649) sample of U.S. Naval Academy students. Scores on relatively highly speeded verbal tests had higher simple correlations with the GPA criteria than did scores on less speeded tests. Lord found four "speed factors" (called number-speed, perceptual-speed, verbal-speed, and spatial-speed).

> (The primary vectors) were found to be positively correlated, demonstrating the existence of a general speed factor at the second-order level. All correlations between course grades and the four speed factors, with one small exception, were found to be positive, although not large. It is to be concluded that speed of various kinds plays some part in the course grades studied, and that speededness in the admissions examinations is to this extent justified (p. 49).

Lord noted that very highly speeded tests apparently were needed to evoke the pertinent factors. It is of incidental interest that no "arithmetic-reasoning speed" factor was identified.

The experimental verbal tests used by Lord were composed exclusively of items requiring examinees to find " . . . among the choices a word opposite in meaning to the given key word" (p. 33)--that is, they were pure vocabulary tests. Thus, the findings should be thought of only as providing evidence of the presence of a speed component in

both vocabulary tests and grades in the sample studied. Unfortunately (for present purposes) no reading comprehension tests were included in the analysis.

(Secondary School Achievement Test validity data [ETS, 1987]). Evidence regarding the comparative validity of SSAT reading compreension (RC) and verbal aptitude (V) scores appears to be quite pertinent, despite the fact that it is based on data for early-secondary-school-level samples, for the following reasons:

1. The SSAT Reading Comprehension measure evolved along psychometric lines represented by the Cooperative Reading Comprehension Test model that provided the conceptual basis for the operational measures of speed and level developed for the present study.

2. The SSAT RC test historically has been designed to generate a score reflecting " . . . the ability to read rapidly with understanding" (e.g., ETS, 1969)[40] and it is considerably more speeded according to ETS criteria than either the SSAT verbal aptitude or the quantitative ability measure. Based on results of internal test analyses, for example, the percentage of examinees completing the SSAT RC test typically is only about half as great as the percentage completing either the verbal or the quantitative sections.

3. A consistent distinction has been maintained between "reading comprehension" and "verbal aptitude" (defined by antonym and analogy items) for purposes of test development and score reporting.

The validity study findings outlined below appear to be most pertinent.

o Grades in ninth-grade English and mathematics courses, and an overall GPA, were employed as criteria in a study involving a sample of 1,182 students from 21 SSATB member schools (ETS, 1987: pp. 13-15). For present purposes it is sufficient to consider selected regression findings for these criteria (using pooled-within school data for the total sample), shown in Table 10.

o The regression coefficient for SSAT-RC was larger than that for the Verbal Aptitude (V) measure regardless of the criterion under consideration. The regression coefficient for the verbal measure was negative in the analysis involving Math GPA as the criterion. RC and V were relatively closely related (r = .78), and the simple correlation of RC with the criterion was higher than that for V (coefficients were .35 [RC] and .30 [V]) resulting in a "suppression" effect.

These results unambiguously extend evidence indicating that reading comprehension measures tend to be more valid than discrete verbal measures (in this case, a measure composed of antonym and analogy item types) for predicting academic performance criteria.[41]

Table 10

Regression Results for SSAT Reading, Verbal, and Quantitative Scores in Analyses Involving Designated GPA Criteria (from ETS, 1987, Table 8)

Criterion	Standard regression wts. RC	V	Q	Multiple correlation
English GPA	.25	.14	.26	.56
Math GPA	.14	-.06	.50	.54
Overall GPA	.22	.04	.38	.56

Note. All SSAT measures are formula scored. The SSAT-RC measure is more speeded than either the SSAT verbal (antonyms and analogies) or quantitative test.

With regard to the contribution of "speed," it is clear only that the SSAT-RC measure is considerably more highly speeded than the SSAT-V measure, and that it appears to be more valid for predicting grades. Whether differences in speededness contributes to this result is not clear, of course. It is possible that scores obtained under "pure power" conditions (or under more highly speeded conditions) might tend to be more valid than those obtained under current conditions that lead to RC scores with some difficult-to-measure mixture of speed and power. Validity data for early versions of the SSAT--for which both a level-of-comprehension score and a speed-of-comprehension score were reported--are limited and do not help to resolve the question at issue here (see, for example, Pitcher, 1962, for results for two schools).

On balance, it is believed that the foregoing evidence lends credibility to the interpretive inference that the higher correlations of RC2 scores than RC1 scores with the SR-UGPA criterion may be attributed, at least in part, to the fact that there was a larger "speed of response" component in RC2 scores than in the RC1 scores.

Why higher validity for "power-like" scores for nonnative-speakers? On the basis of the evidence and lines of reasoning developed above, it appears plausible that the speed-of-response-variance component in RC2 scores had a validity-enhancing effect in samples composed predominately of native-English speakers.[42]

But why should the speeded RC2 scores be less valid than RC1 scores, obtained under "power-like" conditions, for predicting the same criterion in samples that included significant proportions of nonnative-English speakers? One line of reasoning about this outcome involves the following assumption: For nonnative-English speaking examinees, but not for native speakers, RC2 is a significantly less reliable measure of reading comprehension than is RC1. What is the

basis for this assumption? Briefly, it rests on the following line of reasoning:

o On logical and evidential grounds, verbal admission tests such as the GRE or the SAT--tests that are administered with time limits set so as to minimize speed-of-reponse variance in the general examinee population (predominately native-English speakers)--may be expected to be substantially more speeded (by usual test-completion criteria) for nonnative-English speaking examinees.

o Due to slow average speed of processing general reading matter in English, nonnative speakers as compared to native speakers were able to attempt proportionately fewer items in the second half of the RC test. As a consequence, RC2 was, in effect, a substantially shorter, less reliable measure for nonnative- than for native-English speaking examinees.

Comparative test-completion data for U.S. GRE examinees and a sample of foreign-ESL examinees (Angelis, Swinton, & Cowell, 1979: p. 30) are directly illustrative.

o For GRE reading comprehension (40 items, in the pre-October 1977 separately timed format), completion indices for foreign-ESL and native-speaking examinees, respectively, were as follows: completed the RC section (47% ESL versus 61% EPL); completed 75% of the questions (76% versus 95%); items reached by 80% of the examinees (35 of 40 versus 27 of 40). Estimated reliabilities for the entire 40-item RC section were .84 and .47 for native-speaking (EPL) and nonnative-speaking (ESL) examinees, respectively.

On the basis of the foregoing, higher validity for RC1 than for RC2 in samples of non-U.S. citizens, and in samples of Hispanic Americans and U.S. ESL-examinees, may be explained in terms of relative measurement efficiency: RC1 was a longer, more reliable measure of reading ability than was RC2 in the samples that included nonnative-English-speaking examinees; other things equal, increasing the length of a homogeneous test adds to its reliability and validity.[43]

Generally speaking, it seems reasonable to assume that in samples of nonnative-English speakers, the ability to comprehend and answer questions about GRE reading passages, is likely to be measured more validly and efficiently under power conditions than under speeded conditions. It has been argued elsewhere (for example, Wilson, 1984) that due to the likelihood of an atypically large speed component--associated with less-than-native levels of proficiency in English--the GRE verbal scores of foreign ESL examinees may tend to underestimate their ability to perform relevant academic tasks. Why? In part, because " . . .(u)nder normal conditions of academic life, foreign ESL students typically may be able to compensate for relatively low speed of English language verbal processing (e.g., reading speed) by additional time on task (pp. 21-22). There is evidence suggesting that foreign ESL examinees may tend to earn somewhat higher grades than

their U.S. counterparts, despite markedly lower average scores on GRE verbal and analytical ability measures (Wilson, 1986b: p. S-9; passim).

