DOCUMENT RESUME

ED 248 827 HE 017 889

AUTHOR	Adelman, Clifford
TITLE	The Standardized Test Scores of College Graduates, 1964-1982.
INSTITUTION	National Inst. of Education (ED), Washington, DC.
PUB DATE	Dec 84
NOTE	101p.; Prepared for the Study Group on the Conditions of Excellence in American Higher Education. For related document, see ED 246 833.
PUB TYPE	Statistical Data (110) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Academic Achievement; *College Entrance Examinations; *College Graduates; Graduate Study; Higher Education; Majors (Students); Professional Education; *Scores; *Standardized Tests; Student Characteristics; *Test Results; Trend Analysis
IDENTIFIERS	*Excellence In Education; Graduate Management Admission Test; Graduate Record Examinations; Law School Admission Test; Medical College Admission Test

ABSTRACT

Scores from 23 standardized tests that are used in application to graduate and professional schools are analyzed, primarily from the 1964-1982 period. The 23 examinations include tests of advanced achievement in 15 subject areas, along with tests of general learned abilities (the Graduate Record Examination/Verbal and Quantitative, the Law School Admissions Test, the Graduate Management Admissions Test, and Medical College Admission Test Reading and Quantitative Analysis subtests. Major conclusions include: (1) the quality of available data on test scores and on the background characteristics of test-takers is highly variable; (2) changes in test scores over a period should be measured in terms of standard deviation units, and not in points or percentages; (3) of 23 examinations, performance declined on 15, remained stable on 4, and advanced on 4--the greatest declines occurred in subjects requiring high verbal skills; (4) none of the basic demographic characteristics of the test-takers (age, race, gender, citizenship, or native language), in themselves, explain the observed changes in performance over the period; and (5) different undergraduate majors provide convincing explanations of observed changes in performance. Issues concerning the measurement of scaled test scores and the magnitude of change are addressed. Data on test performance are appended. (Author/SW)

The Standardized Test Scores of College Graduates, 1964-1982

Clifford Adelman
Senior Associate
National Institute of Education

Prepared for
The Study Group on the Conditions of Excellence in
American Higher Education

December, 1984

The opinions, observations and conclusions offered in this paper are those of the author as an individual. They should <u>not</u> be taken as representing the opinions or positions of the National Institute of Education, the U.S. Department of Education, the National Commission on Excellence in Education or any of the Commissioners, or the Study Group on the Conditions of Excellence in American Higher Education or any of its members.

TABLE OF CONTENTS

Acknowledgements

Abstract

Text:

1. Background and Purposes of this Paper	1
2. General Approach	3
3. Limitations for Policy Makers and Trend Watchers	4
4. Sources of Data	5
5. Inconsistencies in Reporting Data	7
6. Size of the Test-Taking Population	8
7. Are Test Scores Reflective or Predictive?	9
8. How Should We Measure Change in Scaled Test Scores?	12
9. How Should We Judge the Magnitude of Change?	16
10. Changes in Performance, 1964-1982	16
11. Three Periods of Change, 1964-1982	18
12. Other Internal and Administrative Influences	21
13. Explaining the Changes, 1: Numbers of Test-Takers	22
14. Explaining the Changes, 2: Age, Race, Gender	24
15. Explaining the Changes, 3: Citizenship, Native Language	26
16. Explaining the Changes, 4: Undergraduate Major	30
17. General Conclusions	35
18. A Message to the Testing Services: Gathering Consistent Data	36
19. A Message to the Commentators: a Plea Against Excuses	40

Notes 43

Bibliography 46

Tables and Graphs Appendices

ACKNOWLEDGEMENTS

The author is deeply grateful for the extended commentaries on earlier drafts of this paper by William Turnbull (former President and Distinguished Scholar in Residence at the Educational Testing Service) and Jonathan Warren (former Senior Research Scientist at ETS/Berkeley), and for assistance in data collection and explanation of test development, reporting and allied procedures to Leonard Ramist, Eldon Park, Cheryl Wild, Nancy Burton, Jerilee Grandy, Linda Cole, and Lawrence Hecht of ETS and Rennee Kennish and Janice Wilson Plante of the Law School Admissions Service. For comments and suggestions on the final draft, I am grateful to Alice Irby, William Turnbull, Nancy Burton and Eldon Park of ETS, Zelda Gamson and Alexander Astin of the Study Group on the Conditions of Excellence in American Higher Education, and my NIE colleagues John Wirt and Charles Stalford. I also wish to thank Dr. Manuel J. Justiz, Director of the National Institute of Education, for providing agency funds to analyze the LSAT data and for his interest and support for this and other NIE initiatives in the field of higher education.

ABSTRACT

This is a secondary analysis of available and/or published data on the performance of college graduates (and soon-to-be graduates) on 23 standardized tests used in the process of admission to graduate and professional schools. For most of these tests, the analysis covers the years, 1964-1982. The 23 examinations include tests of general learned abilities (the Graduate Record Examination/Verbal and Quantitative, the Law School Admissions Test, the Graduate Management Admissions Test, and sub-tests in Reading and Quantitative Analysis of the Medical College Admissions Test) and tests of advanced achievement in 15 specific subject areas.

The approach taken is purposefully heuristic--an attempt to illustrate both the virtues and limitations of common-sense empiricism. Its major conclusions are:

1) The quality of available data on test scores and the background characteristics of test-takers on these examinations is highly variable. The data have been inconsistently gathered and reported over the years.

2) Some 550,000 U.S. citizens currently take these examinations every year. While the group is self-selected, it represents a significant sample of the potential pool; and we are justified in looking at its performance as reflective of the changing quality of <u>student learning</u> in U.S. colleges and universities. At the same time, we would <u>not</u> be justified in using these test scores as the primary indicator of the quality of American higher education.

3) Changes in test scores over a period such as this should be measured in terms of Standard Deviation Units, not points or percentages.

4) Of 23 examinations, performance declined on 15 (principally GRE Subject Area tests), remained stable on 4 and advanced on 4. The greatest declines occurred in subjects requiring high verbal skills.

5) There were three distinct historical periods of change: one of sharp decline (1964-1970), one of basing or reversal of trend (1970-1976) and one of more modest decline (1976-1982), These periods can be explained, in part, by internal content adjustments of the tests and to changes in methods of test administration and scoring.

6) Most of the relationships between numbers of test-takers and trends in scores on individual examinations over the period in question are counter-intuitive. We cannot explain the observed changes in performance with reference to gross numbers.

7) None of the basic demographic variables--age, race, or gender of the test-takers --in and of itself, can explain the observed changes in performance over the period.

8) Neither citizenship nor fluency in English, in themselves, can explain the observed changes in performance. Only in combination with undergraduate major do these variables begin to offer plausible hypotheses of influence on test score trends.

9) The performance and participation of U.S. students from different undergraduate majors appear to offer the most convincing explanation of observed changes. Students with undergraduate majors in professional and occupational fields (the most rapidly growing group among both degree grantees and test-takers) <u>under</u>perform all others.

The paper concludes with a plea to the testing services to gather and report consistent data to help educators monitor the quality of undergraduate education, and provides some suggestions toward that end. The paper also urges that no one be excessively defensive about test results. The more excuses we make, the less likely our colleges and universities will focus their attention on the bottom line of undergraduate education--what students learn.

Standardized Test Scores of College Graduates
1964-1982
Some Notes and Interpretations

Clifford Adelman, Senior Associate, Program on Educational Policy and Organization, the National Institute of Education

1. Background and Purposes of this Paper

A number of the recent national reports on the status of American education have commented on the widely-watched indices of declining test scores of secondary school graduates. But only the National Commission on Excellence in Education (A Nation at Risk) and the Study Group on the Conditions of Excellence in American Higher Education (Involvement in Learning) mentioned the declining trends in the scores of college graduates on standardized tests. As a staff member to both groups, the author was responsible for gathering the background data on test scores, and has commented elsewhere (1) on the implications of the trends in scores for postsecondary standards and institutional assessment policy. But in the course of writing those commentaries, I developed a sense of little ease: it was obvious that the publicly accessible data on which I drew my conclusions were not very good, and the conclusions themselves were thus more polemical than they should have been. I resolved my frustrations with the data by saying, in effect, "well, if that's the way they report it, then this is the only conclusion one can reach."