The lines of reasoning introduced in evaluating the evidence of differences in patterns of validity for RC1 and RC2 have included consideration of possible effects associated with lack of parallelism in subject matter and writing style for reading passages in RC1 and RC2. In essence, the issue of parallelism does not appear to have any (easily discernible) bearing either on the differential levels of criterion-related validity for RC1 and RC2, or on the finding of consistently higher validity for GRE reading comprehension items than for GRE discrete verbal items. RC2 was slightly more difficult than RC1. Thus, effects associated with the differences in difficulty cannot be ruled out.

Research Needed to Resolve Ambiguities

Generally speaking, despite the venerable status of the topic of "speed versus level of reading ability," remarkably little directly relevant evidence appears to be available to help resolve the ambiguities that have been noted. There have been no previous studies of major-area-related or EPL/ESL-related differences in performance on differentially speeded GRE measures (including reading comprehension subtests). Similarly, no previous work appears to have been undertaken for the purpose of assessing the effect on predictive validity of increasing or decreasing the speed component in scores on reading comprehension, or on any separately timed section of any of the ability measures provided by the GRE General Test (or similar tests such as the SAT and the GMAT).

Studies involving GRE reading comprehension subtests that are differentially speeded, but otherwise parallel, are needed to evaluate the tentative speed-versus-level interpretation of the findings of this exploratory study with respect to (a) RC2/RC1 differences in average performance for subgroups and (b) differences in criterion-related validity, favoring RC2 for native-English speaking examinees, and RC1 (level) for examinees for whom English is not the native language.

Further research is needed to resolve interpretive ambiguities associated with lack of parallelism in the content of the measures used in this exploratory study. A model involving the development of parallel versions of reading comprehension subtests, to be administered in differentially timed experimental sections of the GRE, would seem to be appropriate--an adaptation of the model employed by Wild and Durso (1979) in studying the effect of changes in time limits on subgroup performance, for example.[44]

It is important to obtain concurrent data for an essentially unspeeded RC measure, an RC measure reflecting "normal" time-per-item conditions, and one or more relatively highly speeded RC measures. It would be useful to assess the differential criterion-related validity of such measures in each of the subgroups defined for this study.

The findings of this study reflect conditions--population characteristics, test formats, and so on--that have changed in significant ways since October 1977 (the date of the operational test administration that generated the data employed in this study). The fact that the GRE verbal measure no longer includes separately timed reading and discrete verbal sets, for example, forecloses the possibility of assessing the replicability of the findings using current operational data. However, another study involving older test data could be designed to provide evidence bearing on the generalizability of the findings of this study, possible interactions between "passage characteristics" and membership in subgroups such as those defined for this study, especially subgroups based on discipline.

Extending speed/level inquiry to other GRE ability domains. This study has been concerned exclusively with the GRE verbal measure and, insofar as the speed/level questions are concerned, only with evaluating hypotheses involving speed versus level of GRE reading comprehension. GRE reading comprehension was the logical choice for this exploratory inquiry: the concept of assessing individual differences in rate or speed of reading (with understanding) is well established, and a model was available for the purpose of generating plausibly interpretable "level" and "speed" of reading comprehension scores from data available in GRE files.[45]

However, there is a speed component and a power component in scores on the items in each timed section of each GRE ability measure, as well as in the respective total ability scores. It is possible that there may be population differences in relative standing on differentially speeded, but otherwise parallel versions of subtests based on GRE quantitative or analytical ability item types. Such subtests may prove to be differentially valid for predicting external criteria.

It is logical to extend speed/level inquiry to the other ability domains tapped by GRE General Test items. Use of last-item-attempted (LIA) indices, reported by Lord (1967), to be the purest measure of "speed," is complicated by "rights only" testing conditions. At the same time, it is possible to conduct studies of relationships among "LIA scores" for separately timed GRE sections, using data from pre-rights-only test administrations.

Concluding Observations

The findings of this exploratory study, apart from issues pertaining directly to speed versus level of reading comprehension, add to a growing body of evidence indicating that a useful distinction can be made between GRE reading comprehension sets and GRE discrete-verbal sets. There are pervasive major-area-related differences in performance on these item sets, and there is evidence (based on self-reported UGPA) suggestive of differential predictive validity for subtests based on these item types--"suggestive" only, because this has not been demonstrated to be true in predictive studies involving graduate-level

GPA criteria. Graduate-level predictive validity data for GRE reading comprehension and discrete-verbal subtests are needed. However, the available evidence appears to be reasonably persuasive.

Lord and Wild (1985) concluded that " . . . reading comprehension is measuring something different from what is being measured by the other verbal item types" (p. 18). Factor analysis results have identified a factor defined primarily by reading comprehension sets in both the GRE examinee population (e.g., Kingston & Dorans, 1982) and the SAT examinee population (Dorans & Lawrence, 1987)--who suggested that if the format of the SAT were to be revised, adding more reading comprehension items would seem to be desirable.

GRE reading comprehension sets have clear face validity: they are measuring under standard conditions a complex functional ability that is exercised naturally by students in performing comparable aspects of their academic work. On balance, such evidence lends support to the notion that the overall utility of the GRE verbal measure would be enhanced by reporting a score based on the reading comprehension sets and a score based on the discrete-verbal sets--reading comprehension and verbal ability scores (along lines represented by the SSAT model, for example).[46]

Clarifying the Role of "Speed" and "Power" in GRE Scores

The GRE General Test is intended to measure "level of developed ability" (amount of knowledge, skill, understanding, and so on), in ability domains represented by specific combinations of verbal, quantitative, and analytical reasoning item types. However, because significant numbers of examinees are unable to attempt all the test items within specified time limits (limits that are set for practical, administrative reasons), there is a "speed of response" component, as well as a "power" or "level of ability" component, in score distributions generated for each separately timed section of each of the GRE ability measures.

The pragmatic response to this "dilemma"--that is, the presence of an apparently inescapable "speed" component in a test that is intended to be a test of "power"--has been to adopt procedures designed to standardize the amount of speed-related variance in successive editions of the GRE General Test. Each separately timed section of each ability measure in each edition of the GRE General Test (and other major ETS-based admission tests as well) is expected to meet a common set of test-completion standards.

No clear a priori or evidential grounds have been advanced for what appears to be an implicit assumption that GRE scores that might be obtained under "pure power" conditions are likely to be more valid for intended purposes than are GRE scores obtained under currently speeded conditions, or scores that might be obtained under more-speeded conditions.

In order to clarify the role of speed in GRE scores, it is important to advance explicit theoretical and pragmatic arguments for eliminating, varying, or continuing to standardize the speed component in GRE scores--from the perspective of effects on "validity for intended purposes."[47] A strong rationale for action along these lines has been offered by Donlon (1980, p. 1)).

> There are three broad reasons for attending to speed and power: (1) issues of fairness or equity, (2) issues of psychometric efficiency, and (3) issues of administrative efficiency. These three facets of the problem are interrelated but differentiable. A testing program may design its tests, or modify them, with respect to speed and power in order to achieve goals in each of these three areas.
>
> The first area, the notion of fairness or equity, is a fundamental one. If two candidates work on an examination of 100 items for 40 minutes, and candidate A reads and responds to 80 items while candidate B reads and responds to 40 items, there is a clear potential advantage to A. The test developer cannot overlook this possible advantage. Did B understand the test? Is B familiar with the testing situation? If B has truly a characateristically slower rate of work that is not easily accelerated, is the resulting distinction between B and A valid, in the sense that B will not do as well on criterion tasks for which the test may be predictive?
>
> This note of validity blends the discussion of equity into the discussion of psychometric efficiency. Speed and power can be established as separate factors or sources of variance in test scores. To the extent that these factors are differentially predictive of a criterion, they may be differentially valuable to the test designer. If a speeded test performance is more predictive of a criterion than a power test performance is, then the test planner will establish conditions, in terms of number of items and time allowed, that foster a speeded performance.
>
> Even if there is no difference between a speeded test and a power test in the prediction of criterion performance, the time efficiency of a speeded test may be of value. That is, it may offer more measurement time per minute. In the design of multi-test batteries, requiring several hours of testing, where the proper allocation of time to tests is a problem, the ability to elect a speeded test may be a distinct advantage (p. 1).

The results of the present study suggest that there may be some "psychometric-efficiency merit" in generating two GRE reading comprehension scores--one with a "speed of response" component and the other without such a component. Is it not plausible that this may be true, to some extent, for scores in other GRE ability domains?

In stating the aims of his study of speed factors in tests and academic grades, Lord (1956) commented as follows:

> Much remains to be learned about 'speed,' in spite of the fact that it is commonly an element in test scores. Is speed on cognitive tests a unitary trait? Or are there different kinds of speed for different kinds of tasks? If so, how highly correlated are these different kinds of speed? How highly correlated are speed and level on the same task? How do various criteria relate to speed, and how speeded should tests to predict these criteria be? (p. 31).

This constitutes what appears to be a challenging, currently pertinent agenda for "speed-related research" involving GRE verbal, quantitative, and analytical ability subtests. The research questions are framed from the perspective of differential psychology. It seems probable, however, that the most satisfactory models for investigating them (a) will reflect both psychometric and cognitive-process perspectives, and (b) will be developed most effectively within the framework of interactive testing models that will permit the assessment of both performance and process.

NOTES TO TEXT

1. The GRE General Test also provides measures of quantitative and analytical abilities. The present study, however, is primarily concerned with analyses based on items included in the GRE verbal measure.

2. A "last-item-attempted" (LIA) model has been employed for assessing test completion. The last item marked by an examinee is considered to be the last item attempted. Use of the LIA model for monitoring test completion rates is limited under "rights only" scoring conditions, currently in effect for the GRE. However, these limitations are not directly at issue in the present study.

3. The GRE verbal measure as currently formatted includes two separately timed sections, each made up of a balanced representation of antonym, sentence completion, reading comprehension, and analogy questions, in the sequence indicated. For purposes of the present study, it is necessary to assess performance on a timed GRE reading comprehension section, hence the need to employ pre-October 1977 data.

4. Selected results of research conducted by Hunt (1978), involving University of Washington (UW) students, illustrate basic patterns of findings regarding relationships between "decoding time" (with simple verbal materials) and performance on a verbal ability test (reported [p. 109] to be comparable psychometrically to the SAT verbal measure). Verbal scores of UW students on the Washington Pre-College Test (WPCT), taken by them as high school juniors, were related to an "NI-PI" index: "NI-PI = (reaction time required to classify an item as "same" under name identity instructions) minus (reaction time required to classify an item as same under physical identity instructions)." "Aa" is illustrative of a name identity item, and "AA" is illustrative of the physical identity counterpart. A negative correlation with verbal ability is expected for this index if high verbal ability is associated

with rapid decoding. Hunt reported correlations of about -.30, typically, between the NI-PI index and WPCT verbal scores (presumably obtained under "speed" rather than "pure power" conditions, by inference from reported similarity to the SAT verbal measure).

5. In a study by Wild and Durso (1979), experimental sections of the GRE verbal and quantitative measures were administered with a 20-minute time limit, representing the same time-per-question allotment as in the corresponding operational sections, and with a 30-minute time limit. Although a larger proportion of examinees completed the experimental tests under the more liberal time limits, it was found that the extra time did not differentially help any of the subgroups involved. A similar pattern of findings was reported by Evans (1980) for a similarly designed study involving samples of students taking experimental verbal and mathematical sections of the Scholastic Aptitude Test, classified by sex, ethnic group membership, and rural versus urban environment. The tests were administered under 20-, 30-, and 40-minute time limits; there were no significant interactions involving speed and group membership.

6. Simple correlations of designated test variables, including the verbal section of the Naval Academic admission battery, are summarized below. Underscoring indicates the highest simple correlation.

End-of-course GPA	Test variables* 2	3	4	5	6	7	8	9
English	.560	.497	.568	.537	.519	<u>.590</u>	.540	.373
For Lang	.210	.172	.205	.192	.186	.220	<u>.226</u>	.204
Engin. Dr.	.184	.084	.186	.138	<u>.247</u>	.192	.221	.182
Chemistry	.230	.172	.238	.196	<u>.270</u>	.258	.248	.228
Math	.156	.119	.145	.128	<u>.213</u>	.211	.210	.258

*2. Regular verbal admission test (analogies and sentence completions)
3. Unspeeded antonyms (<u>15</u> items/ <u>7</u> minutes/ <u>97%</u> finishing)
4. Unspeeded antonyms (% finishing = average for 3 and 4)
5. Moderately speeded antonyms (<u>30</u> / <u>5</u> / <u>71%</u>)
6. Speeded antonyms
7. Speeded antonyms (<u>75</u> / <u>5</u> / <u>2%</u>: average of 6-8)
8. Speeded antonyms
9. Last item attempted score for test 7.

7. For a number of years, separate SAT reading comprehension (RC) and vocabulary (VO) scores, as well as a total SAT verbal score, have been reported. The RC score is based on both reading comprehension and sentence completion items; the VO score is based on the antonym and analogy items.

8. These findings were anticipated on the basis of evidence provided by Ramist (1981a, 1981b) indicating that the formally reported reading comprehension subscore of the Scholastic Aptitude Test (SAT) verbal

measure (based on sentence completion and reading comprehension items) was (a) more valid than the vocabulary subscore (based on antonym and analogy items) for predicting college GPA, and (b) as valid as the total SAT verbal score for predicting this criterion. See also Note 41.

9. The Cooperative Reading Test is included in the Cooperative English Tests (CET) series. The present study is concerned only with those aspects of the CET that make up the reading comprehension component. Other components of the CET series are described in the references cited.

10. The CRCT technical manual (ETS, 1960b) cites illustrative test completion rates for the sections that contribute to the total Reading score. For example, in one study, 78% of college freshmen completed the vocabulary test; 93 percent finished the first 30 RC items, but only 15 percent reached item 60 in the RC section. In this sample, intercorrelations were r(v,l) = .71, r(v,s) = .74, and r(l,s) = .83, where v = Vocabulary, l = Level, and s = Speed. Correlations of V, L, and S scores with scores on the School and College Ability Test were .88, .76, and .79, respectively. Alternate form reliability coefficients for the 30-item Level score were in the mid-.70s, lower than those for the Speed and Vocabulary scores, each based on 60-item tests.