This paper is intended, in part, both to revise and expand upon those earlier analyses. It is based on a great deal more data on both test scores and the background characteristics of test-takers, as well as discussions of potential interpretations of this data with experts in the testing community. The ultimate intention of this paper is to plea for better data on both reported standardized test scores and the characteristics of test-takers, and for data that can be understood by a broad audience, as well as by small groups of technicians and clients with narrow interests. More importantly, it is to urge the higher education community to use assessment data on college graduates to improve our understanding of the influence of college curricula on student learning.

Given that residue of polemical purpose, it is also important, at the outset, to state what this paper is not:

- It is not a commentary on the generic uses of tests and test scores or particular uses of the tests under discussion;
- It is not an analysis of the quality, reliability, validity or any other technical property of the tests under discussion;
- It is not an inquiry into the question of whether these particular tests and the methods of assessment they assume are the best means of measuring college student learning.

The reader may find that some of these issues are implicit in the discussion; but if this initial warning sign is flashed, I am confident that this paper will not be misused.

Even in the best of times, college educators do not like to employ scores on standardized examinations to reflect the quality of higher education. Since the objectives of baccalaureate education include far more than learning whatever it is that is measured by multiple-choice examinations, and since life itself places a higher premium on careful insight than on the speed of response that governs American testing, their objection is understandable. The objection is understandable, too, in light of the fact that those who take examinations such as the Graduate Record Examinations (General and Subject Area), the Law School Admissions Test, the Graduate Management Admissions Test, and the Medical College Admissions Test tend to have higher educational aspirations than their peers and are probably the more able students. The tests results, some might say, do not reflect the achievement of the average or below average college student, hence should not be used to judge the quality of American higher education.

While granting the general validity of these objections, I would point out, first, that there is a distinction between measuring the quality of student learning and assessing the quality of higher education--the former being but one of a number of tasks that one would employ in the latter. One should note, too, that the graduate and professional programs of our universities use these tests in much the same way as do the undergraduate admissions offices--sometimes with an even more actuarial bias.* As Lauren and Daniel Resnick have pointed out, "graduate schools, like employers, do not treat college diplomas as equivalent, although it is still considered somewhat impolite to talk very openly about the differences in standards among colleges" (2)--for which reason, in part, the tests are used in the graduate school admissions process. At the least, they are a recognized common currency, something that we cannot say about either credits or grades. Somehow, though, we allow colleges and universities to criticize secondary education on the basis of admissions test score trends, and decline to comment on trends in graduate school admissions test scores.

This paper attempts to organize and provide interpretive frameworks for publicly accessible measures of college student achievement---on 18 years (1964-1982) of scores on such examinations as the Graduate Records, the Graduate Management Admissions Test, the Law School Admissions Test, and the Medical College Admissions Test (though I place less emphasis on the MCATs because the entire test battery was changed in 1977). All of these examinations are taken predominantly by college graduates and soon-to-be graduates.

The information that forms the core of the analysis is presented in six tables appended to this text:

---

* For a comprehensive review and discussion of the uses of all these examinations in the graduate and professional school admissions process, see Rodney Skager, "On the Use and Importance of Tests of Ability in Admission to Postsecondary Education," in Alexandra Wigdor and Wendell R. Garner (eds.), Ability Testing: Uses, Consequences and Controversies. Washington, D.C.: National Academy Press, 1982, Part II, pp. 286-314.

Table A presents the basic data on mean scores, Standard Deviations, and number of test-takers for the Verbal and Quantitative sections of the Graduate Record exams, for 15 Graduate Record Subject Area Tests, for the Law School Admissions Test, for the Graduate Management Admissions Test, and for four of the sub-tests of the Medical College Admissions Test.

Table B presents selected available data on the background characteristics of the GRE/General, LSAT, MCAT and GMAT test-takers in recent years, and focuses on key variables such as age, citizenship, native language, sex, race, and post-college experience.

Table C indicates the preferred method of determining trends in scaled scores on all the tests at issue, and provides information concerning the scales themselves, and other data that bear upon our interpretation.

Table D follows up on Table C by honing in on the changes in test scores as determined by the preferred method.

Table E identifies turning points in the trends of test scores on the Graduate Record examinations (Verbal, Quantitative, and 15 Subject Area Tests) during the 1970s.

Table F illustrates the comparative performance of students, by college major, on the Verbal and Quantitative Sections of the Graduate Record Examination, the Law School Admissions Test and the Graduate Management Admissions Test, during the period 1977-1982.

Other tables illustrate very specific issues raised in this paper, for example, distinctions between the performance of U.S. citizens and foreign students on the GRE/General and GMATs, or changes in numbers of test-takers on the GRE Subject Area examinations following changes in the method of test administration.

2. <u>General Approach</u>

My approach to analyzing this data is partially naive. That is, I do <u>not</u> want to take the role of the psychometrician (test expert)--which I am not--rather, that of the informed citizen, the individual who tries to bring a common sense empiricism to bear on what otherwise appears as a chaotic collection of data. I want to look at some numbers and try to reason through them in a way that would reflect knowledge of at least a modicum of the basic literature on testing. In following this heuristic journey through reported data, the reader should remember that the testing services have five constituencies:

1) College admissions officers and graduate school admissions committees (the principal constituency which, as noted below, determines the ways in which scores are reported);

2) Students who take the various examinations (and their families);

3) Policy-makers in schools and school districts, colleges, legislatures and executive offices, etc. who use test scores in symbolic

ways to guide analyses of existing conditions in education and proposed changes in educational policy;

4) The general public (and the media), which also uses test scores in very raw and symbolic ways so as to create the environment for change or stability, an environment that influences the ways in which principals and school boards, college deans and presidents, state legislators and others will subsequently act;

5) Test developers and administrators, researchers and other scholars.

In these comments, I am basically adopting the perspective of #3, and, in light of recent national assessments of the state of education, am naturally interested in historical trends.

3. Limitations for Policy Makers and Trend Watchers

But the way in which the testing services report scores and data on test-takers is not intended for any of the above constituencies except the first.

That is, the primary clients of the testing services are admissions officers or committees or Graduate Deans who want to compare applicants to their institutions or programs in any one year. These clients are not interested in historical trends (nor are the students who take the tests in any given year). Their interests are wholly understandable; and the data are generated for them. In fact, they own the testing programs and control testing policies.

How does the client affect the way in which the data are reported? One example from the college admissions level may suffice: a reader of an annual series of mean scores of College Board achievement tests will notice an occasional blip--up or down 20 points, let us say--in what otherwise seems to be a fairly stable series of numbers. The explanation provided by ETS is that a particular achievement test was re-scaled to the SAT in that particular year. Rescaling is a perfectly legitimate statistical procedure that removes the differences in verbal and mathematical ability among populations taking particular achievement tests. It "re-sets" the mean score on an achievement test to the mean SAT scores of those who took that achievement test.

The college admissions officer is not interested in what that rescaling does to the historical trend of scores: he or she simply wishes an accurate way of comparing applicants who present different CEEB achievement test scores. While there may be other reasons for rescaling, the fact is that it occurs with some regularity on achievement tests, at least on the college entrance level. As indicated by internal ETS studies, the cumulative effects of rescalings are rather significant, though both the magnitude and direction of the effects differ by test. (3)

The point of the example, though, is that because the data are not designed for us, they are hard for us to interpret. Thus, too, College Board achievement test means are not comparable over any extended period of time, so we cannot use them to reach conclusions about trends in the achievement of high school graduates in specific subject fields.