11. In studies where validity coefficients were reported for the Speed and the Level score, the coefficients were about the same--noteworthy, in part, because the Level score is based on a test of only 30 items whereas the Speed score is based on a 60-item test that includes the 30-item subtest. This pattern is illustrated in a study (Frederiksen, 1952) involving a relatively large sample (N > 400) of Princeton University freshmen.

o Correlations with first-year grades and with SAT-verbal scores were reported for this sample, as follows:

	Correlation with Vocab	Speed	Level	Total
Grades	.38	.36	.37	.44
SAT-V	.81	.68	.60	.80

The speed/level coefficient was .65; as compared to .83 in the less highly selected "technical manual" sample (ETS, 1960b). Coefficients with first-year grades were very similar for the three components of the total reading scores (V, S, and L); the coefficient for the 30-item Level score was comparable to that for scores on Vocabulary and Speed of Comprehension (both based on "speeded" performance on 60-item tests).

12. Investigators concerned with the effects of the degree of test speededness on predictive validity (e.g., Kendall, 1964; Lord, 1956) have noted that questions regarding the effect of degree of speededness

on predictive validity need to be addressed empirically--that is, it should not be assumed that the validity of aptitude tests decreases as their degree of speededness increases.

13. Experimentally independent speed or rate scores have been developed in a variety of ways. For example, in the Gates-MacGinitie Reading Tests: Survey F (Gates & MacGinitie, 1969), "speed" and "accuracy" scores (as well as vocabulary and comprehension scores) are obtained. The speed and accuracy scores are based on 36 items to be completed in four minutes. See Buros (e.g., 1965, section on "Tests & Reviews: Reading") for other examples.

14. As presently constituted, the verbal section of the GRE General Test has two 30-minute sections composed of 38 questions each: 7 sentence completion, 9 analogy, 11 reading comprehension, and 11 antonym questions. It is thus not possible to develop subscores reflecting performance on a specifically timed reading comprehension test using a current operational form of the General Test.

15. For each test form, ETS routinely conducts standard item analyses (IA) designed to provide evidence about the difficulty of each item, proportions choosing various options, percent reaching each item, and so on. Data are analyzed for each separately timed test section by level (quintiles) based on the total score for the ability involved. The IA results reported herein for reading comprehension and discrete verbal items were based on a sample of 1,960 examinees, 392 from each quintile based on the total verbal score.

16. "Although all questions in the verbal test necessarily refer to some area of human thought, answering questions correctly does not depend upon specific subject-matter knowledge in any of these areas, other than a reasonable familiarity with the basic elements or processes in a particular area. Rather, to the extent that each question draws upon subject-matter domains, the question or its related stimulus material [e.g., content of a reading passage] provides the context or information necessary to furnish the subject-matter background for answering the question" (from internal ETS documentation of test specifications). Of course, even if prior knowledge is not required to answer the question, examinees who happen to have relevant prior knowledge may benefit therefrom--for example, they may be able to process related verbal material more efficiently.

17. GRE examinees are asked, "Do you communicate better in English than in any other language?" Those who answer "Yes," include both native-English speakers and nonnative speakers. The verbal performance of foreign native-English speakers is fully comparable to that of U.S. examinees whose native language is English; the verbal performance of foreign nonnative-English speakers who say they communicate better in English is lower than that of native-English speakers but is higher than that of nonnative speakers who report that they communicate better in a language other than English.

18. It is useful to recognize that the scoring rationale employed in this study is applicable to reading scores that might be obtained from a separately timed GRE reading comprehension test administered under "rights only" instructions.

19. It is not assumed that scores on RC1 are unaffected by "speed of reading." For example, faster-working examinees who complete the RC section in less than the amount of time allowed have additional time to review their work.

20. The decision to use RCodd as a surrogate for the total score on the 40-item RC section (the "speed of comprehension" score used in the CRCT model), and to use DVodd as a surrogate for the total DV, or "vocabulary" score, was designed to facilitate the evaluation of differences in means and correlation coefficients. By creating tests of approximately the same length, some degree of control is introduced for differences in reliability.

21. Gulliksen (1950) commented on the use of odd-even items as comparable halves of a test, in part, as follows: "It can readily be seen that, if the items are in difficulty order, the odd items will have about the same average difficulty and spread of difficulty as the even items. If there is any bias, it is likely to be that the odd items will be on the average very slightly easier than the even items" (p. 205).

22. These patterns were consistent across subgroups. Coefficients for various subgroups of U.S. examinees are provided in Appendix A.

23. Generally speaking, major-area differences in speed versus level of reading comprehension are expected to vary with degree of verbal-relative-to-quantitative emphasis as follows: humanities, social sciences, biosciences, math-science/physical sciences. In this and other comparisons involving major-area subgroups, it is expected that the specified outcome will be most clearly evident for the two major-area subgroups that are most sharply differentiated with respect to verbal-relative-to-quantitative emphasis--that is, for examinees majoring in the humanities and in the math-science/physical-science disciplines.

24. The statistical significance of differences between the correlated RC1 and RC2 means for the respective subgroups was assessed using a standard formula (e.g., Guilford, 1950: Formula 9.31, p. 216). Correlations, not shown in the table, centered around the values shown in Table 5.

25. In this connection, it is useful to recall both (a) that Ns for non-U.S. subgroups are relatively small and (b) that the verbal subtest standard deviations for non-U.S. examinee subgroups are larger than those for corresponding groups of U.S. examinees (see Table 5).

26. Questions naturally arise regarding the usefulness of undergraduate GPA data, whether actual or self-reported, as a criterion (or as a potential predictor) of academic achievement for foreign students with diverse undergraduate origins. In data for foreign MBA students from 22 U.S. programs (Wilson, 1985b), actual UGPA was uncorrelated with first-year GPA in MBA study (r = .013), while for U.S. students the corresponding coefficient (r = .262) approximated the typical value for such samples. In the present sample of foreign GRE examinees, GRE/SR-UGPA relationships were free of obviously anomalous patterns, hence are reported in this section.

27. These are size-adjusted averages of coefficients computed for subsamples classified by graduate area. It is believed that these coefficients provide a better indication of trends than is provided by evaluation of coefficients for considerably smaller subgroups classified by major area. Generally speaking, the size-adjusted averages corresponded quite closely to coefficients computed in the samples without regard to graduate major area. For example, coefficients for Hispanic examinees generally (that is, without regard to area) as compared to size-adjusted averages of major-area coefficients were: RC1 (.359 versus .361); RC2 (.336 versus .332), DVodd (.304 versus .305); Vform (.381 versus .384).

28. More than 20 percent of the U.S. Hispanic examinees in the sample reported that English was not their better language of communication. Among non-U.S. examinees, as noted earlier, substantial numbers of nonnative-English speakers report that they communicate better in English than in any other language. Such examinees perform less well on English language verbal tests than do their native-English-speaking counterparts. This may be true for U.S. examinees as well. Studies of U.S. Hispanics in the SAT test-taking population (e.g., Pennock-Roman, 1988) have demonstrated interactions between language-learning backgrounds and test performance. Overall, it seems plausible that the common pattern of findings for U.S. Hispanics and for non-U.S. examinees and a different common pattern for U.S. examinee subgroups, except Hispanics or ESL examinees, reflects differential effects associated with differences in speed of processing verbal material in English as the nondominant language on the one hand, and as the dominant language on the other.

29. In the SAT sample studied by Evans (1980), proportionately fewer Blacks than Whites completed the experimental tests under each of several experimentally varied time limits. In current samples of SAT examinees, according to "last-item-attempted" (LIA) criteria, the verbal section of the SAT is more speeded for Blacks than for Whites (e.g., Dorans, Schmitt, & Bleistein, 1988). Generally speaking, lower-scoring examinees without regard to group membership have slower rates of responding as indicated by LIA indices (as shown for GRE examinees in Figures 1a and 1b, herein).

30. Note that among non-U.S. examinees, physical science majors have markedly lower means than humanities majors on both RC2 (speed) and DVodd, but that these subgroups have generally comparable RC1 means, suggesting that under "power-like" conditions they are not very different with respect to ability to read and comprehend general English prose. It seems logical that foreign nationals who are pursuing or planning to pursue graduate work in the humanities will tend to be considerably more highly selected than are their counterparts oriented toward quantitative fields, in terms of developed proficiency in English--including speed of processing general English prose.

31. In an experimental study of component skills in reading, Frederiksen (1980) found that " . . .subjects' use of context in generating and evaluating hypotheses was . . . associated with high reading speed in the comprehension section of the Nelson-Denny test. The picture we gain is that of a proficient reader who constructs a discourse model while reading and utilizes the model to generate hypotheses about likely occurring propositional and syntactic forms that are to follow. The processes of lexical retrieval in such a reader are to a large extent guided by hypotheses derived from context" (p. 136). It is plausible that these "lexical retrieval" processes are expedited by a larger store of familiar words.

32. It seems useful, in this context, to think of examinees in a given major field as "experts" in a particular subject or subject area, and other examinees as being (relative) "novices." If this comparison is tenatively accepted as reasonable, evidence from cognitively based inquiry into "expert/novice" distinctions becomes quite relevant. The following brief commentary from Rigney (1980), for example, is illustrative: "Greater speed and fluency of performance certainly seem to be a general difference between the expert and the novice. . . . A second general distinguishing characteristic of the expert seems to be an enormously richer store of appropriate knowledge in LTM [long-term memory] . . . " (pp. 335-336). Rigney goes on to suggest that the array of differences between the novice and the expert reduce the "amount of uncertainty involved in answering . . . six questions . . .: 'What is it?' 'What should I do about it?' 'How do I do it?' 'Can I do it?' "How am I doing?' and 'Am I through?'" Without stretching the comparison, it seems plausible that, on the average, majors in a given field ("subject-matter experts") will tend to exhibit greater speed and fluency of performance than nonmajors ("comparative novices") when confronted with a GRE reading comprehension set with field-related subject matter. This appears to be recognized implicitly in information supplied to GRE examinees: "Since reading passages are drawn from many different disciplines . . . you should not expect to be familiar with the material in all the passages . . .. You may, however, want to do last a passage that seems to you particularly difficult or unfamiliar" (ETS, 1988, p. 31, emphasis added).

33. There is evidence indicating major-field-related differences in the interpretation of "ambiguated" textual material. Anderson, Reynolds, Schallert, and Goetz (1977: p. 367), for example, reported significant

differences in scores on "disambiguating" measures of various kinds for 30 physical education majors and 30 music education majors after reading passages that could be given alternative inter- pretations (e.g., as either an evening of card playing or a rehearsal session of a woodwind ensemble). The patterns of interpretations reportedly were strongly related to the backgrounds of the subjects involved.

34. In a personal communication, Hale (1988a) noted that on the basis of an informal, internal analysis he was unable to find evidence of significant interaction between passage content of SAT reading comprehension sets and the major-field orientations of SAT examinees. It is, of course, possible that such interactions might be present in data for graduate-level students, who are much more sharply differentiated along disciplinary lines than are high-school seniors.

35. For nonnative-English-speaking examinees, narrative or argumentative passages may make relatively heavier demands on general level of developed proficiency in English than does expository prose involving subject-matter content, regardless of degree of curricular specificity. For example, foreign ESL examinees have been found to perform relatively better on GRE Subject Tests in humanities and social science fields--tests that require extensive processing of discipline-specific verbal material--than on the verbal section of the GRE General Test--whose content is intended to be essentially "curriculum free" (Wilson, 1987). As for possible major-area effects, it seems pertinent that "analysis and evaluation of arguments was judged to be most important (among several reasoning skills, by faculty) in English," in a survey conducted by Powers and Enright (1986: p. 11).

36. The following commentary on differences between the GRE discrete-verbal and reading comprehension sets, from the GRE Technical Manual (Conrad, Trismen, & Miller, 1977) is pertinent: "Discrete questions are notable for their efficiency (contributing high reliability for the amount of time invested), and reading comprehension questions are distinguished by the close link they provide between the test and the actual reading activities of graduate students" (p. 11, emphasis added).

37. For example, GRE Subject Tests (which measure discipline-specific knowledge) tend to be better predictors of graduate school GPA (an alternate measure of discipline-specific accomplishment) than the GRE General Test (e.g., Burton & Turner, 1983). Similarly, the average of scores on two or three College Board Achievement Tests has been found to be a better predictor of freshman-year GPA than general SAT verbal and SAT mathematical scores in samples of first-year students in relatively selective colleges (e.g., Wilson, 1974). Tests that measure subject-matter achievement logically should tend to predict grades (an alternate measure of subject-matter achievement) somewhat better than tests that measure general verbal or quantitative abilities.

38. In evaluating findings regarding differences in criterion-related validity for RC1 and RC2, it is assumed for working purposes that the differences are not plausibly attributable to effects associated with the lack of parallelism in test content discussed earlier. RC2 was a slightly more "difficult" test than was RC1.

39. The Secondary School Admission Test is an academic ability test, developed and administered by ETS under the policy guidance of the Secondary School Admission Test Board, that is used for admissions purposes by selective college preparatory schools. Upper-level (grades 8, 9, and 10) and lower-level (grades 5, 6, and 7) editions are available. It provides measures for reading comprehension, verbal aptitude, and quantitative aptitude.

40. An earlier description (ETS, 1964: p. 6) provides the following additional detail: "The score on the Reading Comprehension Test reflects the amount of prose that can be read and comprehended within a limited period of time. In order to provide adequate measurement for the most rapid readers, the test is constructed in such a way that most students will not complete all the questions" (p. 6, emphasis added). Early editions of the SSAT-RCT provided a score for Level of Comprehension and a Speed of Comprehension score--both were precisely as defined for the Cooperative Reading Comprehension Test. A "Level" score is no longer reported--the current RC score is, in effect, the original "Speed of Comprehension" score. Degree of SSAT RC speededness has not been constant, and today's versions of the SSAT Reading Comprehension Test are somewhat less highly speeded than the version referred to in the 1964 publication cited above (internal ETS documentation). However, the SSAT RC section is currently (ETS, 1987) described as being " . . . comprised of 40 questions based on about seven reading passages that measure the ability to read quickly with understanding . . ." (p. 1, emphasis added). No other ETS-based admission test appears to have been deliberately designed to evoke "speed of response" variance.