On the other hand, with the exceptions of the MCATs, none of the tests taken by applicants to graduate and professional schools was re-scaled after its introduction (or in some cases, revision) until 1982-3. A rescaling of the GRE Subject Area Tests was proposed and considered by the GRE Board in 1969, but was rejected on the grounds that across-field comparisons were not as necessary or helpful as historical patterns within fields. So, it is wholly appropriate to look at historical trends in these scores, aided by the data that have been collected since 1975-6 on the background characteristics of the test-takers. In a sentence, this stability is one of the principal statistical justifications for undertaking this study.

4. Sources of Data

This is a secondary analysis of existing data, and the data used were drawn from a variety of sources. It is best to describe the sources first, and then to comment on the adequacy and quality of the data provided in the course of our analysis.

Graduate Record Examinations: General and Subject Area

The basic data on mean scores, standard deviations and candidate volume were drawn from an unpublished chart provided by the Educational Testing Service, covering the years 1963 to 1983. More detailed data on the background characteristics of test takers and cross-tabulations of scores according to different background variables were drawn from the published series, A Summary of Data Collected from Graduate Record Examinations Test Takers During 19--. (hereafter referred to as GRE Summary). This series has been published every year since 1975-1976. The data contained in this annual publication apply only to the GRE "General" Examinations (Verbal, Quantitative, and Analytic). If there is comparable public data for the Subject Area tests, I am unaware of it (though it is obvious from the Guide to the Use of the Graduate Record Examinations that at least some of this data exists). The reader should note that I do not include the GRE/Analytic test scores in this analysis because the examination was introduced recently (1977) and (more importantly) has undergone one major reconstitution since (in 1981). The amount of comparable performance data on the GRE/Analytic is hence too limited for our purposes.

Graduate Management Admissions Test

The basic data on mean scores and candidate volume for the years 1965-1982 were drawn from an unpublished chart provided by the Educational Testing Service. Standard Deviations for the years

1965-1977 were obtained by telephone conversation with ETS staff. Standard deviations, other test score data, background characteristics of test-takers and cross-tabulations of scores by different background variables were drawn from an unpublished set of tables prepared each year since 1977-1978 under the copyright of the Graduate Management Admissions Council, and used here with its permission. The data for 1980-1981 were obtained from a published version of these tables, A Demographic Profile of Candidates Taking the Graduate Management Admission Test During 1980-1981 (Princeton, N.J.: Graduate Management Admissions Council, 1982).

Medical College Admission Test

All the data on mean scores, sub-test scores, standard deviations, candidate volume, background characteristics of test-takers and cross-tabulations of scores by different background characteristics were drawn from an annual series of reports prepared by the Division of Educational Measurement and Research of the Association of American Medical Colleges, Medical College Admission Test: Percentile Rank Ranges for MCAT Areas of Assessment [and] Summary of Score Distributions. This particular information on the background characteristics of MCAT test-takers is less detailed than that on the other examinations. I focus principally on the annual reports since 1977, as that was the year in which the current version of the MCATs was introduced.

Law School Admissions Test

The basic data on mean scores, standard deviations and candidate volume were drawn from an unpublished chart provided by ETS to the National Commission on Excellence in Education in 1983, and subsequently by the Law School Admissions Services (which took over the administration and records of the LSATs from ETS in 1979). In addition, and with the permission of the Law School Admissions Council, LSAS prepared a separate set of data analyses especially for this paper, covering characteristics of test takers and cross-tabulations of scores by different background variables for the years 1975-1983.

On the basis of the LSAS data analysis, it was obvious to me that the ETS data excluded some test-takers who should have been included in the aggregate numbers and included other data that should have been excluded.* Thus, the LSAT data used in this paper are broken into two periods, by source: 1964-1974 (ETS) and 1975-1982 (LSAS).

---

* We were able to control what was in the LSAT data from 1975-1982 by defining the universe as those who actually took and completed the examination (without subsequent cancellation of scores) at all regular administrations (including those who, for religious reasons, use special Monday administrations). For those individuals who took the LSAT more than once in a given testing year, only the most recent score (and accompanying background data) is included in our universe. This definition precluded some of the irregularities that appear to be in the pre-1975 data.

5. Inconsistencies in Reporting Data

As the LSAT example indicates, there is a considerable lack of consistency in reporting both test scores and data concerning the test-taking population. A quantitative historian would spend years untangling the mess in order to obtain comparable data over time.

So as not to be misinterpreted, however, I should stress that individual scores reported to students and graduate or professional schools are accurate. Those with a personal interest (students or admissions committees) should not be excessively concerned if the aggregate data are inconsistently reported to those with academic or policy interests. But we should be concerned, nonetheless, because policy decisions are often made on the basis of our perceptions of the aggregate data.

For an example of inconsistency, the reader will note that on Table B (data concerning the background characteristics of GRE, GMAT, MCAT, and LSAT test-takers) the figures for some variables are reported for all test-takers, while for others, the figures are those for first-time test-takers only, and for others, still, for first-time test takers who are U.S. citizens. One can reasonably assume that there will be a difference in the aggregate characteristics (and test scores) of all test-takers versus first-time test-takers, let alone other universes used in the reporting. The more complex charts used in the GRE Summary bear that assumption out.

For example, again, in the annual GRE Summary there was no simple figure indicating the percentage of women test-takers until 1978-1979!!! And as for the percentage of women aspiring to doctoral or post-doctoral study, the universes for reporting changed considerably from 1975-1979 to 1980-present.

While it is difficult to require the administrators of different examinations to ask background questions in the same way, one is dismayed that in the simple matter of identifying a student's undergraduate major field, no two of the major examinations are comparable. As the reader will note later in this discussion, that inconsistency hampers our insight concerning the movement of students from one field to another.

Yet another complication (beyond the control of the testing services) is that not all the test-takers will answer all the questions on background information questionnaires that accompany each test. On some questions --depending on where they are and how they are phrased--the number of non-respondents is significant. As ETS researcher Jerilee Grandy notes, "examinees who complete the background questionnaire . . . tend to have somewhat higher test scores," a fact that is particularly noticeable when one analyzes test scores by undergraduate major. (4) As long as we understand this phenomenon, readers should not be frustrated or puzzled when they find that the subject-universe for one variable is different from the subject universe for another, or where my tables disagree as to the total number of test-takers in a given year or the mean scores and Standard Deviations for a given test.

6. Size of the Test-Taking Population

Despite the inconsistencies, there is another very simple statistical justification for engaging in this inquiry: a lot of people take these tests. In the most recent year for which I am including data on all of the tests (1981-1982), the total number of test-takers (minus those whom we know for sure are not graduates of U.S. colleges) was as follows:

	Total	Minus Non-U.S. Citizens*	% Receiving Degree in 1981 or 1982
Grad. Record General Exams:	256,381	17,291	54%
Grad. Management Admiss. Test:	203,304	38,807	42%
Law School Admissions Test:	99,928	?????	58%
Medical College Admiss. Test:	47,597	?????	N.A.
TOTAL:	607,210	- 56,098	= 551,112

(In addition, 80,149 scores were recorded on 19 GRE Subject Area tests. For purposes of this calculation, I assume that all those who took the subject area tests also took the GRE General Exams and should not be counted twice).

Is that number (about 550,000) substantial? The question is less simple-minded than it appears. While there may be some overlap (though it is impossible to determine how many people take more than one of these tests), this number should be set against a "potential pool of examinees" determined as follows: (5)

(a) Since half of the test-takers in 1981-82 received their bachelor's degrees in either 1981 or 1982, we start with a base of bachelor's degrees awarded in those years: 1,880,000

(b) Since we are trying to determine a participation rate for U.S. citizens, we should remove from (a) all non-resident aliens who received bachelor's degrees in 1981 and 1982. (45,000)

(c) Since, as table B shows, some 23% of the GRE examinees and 14% of the LSAT examinees were already enrolled in graduate or professional school, we should add a percentage of graduate and first-professional enrollments minus enrollments of non-resident aliens at those levels (1,386,000 - 91,000). I have taken 20% as the product of aggregation based on the two examinations. Hence, 20% of 1,295,000: 259,000

Thus, one very rough measure of the potential pool of these test-takers is 2,094,000. When 551,000 take the tests, that means 26% of the pool (v. 65% of the pool of secondary school students who take either the SATs or the ACTs).

---

*The GRE data distinguish between non-resident and resident aliens. I use the non-resident figure. The GMAT, on the other hand, does not distinguish between resident and non-resident aliens. And neither the LSAT nor the MCAT report separate data for non-U.S. citizens.

Most of the adjustments that one might make to this very rough estimate would have the effect of lowering the number in the pool, hence raising the percentage of examinees. For example, William Turnbull writes that my estimate for graduate enrollments probably overstates the pool "since most of the enrolled graduate students who take the GRE are likely to be in their first year." (6) While I do not have the data to prove the case, Turnbull's hunch is probably right, and might bring the proportion of examinees up to 28% of the potential pool.

A small number of graduate departments do not require or recommend the GRE/General examinations, but do require the Subject Area Tests. Thus, were we to modify our original assumption and estimate that 10-15% of the Subject Area examinees did not take the GRE Generals, we would raise the number of test-takers, and add perhaps a half percentage point to the proportion of examinees relative to the potential pool.

Yet another adjustment would be based on the notion of a "likely pool." In other words, instead of all baccalaureate degree recipients, we might eliminate those who received degrees in fields for which the baccalaureate has historically been the terminal degree. But that would involve some arbitrary decisions and would load the dice. I would rather be conservative and, on the basis of historical data (at least as far back as 1971)*, posit a range of 25-30% of the pool.

Let us not complexify this matter: 550,000 is a big number and what the statisticians call a "robust sample." It is a self-selected sample, and also one that is partly driven by the entrance requirements of professional schools and research universities and by the application requirements of fellowship sponsors. Virtually all accredited medical, law, and business schools require their respective examinations, and, as Table J indicates, with very few exceptions, 65% or more of the Ph.D.-granting graduate departments in the major disciplinary fields either require or recommend the GRE Subject Area tests (furthermore, these percentages have not changed significantly since 1971). So in addition to a robust number, there is a surface stability to this population--at least in terms of presumed ability. Whatever quibbles we might have, this is the only substantial annual sample we possess from which to make some inferences concerning the quality of undergraduate learning in the United States in recent years.

7. Are Test Scores Predictive or Reflective?

Some may argue that the tests in question here are all predictive rather than reflective. That is, they may say that most questions on an examination are designed to determine how well a student will perform at the next higher level of education, not how well a student has learned the material taught at his/her current level of education.

---

* In 1971-72, for example, there were 546,145 test-takers, of which (though one must estimate from trends here) 518,500 were U.S. citizens. Using the same formula to determine the potential pool, we come up with 1,961,000. The ratio of test-takers to the potential pool was thus 26.4%--almost exactly the same as it was a decade later.

From one point of view, this distinction is very subtle; from another, it is bogus. What I mean may be best illustrated by Jonathan Warren's comment that the GRE Subject Area tests are "too generalized" to be reflective. "Generalized," as Warren admits, is not a good term. Instead, one might say that in any discipline, there is a common denominator or "core" of subject matter that a graduate admissions committee expects a student to have mastered, and it is that common core that is reflected on the GRE Subject Area test. But some college departments pride themselves in specialties that lie outside that common core. Their students will thus not perform as well on the examination as others. Warren thus suggests that our existing tests are unsuitable for program evaluation or assessment, and are not really reflective--even though (as I would add) the "core" is evidently sufficient for predictive purposes. (7)

But are the examinations unsuitable? The Committee of Examiners for each GRE Subject Area Test recommends test specifications that weight different sub-fields within the overall set of questions. It then draws questions "from the courses of study most commonly offered." (8) For Political Science, for example, the specifications for the 170 questions used on current (1982-1984) versions of the GRE test are as follows:

"I. U.S. Government — 30-35%
Questions will cover the major subfields of United States government, including institutions, processes of national and subnational politics and public administration.

II. Comparative Political Systems — 20-25%
Some questions will be area or country specific; others will be concerned with comparative institutions and administrative processes. Questions will deal with developed and developing states.

III. International Relations — 15-20%
Questions will cover international politics, theory, organizations, law and political economy as well as United States foreign policy.

IV. Political Theory and History of Political Thought — 20-25%
Questions will deal with normative and empirical, conceptual and analytic matters, as well as the ideas of major political thinkers.

V. Methodology — 10%
Questions will deal mainly with methods and techniques involved in empirical research. Many of these questions will also involve the subfields described above." (9)

The comparative weights of these subfield sections may not match the political science program at every college, but that is not the intention. Nor will they likely match the weights of the programs taken by individual students. To be sure, a student who took 12 semester courses in political science, of which 8 were in comparative political systems and international relations, will be at a disadvantage on this examination. The test does not cater to undergraduate sub-field specialists.

What it reflects instead is an ideal balance of the undergraduate curricula offered. If the Committee of Examiners--all of whom are political scientists appointed with the advice of the American Political Science Association--represents the field well, then we have to appeal to their authority. And if they establish a universe of questions "from the courses of study most commonly offered" in undergraduate political science programs, then it stands to reason that the test is reflective.

A more technical term for what is at issue here is "content representativeness," i.e. the degree to which the examinations "draw upon content that is basic to and most important for success in graduate school." (Oltman, 1982) When a study of the "content representativeness" of a Graduate Record Subject Area test is undertaken, faculty are asked to judge an existing version of the test against: (a) what is actually taught in their own departments, and (b) an "ideal" undergraduate curriculum in a discipline. The results of such a study on the Chemistry, Computer Science and Education tests (see Table G) demonstrate that the weighting of the tests ("Committee Specifications") was somewhat different from the judgments of teaching faculty. Whether that should be the case, whether the faculty should be wiser than the Committee, the very question indicates a reflective purpose in the tests. As W. Ann Reynolds, a member of the Committee of Examiners for the GRE Biology test in the early and mid-1970s reflected, "it was always very clear in our minds that we were writing questions to ascertain how well students had mastered various aspects of the biological sciences." (10)

Perhaps the whole issue is a circumlocution of logic; and I would not dwell on it if it had not been raised by others. Of course we expect graduate students to have mastered appropriate subject matter and habits of mind necessary for studying a discipline or profession at an advanced level. Studies of the predictive validity of these tests tell us how well the measures work, not how well the general test-taking population has learned in college. Such studies look only at those test-takers who actually enter and complete at least one year of graduate or professional school; and our understanding of the reflective qualities of the examinations is limited if we look only at that group. As Wigdor and Garner observe:

> "When the accepted group is a select subset of the applicant group, the correlation with the criterion is lower that it would be for all applicants. In extreme cases the reasons for a lower correlation in a selected group are easy to see. In a basketball league that was limited to people with heights of 5'10" and 5'11", height would not be expected to be a good predictor of a player's average number of rebounds per game. If player heights vary greatly, however, the correlation between height and average number of rebounds would be expected to be higher." (11)

Even so, as Anne Anastasi has observed, all these tests measure the "current status" of student learning and development, regardless of "whether their purpose is terminal assessment or prediction," (12) and "current status" is influenced by a student's past course-taking patterns. As Nancy Burton of ETS points out, a score on the GRE Subject

Area Test in Economics may be "affected more by how recently the student took mathematics than by the Economics the student studied. Therefore," she concludes, "the test might be excellent for selection, but poor as a measure of undergraduate Economics curricula." (13) This is a viable hypothesis about curricular co-requisites, one which could be investigated through an analysis of the college transcripts of test-takers (see p. 38 below); but it still does not change the conclusion, as the Economics curriculum has become more quantitative in orientation due to changes in the professional practice of Economics. The Committee of Examiners itself acknowledges that evolution in its description of the GRE Economics test (14), and one suspects that the vast majority of undergraduate Economics programs now require at least Economic Statistics if not other mathematics and quantitative Economics courses. So we would expect the examination to measure the "current status" of student learning in those programs.