41. Evidence that scores on reading comprehension (RC) items tend to be more valid for predicting academic criteria than are scores on "vocabulary" (VO) items (verbal item types that do not involve discourse-level analysis) in the SSAT context (involving ninth-grade students) extends and confirms evidence from validity studies involving (a) samples of GRE examinees (e.g., Wilson, 1985a, 1986b), and (b) undergraduate samples (e.g., Ramist, 1981a, 1981b; Burton, Morgan, Lewis, & Robertson, 1989). All the foregoing were "short-term" validity studies (that is, they employed temporally proximate criteria). In a long-term" validity study, Loyd, Forsyth, & Hoover (1981) found that both college freshman GPA and high school GPA were predicted better by RC scores than by VO scores in a sample of Univerity of Iowa freshmen in which the RC and VO scores had been obtained from one to eight years earlier--from tests administered at grades 4, 6, 8, 9, 10, 11, and 12, respectively.

42. RC2 items were slightly more "difficult" than RC1 items. It is conceivable that this may have had some validity-enhancing effect, apart from "speed," for the native-English speaking examinees.

43. Lord (1974) computed "test-score information curves for hypothetical (SAT) examinees who omit no item . . .," assuming certain specified changes were to be made in an existing test, including scoring only the most difficult half of the test items (that is, the second half of the items). He noted that " . . . discarding the most difficult half of the test items greatly improved measurement efficiency for low-ability examinees [because] random guessing by low-level examinees on the harder items adds so much 'noise' to the measuring process that it would be better simply not to score these items for low-ability examinees. The half test actually measures better than the full-length test at low ability levels" (p. 6). In the case of non-native-English speakers, it is not assumed that low average scores on GRE verbal items reflect "low ability," in the construct-relevant sense of that term. However, if they are not able to reach and consider a representative range of items in the second half of a reading test, the probable effect (in terms of reduced measurement efficiency) appears to be comparable to that described by Lord for "low ability" examinees.

44. From a theoretical perspective, it would be useful to conduct research designed to study the relationship between measures of speed of performing reading-related cognitive processes and RC scores obtained under differentially speeded conditions, ranging from pure power to highly speeded. Based on the work of Hunt and his associates (e.g., Hunt, 1978, 1980, 1987; Hunt, Lunneborg, & Lewis, 1975), Frederiksen (1980), and other cognitive psychologists, it is known that performance on omnibus verbal tests and tests of reading comprehension is related to speed of decoding cognitively undemanding material--measured, for example, by "name identity minus physical identity" (NI-PI) response time. Theoretically, RC/NI-PI correlations should tend to increase with RC speededness.

45. The "first-half versus second-half" model employed in this study appears to have potential value for exploring speed versus level differences primarily for separately timed tests or subtests that are homogeneous with respect to item type. Otherwise, the assumption of parallelism (except for speed) would not be plausible. Moreover, if there are systematic subgroup differences in performance on the item types that are included in a timed test section, systematic interactions between performance on item-type parcels and subgroup membership may be expected (witness, for example, the pervasive major-area related differences in performance on GRE reading comprehension and GRE discrete verbal sets). The fact that the two item-type sets necessarily must be differentially positioned in the section complicates assessment of speed.

46. It is relevant in this general context to note that interest in assessing the properties of GRE reading comprehension sets was prompted in large part by serendipitously discovered evidence of differential

validity for separately reported SAT RC and VO scores that are based, respectively, on reading comprehension and sentence-completion items versus antonym and analogy items (Ramist 1981a, 1981b)--the same item types that are used in the GRE verbal measure.

47. As noted earlier, the only ETS-based admission measure that (implicitly) reflects an assumption that scores with a speed component are more valid for intended purposes than are scores that might be obtained under "pure power" conditions, appears to be the SSAT reading comprehension test. This test is avowedly designed to measure ability to "read quickly with understanding" (ETS, 1987: p. 1). It is noteworthy, however, that no specific rationale for "speeding" the RC measure is provided, and that no empirical evidence appears to have been adduced to support a decision to make this a "speed" rather than a "level" measure (just as none appears to have been adduced to support the [implied] assumption that a speed component in GRE scores is undesirable.

References

Anderson, R. C., Reynolds, R. E., Schallert, D. L, & Goetz, E. T. 1977). Frameworks for comprehending discourse. American Educational Research Journal, 14, 367-381.

Angelis, P. J., Swinton, S. S., and Cowell, W. R. (1979). The performance of non-native speakers of English on TOEFL and verbal aptitude tests (TOEFL Research Report No. 3). Princeton, NJ: Educational Testing Service.

Benton, S. L., & Kiewra, K. A. (1987). The asssessment of cognitive factors in academic abilities. In R. R. Ronning, J. C. Conoley, J. A. Glover, & J. C. Witt (Eds), The influence of cognitive psychology on testing: Buros Nebraska Symposium on Measurement and Testing, Volume 3 (145-189). Hillsdale, NJ: Erlbaum.

Buros, O. K. (1965). (Ed). The Sixth Mental Measurements Yearbook. Highland Park, NJ: Gryphon Press.

Burton, N. W., & Turner, N. J. (1983). Effectiveness of the Graduate Record Examinations for predicting first-year grades: 1981-82 summary report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.

Burton, N. W., Morgan, R., Lewis, C., & Robertson, N. J. (1989). The predictive validity of SAT and TSWE item types for ethnic and gender groups. Princeton, NJ: Educational Testing Service.

Conrad, L., Trismen, D., & Miller, R. (1977). GRE Technical Manual. Princeton, NJ: Educational Testing Service.

Curtis, M. E., & Glaser, R. (1983). Reading theory and the assessment of reading achievement. Journal of Educational Measurement, 20, 133-147.

Donlon, T. F. (1980a). An annotated bibliography of studies of test speededness (GRE Board Research Report GREB No. 76-9R). Princeton, NJ: Educational Testing Service.

Donlon, T. F. (1980b). An exploratory study of the implications of test speededness (GRE Board Professional Report GREB No. 76-9P). Princeton, NJ: Educational Testing Service.

Dorans, N. J., & Lawrence, I. M. (1987). The internal construct validity of the SAT (ETS RR-87-35). Princeton, NJ: Educational Testing Service.

Dorans, N. J., Schmitt, A. P., & Bleistein, C. A. (1988). The standardizaton approach to assessing differential tests speededness (ETS RR-88-31). Princeton, NJ: Educational Testing Service. Author.

Dornic, S. (1980). Information processing and language dominance. Applied Psychology, 29, 120-140.

Educational Testing Service (1960a). Cooperative English Tests, technical report. Princeton, NJ: Author.

Educational Testing Service (1960b). Cooperative English tests, manual for interpreting scores. Princeton, NJ: Author.

Educational Testing Service (1964). Secondary School Admission Test: A guide for interpreting scores on the SSAT. Princeton, NJ: Author.

Educational Testing Service (1969). Secondary School Admission Test: A counselor's guide for interpreting SSAT scores. Princeton, NJ: Author.

Educational Testing Service (1987). Secondary School Admission Test: Interpretive guide for school personnel and educational counselors. Princeton, NJ: Author.

Educational Testing Service (1988). GRE information bulletin, 1988-89. Princeton, NJ: Author.