Reinforced by such common-sensical distinctions, the lay public will reason that the GRE Subject Area Tests are the only nationally validated measures to assess the "current status" of undergraduate learning in the disciplines (and their co-requisites) and that the LSAT, GRE/General Examinations (Verbal and Quantitative), GMAT and portions of the MCAT examinations are the only nationally validated measures of student competence in analysis, problem-solving, and verbal facility at the college level. Therefore, when over a half-million U.S. citizens take these tests annually, we are justified in looking in changes in scaled test scores as reflective of the quality of college student learning. Whether these particular tests are the best of all possible measures or whether the quality of test performance predicts performance in graduate school, professional school, or subsequent career is beside the point in the symbolic environment of public interpretation. (15)

8. How Should We Measure Change in Scaled Test Scores?

For reasons of easy public reference, we would like to be able to make statements such as "The mean SAT scores have declined by X per cent over the past 10 years," or that "The mean LSAT score has risen by X per cent over the past 10 years."

The Point Approach

But the testing services never make such statements. Instead, they say that the SAT has declined by so many points or the LSAT has risen by so many points. Let us adopt a piece of the terminology from testing to discuss this issue and ask whether these statements--phrased in terms of "points"--have "face validity." That is, is what they are saying obvious?

If you observed that the temperature fell 10 degrees last night, the meaning of your statement is very different depending on the scale you are using. Fahrenheit? Celcius? Reaumur? For a farmer or a chemist (in the case of the Reaumur scale), the difference is critical.

All the tests we are talking about involve raw scores which are then statistically translated onto a scale. If all the tests had the same

number of questions and the same scales, a statement such as "the mean GRE/Verbal score declined 61 points between 1964 and 1982 while the mean score on the GRE Area Test in Sociology declined by 113 points" might provide some reliable clues about comparative academic performance on these two measures over the same period of time.

But the scales for each one of these tests are different--as are the number of questions (and the fewer the number of questions, the more volatile the scores). Most of this data is presented for all 24 tests and sub-tests under examination in Table C. An excerpt may help to underscore the point:

Test	(1) Maximum Reported Score	(2) Minimum Theoret. and Reported Score	(3) Number of Points on Scale (#1 - #2)
GRE/Verbal	850*	210	640
Chemistry	990	440	550
Physics	990	370	620
Economics	990	400	590
Sociology	990	210	780
English Lit.	810	250	560

*Was 900 from 1952-1980. Reduced to 800 in 1981.

It is obvious that each scale is different, and that a 61 point decline on the GRE/V scale (210 to 850--or a 640 point scale) would probably not be as severe as a 61 point decline on the GRE Chemistry test scale (440-990--or a 550 point scale).

Given the difference in scales, a statement about changes in mean scores that is phrased in terms of points does not have face validity.

Arithmetic Approaches

The attempt to make statements about percentage declines or increases in mean scores of tests (and I have made my share of those statements over the past few years) involves simple, arithmetic approaches. Let us use the GRE/Verbal scores for 1964 and 1982 (the extremes of the period with which this paper deals) to illustrate these approaches:

	1964-5	1981-2	
Mean	530	469	-61 points
Stand. Deviation	124	130	

The decline in the mean score between these two historical benchmarks was 61 points (we will deal with the Standard Deviation in a moment). What kind of percentage decline is that?

(1) Is the percentage decline 61/530--or 11.5%??? That might make sense if the scale started at zero. But the scale does not start at zero: it starts at 210 (the average minimum score on seven forms of the GRE used between 1973-1977).

(2) If we say that 210 = 0, then is the percentage change 61/(530-210)--or minus 19 1%?? That would appear to make sense if the scale was 1000 points in value, i.e. 210 to 1210. But the scale for the GRE/Verbal is 210 to 850. Instead of 1000 point in value, it is 640 points in value.

(3) If the scale for the GRE/Verbal is 640 points in value, each point is worth 1/640, or .00156%. So, does the percentage decline for the GRE/V = 61 x .00156%--or 9.5%??? That sounds fairly sensible, and perhaps is the most sensible of the otherwise misleading arithmetic methods.

These three arithmetic methods, each referring only to changes in mean scores, yield three very different results: (11.5%), (19.1%) and (9.5%). Most of us would stare at those figures and conclude that there has to be something wrong with the method.

There is: we have not described what is really being measured because we have not identified the basic point of reference accurately.

The Geometric Approaches

If we were dealing with raw scores on these tests, the arithmetic approaches might make more sense. But we are looking at arbitrary scales, scales with historical baggage*, and benchmarks (scores) along those scales that are being "equated" (smoothed out by statistical adjustments) from year to year as the number of questions on a given test changes, as the distribution of those questions across fields and tasks is reweighted, and as the level of difficulty of those questions and the mix of difficulty levels shifts. The scale is sacred to the testing community; but these scales are not universal measures like Fahrenheit degrees. When one changes the scale itself (as the LSAT did in 1982 and as the MCAT did in 1977 to accompany the change in the entire character and construction of that examination), we lose one history and must begin another one.

The measurement of change on a scale demands attention to some very basic statistical constructs, the most basic of which is the Standard Deviation. The Standard Deviation is a measure of variance--or geometric dispersion--of scores that accounts for the abilities of students who take an examination. We ignore it in this regard at our peril--and we ignore it to the detriment of students. Ideally, the Standard Deviation should tell us the range of scores around the mean within which roughly 2/3rds of the cases on a given test in a given year, fall. That may not be a technically eloquent way of phrasing it--but it will have to do.

---

* The GREs, for example, were standardized in 1952.

The Standard Deviation thus provides us with interpretive guidelines for the scores of a majority of students. When the change in scores over time <u>exceeds</u> the range of expectations inherent in the Standard Deviation, then--depending on the direction of change--we ought to be worried ught to be pleased. Again, that is not a technically accurat ng of the issue, but it's going to have to do.

Changes in mean scores over time, then, can be measured as a fraction of the Standard Deviation in the base year. When these changes approach or exceed one full Standard Deviation (1.00), then we are observing <u>very</u> significant change. That is a theoretical benchmark which, as we will shortly see, does not always fit the circumstances.

There are two ways of calculating this fraction:

(1) If we use the Standard Deviation for the base year (1964) only, then we get a computation for the GRE/Verbal that looks like this:

$$\frac{-61 \text{ (change in scale points)}}{124 \text{ (Standard Deviation in 1964)}} = -.49$$

Translation: over an 18 period, the change in the mean score of the GRE/V was roughly one-half a Standard Deviation. The direction of change was obviously negative.

(2) There is a slight problem with Geometric Method #1. It is really quite minor, but, in the interests of technical accuracy, ought to be addressed. The Standard Deviation takes account of real groups of students who take the tests and their abilities. Obviously, a different group of students takes the tests each year. When we compare scores from 1964 to those of 1982--or to any year in between--we are <u>not</u> looking at the same group of students.

So it is a technical mistake to base our calculations on the Standard Deviation in the base year. Certainly, over an 18 year period, we would distort or misrepresent the composition of the student groups taking these tests by so doing.

What do we do instead? There is really no way to render the groups of test-takers over 18 years equivalent to each other so that our comparisons of performance can be scientifically accurate. In the absence of statistical guidelines, let us try something that at least sounds common sensical: let us use the <u>mean</u> Standard Deviation for the 18 years in question. Actually using the mean S.D. (as opposed to the base year S.D.) does not change the results that much.

For the GRE/Verbal, the computation would thus be:

$$\frac{-61 \text{ (change in scaled points)}}{128 \text{ (mean S.D., 1964-1982)}} = -.48$$

For better or for worse, the last of our methods (Geometric #2) will be used in this paper, except when we are looking at fewer than 10 years of data on a particular test (and then we will use Geometric #1).