Evans, F. R. (1980). A study of the relationships among speed and power aptitude test scores, and ethnic identity (College Board RDR 80-81 & ETS RR-80-22). Princeton, NJ: Educational Testing Service.

Frederiksen, N. (1952). The influence of timing and instruction on Cooperative Reading Test scores. Educational and Psychological Measurement, 12, 598-607.

Frederiksen, J. R. (1980). Component skills in reading: Measurement of individual differences through chronometric analysis. In R. E. Snow, P. Federico, & W. E. Montague (Eds). Aptitude, learning, and instruction, volume 1: Cognitive process analyses of aptitude (105-138). Hillsdale, NJ: Erlbaum.

Gates, A. I., & MacGinnitie, W. H. (1969). Gates-MacGinitie Reading Tests: Survey F. New York: Houghton Mifflin.

Glaser, R., Lesgold, A., & Lajoie, S. (1985). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. C. Conoley, J. A. Glover, & J. C. Witt (Eds). The influence of cognitive psychology on testing: Buros Nebraska Symposium on Measurement and Testing, Volume 3 (41-85). Hillsdale, NJ: Erlbaum.

Guilford, J. P. (1950). Fundamental statistics in education and psychology. New York: McGraw-Hill.

Gulliksen, H. (1950). The theory of mental tests. New York: John Wiley.

Hale, G. (1988a). Personal communication.

Hale, G. (1988b). Student major field and text content: Interactive effects on reading comprehension in the Test of English as a Foreign Language, Language Testing, 5, 49-61.

Hunt, E. (1978). Mechanics of verbal ability. Psychological Review, 85, 109-130.

Hunt, E. (1980). The foundations of verbal comprehension. In R. E. Snow, P. Federico, & W. E. Montague (Eds), Aptitude, learning, and instruction, volume 1: Cognitive process analyses of aptitude (87-104). Hillsdale, NJ: Erlbaum.

Hunt, E. (1987). Science, technology, and intelligence. In R. R. Ronning, J. C. Conoley, J. A. Glover, & J. C. Witt (Eds); The influence of cognitive psychology on testing: Buros Nebraska Symposium on Measurement and Testing, volume 3 (11-40). Hillsdale, NJ: Erlbaum.

Hunt, E., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? Cognitive Psychology, 7, 194-227.

Kendall, L. M. (1964). The effects of varying time limits on test validity. Educational and Psychological Measurement, 24, 799-800.

Kingston, N. M., & Dorans, N. J. (1982). The feasibility of using item response theory as a psychometric model for the GRE Aptitude Test (GRE Board Professional Report GREB No. 79-12-P and ETS Research Report 82-12). Princeton, NJ: Educational Testing Service.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 67, 293-323.

Lord, F. M. (1956). A study of speed factors in tests and grades. Psychometrika, 21, 31-49.

Lord, F. M. (1974). Practical methods for redesigning a homogeneous test, also for designing a multilevel test (ETS RB-74-30). Princeton, NJ: Educational Testing Service.

Lord, F. M., & Wild, C. L. (1985). Contribution of verbal item types in the GRE General Test to accuracy of measurment of the verbal scores (GRE Board Professional Report, GREB No. 84-6P & ETS RR-85-29). Princeton, NJ: Author.

Loyd, B. H., Forsyth, R., and Hoover, H. D. (1981). Relationship of elementary and secondary school achievement test scores to later academic success. Educational and Psychological Measurement, 41, 1117-1124.

McClelland, J. L. (1979). On time relations in mental processes: An examination of systems of processes in cascade. Psychological Review, 86(4), 287-330.

Pennock-Roman, M. (1988). The status of research on the Scholastic Aptitude Test (SAT) and Hispanic students in postsecondary education (ETS RR-88-36). Princeton, NJ: Educational Testing Service.

Pitcher, B. (1962). SSAT scores as predictors of ninth grade averages: 1959-60, 1960-61 (SR 62-20). Princeton, NJ: Educational Testing Service.

Powers, D. E., & Enright, M. K. (1986). Analytical reasoning skills involved in graduate study (GRE Board Professional Report GREB No. 83-23P & ETS Research Report 86-43). Princeton, NJ: Educational Testing Service.

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

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

Ramist, L. (1981a, February 12). Validity of the SAT verbal subscores (Internal memorandum, Educational Testing Service, Princeton, NJ).

Ramist, L. (1981b, July 7). Further investigation of the validity of SAT verbal scores (Internal memorandum, Educational Testing Service, Princeton, NJ).

Rigney, J. W. (1980). Cognitive learning strategies and dualities in information processing. In R. E. Snow, P. Federico, and W. E. Montague (Eds), Aptitude, learning, and instruction, volume I: Cognitive process analyses of aptitude (pp. 315-343): Hillsdale, NJ: Erlbaum.

Rindler, S. E. (1979). Pitfalls in assessing test speededness. Journal of Educational Measurement, 16, 261-270.

Rock, D. A., Werts, C., and Grandy, J. (1982). Construct validity of the GRE Aptitude Test--An empirical confirmatory study (GRE 78-1P & ETS RR-81-57). Princeton, NJ: Educational Testing Service.

Ronning, R. R., Conoley, J. C., Glover, J. A., & Witt, J. C., Eds. (1987). The influence of cognitive psychology on testing: Buros Nebraska Symposium on Measurement and Testing, volume 3. Hillsdale, NJ: Erlbaum.

Sinnot, L. T. (1980). Differences in item performance across groups ETS RR-80- 19). Princeton, NJ: Educational Testing Service.

Snow, R. E., Federico, P., & Montague, W. E., Eds. (1980). Aptitude, learning, and instruction, volume 1: Cognitive process analyses of aptitude. Hillsdale, NJ: Erlbaum.

Swinton, S. S., & Powers, D. E. (1980). A factor analytic study of the restructured GRE Aptitude Test (GRE Board Professional Report 77-6P). Princeton, NJ: Educational Testing Service.

Wild, C. L., & Durso, R. (1979). Effects of increased test taking time on test scores by ethnic group, age, and sex (GREB No. 76-6R). Princeton, NJ: Educational Testing Service.

Wild, C. L., McPeek, W. M., & Koffler, S. L., with Braun, H. I., & Cowell, W. (1988). Concurrent validity of verbal item types for ethnic and gender subgroups (Final Report GRE No. 84-10). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1974). The contribution of measures of aptitude and achievement in predicting college grades (ETS RB 74-36). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1985a). The relationship of GRE General Test item-type part scores to undergraduate grades (GRE Board Professional Report GREB No. 81-22P & ETS Research Report 84-23). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1985b). Factors affecting GMAT predictive validity for foreign MBA students: An exploratory study (ETS RR-85-17). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1986a). The GRE Subject Test performance of U.S. and non-U.S. examinees, 1982-84: A comparative analysis (GRE Board Professional Report No. 83-20P11 & ETS Research Report 87-04). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1986b). The relationship of scores based on GRE General Test item-type part scores to undergraduate grades: An exploratory study for selected subgroups (GRE Board Professional Report 83-19P & ETS Research Report 84-38). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1986c). The relationship of GRE General Test scores to first-year grades for foreign graduate students: Report of a cooperative study (GRE Board Professional Report GREB No. 81-11P & ETS RR-86-44). Princeton, NJ: Educational Testing Service.