9. How Do We Judge the Magnitude of Change?

The notion that a change of 1.00 in Standard Deviation Units (S.D.U.) is a touchstone of significance is an arbitrary one. One has to develop a sense of degrees in the context of specific data and make some judgment calls. In Investment in Learning, Howard Bowen provides some guidelines for describing "estimated average changes in cognitive learning resulting from college education" when one measures in terms of Standard Deviation Units (16), and I borrow from him here.

Bowen applies his descriptors to change in the learning of the same group of students over a period of four years and uses a total of six gradations. Our figures cover 7-18 years of the performance of different students. As our variations will thus be greater, we need more descriptive categories, and I have chosen the following:

Estimated Change as Expressed in S.D.U.s	Descriptive Judgement
+.75 or above	Extreme increase
+.40 - .74	Large increase
+.20 - .39	Moderate increase
+.10 - .19	Small increase
(.09)- +.09	NO CHANGE
(.10)- (.19)	Small decline
(.20)- (.39)	Moderate decline
(.40)- (.74)	Large decline
(.75) or below	Extreme decline

10. Changes in Performance, 1964-1982

Let us now turn to the bottom line as presented in Tables C and D. There are 24 tests and sub-tests at issue (including the division of the LSAT data into two periods). Of the 24, we have 15-18 years of data on 18, and 6-8 years of data on 6 (all of which are either sub-tests of the MCAT or the two periods of the LSAT). Let us make our task easier for a moment, look only at the long-term measures, and combine the two periods of the LSAT. Then, let us distinguish between tests with large numbers of test takers and those with comparatively small numbers of examinees. The former consist of the tests of "general learned abilities," the GRE/V, the GRE/Q, the GMAT and the LSAT. The latter consist of 15 GRE Subject Area tests. In very gross terms, here's what happened:

	High Volume Tests (GRE,GMAT,LSAT)	Low Volume Tests (GRE Subj. Area)
Advances	1	2
No Change	1	2
Declines	2	11

Now, if that were a summary of stock market action for a day, a week, or a year, one would be hard-pressed to call it bullish.

Remember that these are trends in separate measures, and that most of the students who take the GRE Subject Area tests (low volume) also take the GRE/V & Q (high volume). We are not counting the same students twice, rather are distinguishing between two types of performance, general and subject-specific, and are using the trends as indicators. While the general performances may be more important simply by weight of numbers, no matter which type of performance we examine, the 18-year trend is the same--down. At the same time, as we will see, shorter term measures (the 6-year trend) evidence a more neutral pattern.

Staying with the surface data for a moment, we should ask whether the changes in scores are significant. Note (in Table D) that six of the 15 test scores (longer term or short term) that declined evidence "large" or "extreme" declines, but that none of the four test scores that increased did so by a large amount. So our intuitive perception of significant decline is probably justified.

We know that not everyone who takes the tests attends graduate or professional schools; and we know that not everyone who attends graduate and professional schools has taken the tests. But if the half-million people who take these tests come from the top 25-30% of their college classes, and if the general decline is as severe as it appears, then perhaps we ought to start looking more closely at the quality of what is taught and learned in American higher education (at least in the fields in which the declines are most significant).

No doubt an objection will be raised: these declines, it will be argued-- even among the top 25-30% of our college graduates-- simply reflect declines in the preparation of of entering college students, and what we really should do is to control the results for student ability at the point of college entrance, or, at the least, compare these trends to those on the SAT and College Board Achievement tests.

These objections should be answered in reverse order: in a secondary analysis of aggregate data, one cannot compare trends in the test scores of college graduates to those in the test scores of high school graduates. As I pointed out previously, the College Board Achievement tests are not good historical guides because the scores are periodically rescaled. As for the SAT and the ACT as compared to the GRE, LSAT, GMAT, and MCATs, we have a classic case of apples and pears. These are different examinations being taken by very different populations.

But the SATs and GREs are the most analogous of these fruits, and invite comparison as indicators. Performance on the SAT/Verbal fell by -.41 of a Standard Deviation Unit during the period 1960-1978 (the period under investigation minus four years--to account for the traditional gap between high school and college graduation), while performance on the GRE fell -.48 of an S.D.U. Likewise, performance on the SAT/Math fell -.23 during that period while performance on the GRE/Q did not change at all. But as soon as we remove foreign students from the GRE equation, the results change (see p. 27 below), and the difference between SAT and GRE indicators narrows considerably, to wit:

Change in Performance for U.S. Citizens
(in Standard Deviation Units)

	SAT (1960-1978)	GRE (1964-1982)
Verbal	-.41	-.38
Quantitative	-.23	-.15

Is the GRE population here analogous to the SAT population? No. It is not as homogeneous a group in terms of age, and most will concede that it is a group of higher relative ability.

It is more important to control for ability in primary statistical analyses of GRE test scores by reference to earlier scores on the SATs, but the research literature is surprisingly thin on this issue. (17) But it is equally important not to transform statistical controls for ability into an "input approach" to measuring student achievement, for the "input approach" is of no help in determining what influences students between matriculation and graduation. To determine student learning during the college years--at least on the more general of these measures--we might give the GREs, LSATs, and GMATs to college freshmen, and then again, to college seniors (on the other hand, one could not prove anything by following that procedure with the GRE Subject Area Tests). Or, as Turnbull suggests, we can aggregate existing information on student performance to make a similar assessment. (18)

11. Three Periods of Change, 1964-1982

Even if all we looked at were mean scores, Standard Deviations and numbers of test-takers, the case of decline is not as simple as indicated in Tables C and D. One of the striking characteristics of this data is that there appear to be three distinct periods of change within the 18 year span under investigation:

Period I: 1964-1970. This is a period of sharp declines in scores--at least for the 17 examinations on which we have full data.

Period II: 1970-1976. This is a period of reversal and stabilization, i.e. one during which either declining trends reversed direction or "based out," reaching a low plateau.

Period III: 1976-1982. This is a period of stability and/or modest decline. After "basing out" or reversing trend, the mean scores either held steady or resumed their general direction of change at less dramatic rates than in Period I.

Table E is intended to illustrate this phenomenon for the GRE examinations (Verbal, Quantitative, and Subject Area tests), and should be used in conjunction with Table A. Table E points to the specific years in which change in direction of the test scores occurred, using the criterion of ± .10 or more change in Standard Deviation Units from base years (and, subsequently, from "turning point years"). In examining Table E, the reader will note that not all the tests demonstrate trends corresponding to the three periods. The GRE/Verbal and Quantitative examinations, for example, evidence only two periods, i.e. one "turning

point," as do the GRE Subject Area tests in Engineering, History, Psychology and French. But even in those cases, the turning point occurs sometime during Period II, 1970-1976.

Was there anything that happened to the tests themselves, the methods of administration of tests, or the scoring of the tests during the 1970-1976 period that might account for these changes in trends?

Indeed, the answer to all three questions is "yes." The Graduate Record Examination Board was established in 1966, and, under its direction, a number of alterations in both the tests and methods of administration began to appear. ETS carried out a complete review of the GRE Subject Area tests in 1970-1972 that resulted in changes in the content and weighting of the different parts of the examinations. A number of content validity studies were first conducted; then the Committee of Examiners for each test recommended changes designed to reflect the way in which different subjects were actually being taught in our colleges.

Biology, a discipline that underwent a major revolution in the 1960s, is a classic example. The figures in Table A demonstrate that the mean score of the GRE/Biology Area test declined modestly from 1964 through 1969-71, when it "bottomed out," then reversed trend, rising until 1975-76. In 1971-72, the test was reweighted internally to reflect, e.g. the growing emphasis on cellular and molecular Biology in most college curricula; and consequently, the norms had to be recalculated. The following year (1972-73), the first year of the new weightings, the mean score jumped 13 points (a change in Standard Deviation Units of +.11--which, for a one year change, appears rather significant), while the Standard Deviation itself fell from 115 to 110. Both changes reflect a greater agreement between the preparation of undergraduate Biology majors and the construction of the examination.

Do the effects of these reweightings and other internal content adjustments influence our overall judgment on the general decline in test scores of college graduates over the period in question? To a certain extent, yes; though it is impossible to quantify the judgment and though the judgment applies principally to the Subject Area tests of the GREs. One might argue that if the content and weighting of the examinations changed in the early 1970s to reflect better what was actually being taught in our colleges and universities, then the earlier declines (Period I) may have been exaggerated. Nonetheless, after the period of trend reversal, most of the scores resumed their declines--though, as we have noted, at a more modest rate. And one should add that the trends in mean scores of the general learned ability tests (GRE/Verbal and Quantitative, LSAT, and GMAT) were wholly unaffected.

The second possible explanation of trend reversal in Period II has to do with methods of test administration and score reporting. In 1969, local administrations of the GREs (General and Subject Area) were eliminated, i.e. if you wanted to take the test you had to do it at a national administration where your presence and scores would be included in the nationally reported data. Or, if your college wanted all Psychology majors to take the GRE Subject Area test for purposes of program evaluation, those students would have to troop to a test center on a given administration day. (19) The coincidence between the period of these

adjustments and the stabilizing or reversal of trends in GRE mean scores is rather striking, and is backed up by changes in the Standard Deviations. An increase in the numbers of test-takers reported in the national data is also noticeable during this period, in part as a result of changes in rules vis-a-vis the use of the tests in local assessments.

How much did this shift in method of test administration and score reporting affect the trends we observe? Based on the data available, I cannot tell; but let us speculate. For convenience, let us call those who voluntarily took the GREs (General or Subject Area) as part of institutional evaluations the "experimental group." The GRE annual reports (GRE Summary) do not split this group out from other test-takers, as the questionnaire does not ask why an individual is taking a particular test.

If you are a volunteer test-taker, your motivation to perform well is less than that of someone who needs the best possible score for graduate school admission. If you are being paid, the compensation is for your time, not your performance. So we can reasonably hypothesize that the experimental group will not perform as well on the examination as the "control group" (i.e. the rest of the test-takers). At the same time, though, volunteers for such a task tend to be among the better students. The net effect would probably be a wash.

Even so, how large would this "experimental group" be, and how much would its performance affect mean scores and Standard Deviations? Once again, we have to make some speculative inferences. Between the 1968-1969 testing year and the 1969-1970 testing year, there was a surge across the board in the number of test-takers on the GRE Subject Area examinations, a surge that is rather noticeable when set against both preceding and subsequent years (see Table I). Increases of more than 20% in the number of test-takers in 1969-70 can be observed in History, Biology, Geology, Engineering, Economics, Sociology, Psychology, Education, and Music.

While many factors were responsible for the overall increase in the number of test-takers in the late 1960s and early 1970s, the relative increase in this particular year was probably due to the change in the administration and score reporting of tests taken for purposes of local program assessment.

The impact on scores, however, is more difficult to judge. According to our hypothesis that volunteer test-takers (even though they tend to be the better students) do not perform as well as those who are in the game for competitive purposes, the mean scores should go down and the Standard Deviations should rise with the infusion of any significant number of this group into the overall population of test-takers. On the surface, that appears to be what happened. Of 15 Subject Area tests, 13 mean scores declined between 1968-69 and 1969-1970, and 10 Standard Deviations rose (though three of those rose by a minimal amount). But the changes in mean scores do not differ from those of the previous years--at least in general direction. Thus, one is not wholly sure of the impact of the "experimental group" on score trends.

12. Other Internal and Administrative Influences

Changes in Numbers of Questions. The fewer the number of questions on an examination, the greater the volatility of scores. Table C indicates the approximate number of questions on the different examinations. The reader will note that the GRE Subject Area tests in Mathematics and Physics (which emphasize problem-solving), have fewer numbers of questions than other GRE Subject Area tests, and are the only two GRE tests to evidence increases in mean scores over the 18 year period. The number of questions should not make a difference with the respect to the direction of change, but the relationship is certainly noticeable here. The relationship can only be explained with reference to the quality of students who major in mathematics and physics: they tend to out-perform most of their peers on tests of general learned ability (GRE/V&Q, LSAT, GMAT) as well (see Table F-13).

We should also note that the number of questions on all examinations other than the GRE Subject Area Tests may vary slightly from year to year, and that not all of those questions count in determining raw and scaled scores. Many questions are inserted on a trial basis. From the results, ETS determines the level of difficulty, content validity and other statistical properties of the questions, hence their appropriateness for inclusion in the universe of questions from which the different versions of the test can draw. This variance in number of questions does not affect test results.

Scoring Methods. Nor should a recent change in the method of determining raw scores. Up to 1980, the raw score was determined by adding the number of questions answered correctly and subtracting one fraction of the incorrect answers and another fraction of questions the test-taker did not answer at all. Starting in 1981 for some tests (GRE General) and 1982 for others (e.g. the LSAT), the method was changed so that the students was not directly penalized for questions he/she did not answer. And one can infer from Grandy's comments that the GRE/General Examinations now use a "rights only" method of scoring that does not even penalize the student for guessing wrong answers. (20) The method of determining raw scores should not change scaled scores, since there will be appropriate statistical adjustments to insure comparability and continuity. However, perhaps five years from now, one could subject the data to rigorous statistical analysis to determine whether that hypothesis holds.

Disclosure. Lastly, one might ask whether the recent "truth-in-testing" practices have affected these test scores. Over the period 1979-1981, the governing boards for the GRE, LSAT and GMAT made both the tests and answer sheets available to students following each test administration. Some might assume that the more students who take advantage of disclosure, the better prepared for subsequent examinations the test-taking population will be, and therefore mean scores should rise while Standard Deviations should remain steady or fall.

Given the recency of this development, we have very little information on which to judge the effects of disclosure on the quality, results, and uses of the tests. In fact, practice tests for the GRE Subject Area

examinations were not available until 1982-83, and therefore fall outside this analysis.

The only examination on which test disclosure practice might have made a difference is the LSAT, partly because, as Wigdor and Garner observe, "LSAT scores receive more weight in decisions on laws school admissions than do scores on tests for other programs" (21). Knowing that is the case, we would reason that more prospective law students would seek to take advantage of disclosure as soon as it was possible, and the results would be reflected in the data. Indeed, according to Bruce Zimmer, then Executive Director of the Law School Admissions Council, 55% of "recent" (1981) examinees took advantage of the disclosure option. (22) And circumstantial evidence does suggest a relationship between test performance and disclosure: the mean score on the LSAT jumped 14 points and +.13 of a Standard Deviation Unit in two years (1979/80 to 1981/82). But until we conduct some research, we are never going to know for sure whether this otherwise tantalizing hypothesis can be validated.

13. Explaining the Changes, Round 1: Numbers of Test-Takers

Our judgment on the significance of the overall changes in scores depends on the numbers and characteristics of test-takers, and here there are some confusing trends.

Now, the numbers of test-takers have changed rather dramatically in the period under consideration (1965-1982), and these changes are often attributed to external forces such as the Vietnam War and shifts in the labor market. Regardless of external cause (that analysis lies beyond the scope of this paper, and could not be conducted on the basis of available data gathered from examinees), the conventional wisdom says that there is an inverse relationship between the numbers of test takers and mean scores. That is:

> The greater the number of test takers, the lower the scores;
> The lesser the number of test takers, the higher the scores.

That is supposed to be the way it goes--in very, very bald and simple terms. Of course I have purposefully simplified the relationships. For the ratios to work, many other variables must remain constant, e.g. scales of tests, weightings of sub-test raw scores, etc. Nonetheless, I'd like to keep the analysis very simple for the moment.