Appendix

Tabular Summaries of Correlations of Selected GRE Verbal Subtests (RC1, RC2, REven, ROdd) with a GRE Discrete-Verbal Subtest and Total GRE Verbal Score, for U.S. Examinee Subgroups[a]

[a]See Table 4 and related discussion in the text for perspective.

Table A.1

Correlations of RC1, RC2, RCodd, and RCeven, with Vform (Total Verbal Formula Score), by Graduate Major Area, For Designated Demographic Subgroups

		Correlation of variable with Vform					
	N	RC1	RC2	(Dif)	RCodd	RCeven	(Dif)
Hum	3,692	.797	.840	.043	.826	.841	.015
Male	1,398	.823	.853	.033	.848	.855	.007
Female	2,294	.780	.831	.051	.813	.830	.017
White	3,383	.779	.824	.045	.811	.824	.013
Black	147	.801	.832	.031	.815	.855	.040
Hsp	61	.874	.911	.037	.898	.905	.007
Asian	27	.647	.893	.246	.674	.845	.171
Soc Sci	10,258	.820	.859	.039	.847	.869	.022
Male	4,599	.823	.853	.033	.844	.866	.022
Female	5,659	.821	.862	.041	.852	.867	.015
White	8,963	.796	.841	.045	.827	.849	.022
Black	711	.810	.841	.031	.846	.846	.000
Hsp	263	.838	.852	.014	.850	.881	.031
Asian	98	.794	.865	.071	.854	.841	-.013
Bio	4,429	.786	.836	.050	.815	.845	.030
Male	2,103	.781	.827	.046	.804	.839	.035
Female	2,326	.790	.843	.053	.809	.832	.026
White	4,064	.771	.823	.052	.801	.830	.029
Black	151	.821	.892	.071	.874	.907	.033
Hsp	64	.744	.822	.078	.792	.820	.028
Asian	73	.824	.874	.050	.888	.879	-.009
Phys Sci	2,718	.801	.849	.048	.843	.842	-.001
Male	2,121	.803	.843	.043	.845	.838	-.007
Female	597	.796	.867	.071	.838	.856	.018
White	2,480	.782	.832	.050	.826	.827	.001
Black	71	.867	.887	.020	.916	.863	-.053
Hsp	46	.842	.906	.064	.847	.855	.008
Asian	72	.861	.894	.033	.796	.799	.003

Note: These are part-whole coefficients, reflecting the relationship between designated reading subscores and the total verbal formula score. Each of the subscores is based on 20 reading items: RC1 = 1-20, RC2 = 21-40; RCodd = 1,3, . . ., 39; RCeven = 2,4, . . ., 40. Data are for U.S. examinees only.

Table A.2

Correlations of RC1, RC2, RCodd, and RCeven, with DVodd (Score on Odd-Numbered Discrete-Verbal Items), for Designated Demographic Subgroups, by Graduate Area

Major area/ Subgroup	N RC1	Correlation of variable with DVodd RC2 (a)	(Dif) (b)	RCodd (b-a)	RCev (c)	(Dif) (d)	(d-c)
Hum	3,692	.627	.674	.057	.659	.667	.008
Male	1,398	.676	.698	.022	.700	.697	-.003
Female	2,294	.596	.657	.061	.632	.646	.014
White	3,383	.602	.650	.048	.638	.640	.012
Black	147	.616	.643	.027	.610	.676	.066
Hsp	61	.737	.806	.069	.765	.795	.030
Asian	27	.395	.746	.351	.460	.679	.219
Soc Sci	10,258	.642	.688	.046	.670	.689	.019
Male	4,599	.642	.680	.038	.664	.687	.023
Female	5,659	.642	.693	.051	.675	.690	.015
White	8,963	.604	.653	.049	.634	.654	.010
Black	711	.596	.653	.057	.632	.647	.015
Hsp	263	.661	.674	.013	.672	.696	.024
Asian	98	.631	.699	.068	.686	.675	-.011
Bio	4,429	.568	.618	.050	.595	.619	.024
Male	2,103	.556	.607	.051	.584	.606	.022
Female	2,326	.577	.628	.051	.603	.633	.030
White	4,064	.542	.594	.052	.570	.594	.024
Black	151	.636	.712	.076	.678	.723	.044
Hsp	64	.484	.555	.071	.522	.547	.025
Asian	73	.658	.695	.043	.748	.667	-.081
Phys Sci	2,718	.605	.655	.050	.651	.638	-.013
Male	2,121	.602	.643	.041	.647	.628	-.019
Female	597	.597	.616	.019	.663	.673	.010
White	2,480	.573	.624	.051	.619	.608	-.011
Black	71	.728	.768	.040	.783	.735	-.048
Hsp	46	.664	.751	.087	.698	.686	-.012
Asian	72	.749	.796	.050	.796	.799	.033

Note: The coefficients in this table are between designated reading subscores and scores on DVodd (a verbal subscore based on odd-numbered discrete-verbal items). Each of the reading subscores is based on 20 items: RC1 = 1-20; RC2 = 21-40; RCodd = 1,3, . . ., 39; RCeven = 2,4, . . . , 40). Data are for U.S. examinees only.

Table A.3

Correlations of RC1, RC2, RCodd, and RCeven,
with DVodd (Score on Odd-Numbered Discrete-Verbal Items),
for Designated Demographic Subgroups, by Graduate Area

Major area/ Subgroup	N	Correlation of variable with DVodd RC1 (a)	RC2 (b)	(Dif) (b-a)	RCodd (c)	RCev (d)	(Dif) (d-c)
Major area							
Hum	3,692	.627	.674	.044	.659	.667	.008
Soc	10,258	.642	.688	.046	.670	.689	.019
Bio	4,429	.568	.618	.050	.595	.619	.024
Phys	2,718	.605	.655	.050	.651	.638	-.013
Male							
Hum	1,398	.676	.698	.022	.700	.697	-.003
Soc	4,599	.642	.680	.038	.664	.687	.023
Bioi	2,103	.556	.607	.051	.584	.606	.022
Phys	2,121	.602	.643	.041	.647	.628	-.019
Female							
Hum	2,294	.596	.657	.061	.632	.646	.014
Soc	5,659	.642	.693	.051	.675	.690	.015
Bio	2,326	.577	.628	.051	.603	.633	
.030Phys	597	.597	.616	.019	.663	.673	.010
White							
Hum	3,383	.602	.650	.048	.638	.640	.012
Soc	8,963	.604	.653	.049	.634	.654	.010
Bio	4,064	.542	.594	.052	.570	.594	.024
Phys	2,480	.573	.624	.051	.619	.608	-.019
Black							
Hum	147	.616	.643	.027	.610	.676	.066
Soc	711	.596	.653	.057	.632	.647	.015
Bio	151	.636	.712	.076	.678	.723	.044
Phys	71	.728	.768	.040	.783	.735	-.048
Hispanic							
Hum	61	.737	.806	.069	.765	.795	.030
Soc	263	.661	.674	.013	.672	.696	.024
Bio	64	.484	.555	.071	.522	.547	.025
Phys	46	.664	.751	.087	.698	.686	-.012
Asian American							
Hum	27	.395	.746	.351	.460	.679	.219
Soc	98	.631	.699	.068	.686	.675	-.011
Bio	73	.658	.695	.043	.748	.667	-.081
Phys	72	.749	.796	.050	.796	.799	.033

Note: Data for U.S. examinees only.