Is that the way it actually went from 1964-1982? Let's look at some of the scores and numbers here to see what happened.

A. Graduate Record General Exams: Verbal and Quantitative

- o The total number of test takers has more than doubled since 1965, but has been falling steadily since 1976.
- o The mean score on the GRE-Verbal fell steadily through the whole period, as if it were impervious to the number of test takers.
- o On the other hand, the mean score on the GRE-Q fell with the rising number of test takers through 1976, then rose with the falling number of test takers through 1982--just as conventional wisdom holds.

B. Graduate Management Admissions Test

- The total number of test takers has quadrupled since 1965 (It may have peaked in 1980-1981, but it's too early to tell).
- The mean score followed the paradigm illustrated in Table E, dropping rapidly from 1965-1971, then levelling off, even as the numbers of test takers continued to rise. By the strict constructionalist reading of conventional wisdom, though, the scores should fall with rising numbers.

C. Law School Admissions Test. Here we have two different sets of data, from which we can nonetheless conclude that:

- The total number of test takers tripled between 1965 and 1974, but has declined about 20% since then.
- No matter which set of data one uses, the mean scores on the LSATs rose through the whole period, as if they were impervious to the number of test takers.

*D. Graduate Record Achievement Tests: English and History

- The number of test takers rose dramatically during the period 1964-1970, but has fallen precipitously since then.
- Mean scores fell with rising numbers (just as conventional wisdom would have it), but continued to fall along with the number of test takers, thus running against conventional wisdom.

*E. Graduate Record Achievement Test: Biology

- The number of test-takers rose considerably from 1965 to 1978, and has since reversed field, dropping 32% to 1982.
- Scores first fell, then rose with increasing numbers (particularly from 1970-76), then fell again with decreasing numbers. Only the first of these three movements proceeds according to conventional wisdom.

F. Graduate Record Achievement Test: Mathematics

- The number of test-takers rose from 1965-1970, but has been falling ever since.
- Mean scores fell with the rising numbers and rose with the falling numbers--just as they are supposed to do.

*G. Graduate Record Achievement Test: Sociology

- The number of test-takers rose 150% from 1964-1968, rose another 66% in the next two years, then fell all the way back to its 1964 level by 1982.
- Scores remained stable from 1964-1968, then began falling--first (and briefly) with rising numbers of examinees, and then with falling numbers. Only the second of these three movements proceeds according to conventional wisdom.

In the nine cases I have used for illustrative purposes, there are many instances in which the relationship between score trends and number of examinees does not follow conventional wisdom. The frequency of these

---

*See Table M

"counter-intuitive" situations is such as to dampen one's enthusiasm for an analysis of trends based on the single variable of volume. But there is another reason not to pursue this line of analysis: at the graduate school level, it is not accurate. Conventional wisdom may work when one is talking about moving from one general level of education to another, such as high school to college--as in the case whenever a product or service becomes both attractive and accessible to a previously marginal population of consumers. But, as Turnbull writes, "where you are considering shifts among fields of students at a particular level--entry into one segment rather than another--I would not expect the same logic to hold and would look for volume and scores to move up and down together at least as often as they moved in opposite directions [emphasis added]." (23)

That is, even when we are looking at a population of 550,000, we must realize that the students are selecting not only to continue their education at a very high level, but also to explore a very particular territory. What applies at a more generalized level does not necessarily apply here. Consequently, I would like to set this issue aside for a while, and under the conviction that one cannot explain the changes we are witnessing only by reference to the number of test-takers.

14. Explaining the Changes, Round 2: Age, Race, Gender

The demographic characteristics of the test-taking population are usually regarded as fruitful sources in explaining trends in performance on standardized tests. Let us look at three of the most basic of these characteristics--age, race, and gender--to see whether they can help refine our understanding of what has happened. The basic references are Tables A and B, and we are looking at changes only during the years for which we have information on the demographic background of the test-takers--1975-1982--and only at the tests of general learned ability (GRE, LSAT, GMAT, and the Reading and Quantitative Sub-Tests of the MCAT since 1977).

Let us remind ourselves, first, what happened on each one of those tests during that period:

	Pt. Change	S.D.U. Change
GRE: Verbal	492 to 469 = (23)	(.18)
GMAT:Verbal	25.9 to 26.5 = 0.6	.06
MCAT:Reading	7.98 to 7.74 = (.24)	(.10)
LSAT	530 to 553 = 23	.21
GRE: Quant	510 to 533 = 23	.17
GMAT:Quant	27.0 to 27.4 = 0.4	.05
MCAT:Quant.	7.99 to 7.45 = (.54)	(.21)

There are obviously some distinctly different trends here. Do the basic demographic variables explain them?

Age. The college population is growing older and it is hence not surprising that the population of those who take these tests is growing older. As evident in Table B, the percentage of those in the traditional 19-24 age group fell significantly on both the GREs and

LSATs, while remaining stable for the GMATs. The MCATs do not report data by age, but we can reason that since the percentage of college graduates taking the MCATs has remained stable during a period in which the average age of college graduates rose (see Table B), the MCAT test-taking population is also a bit older.

Should older students perform better on the GRE-Q and worse on the GRE-V, for example? I certainly would not expect so. After all, mathematics is a wholly school-learned subject, and the further away in time one gets from the active study of mathematics, the less likely one would perform well on a test of general quantitative ability. At the same time, verbal facility should increase with age--and the reading, writing, work and social experience (hence, language use) that comes with age.

The MCAT:Reading sub-test trend reinforces that of the GRE-V; but the MCAT:Quantitative trend moves in precisely the opposite direction from the GRE-Q. The explanation is fairly simple. The GRE-Q draws largely on secondary school mathematics; but the MCAT: Quantitative also draws on college-level mathematics, and is more of an achievement test. Pre-Medical students, one can reasonably speculate, pay more attention to their preparation in scientific subjects than they do to mathematics, and the trend in scores on the MCAT sub-tests in Biology and Chemistry supports that speculation (besides, some 77% of the MCAT test-takers major in either the Biological or Physical sciences).

Age does not seem to be an issue with respect to the test-taking population for the GMATs. But what do we say about the LSAT in this regard? The LSAT is an examination that relies heavily on verbal skills, and the trend in LSAT scores moved precisely in the opposite direction from that for the GRE-V during the period under consideration. Age alone cannot account for that difference.

Thus, the first demographic variable in which there has been a significant change over the seven year period in question is not much help in explaining the differences in test score trends.

Racial/Ethnic Characteristics. The racial/ethnic distribution of U.S. citizen test-takers changed but fractionally during the 1975-1982 period, and not enough to affect mean scores. The percentage of whites taking the GREs, for example, declined from 87.4% in 1975 to 86.1% in 1982--hardly a staggering move. During the same time, the percentage of Asian-American test takers increased from 1.3% to 1.9%. While Asian-Americans tend to score higher on the GRE-Q than other groups, that percentage increase would not account for a +.17 change in Standard Deviation Units across 256,000 test-takers.

As for the performance of other minority groups on these tests, there is no single trend. For example, looking at the performance of Afro-Americans and Mexican-Americans compared with that of all U.S. citizens on the LSATs and GMATs, we find:

	Change in Standard Deviation Units		
	Afro-Americans	Mexican-Americans	All U.S. Citizens
LSAT, 1975/76 to 1981/82	+.11	+.09*	+.21
GMAT, 1977/78 to 1981/82	+.16	+.10	+.09

*Approximate, since reporting categories for Hispanics changed in 1979/80.

One could continue with similar examples of variability--greater or lesser change for specific minority groups on different measures of general learned abilities. Whatever separate analyses in which one might wish to engage on minority participation and performance on these examinations, the point here is that this second set of classic demographic variables is not much help in explaining the overall trends in test scores evident in the reported data.

Gender. The percentage of women test-takers rose dramatically over the seven year period: from 29% to 39% on the LSAT and from 27% to 37% on the MCATs, for example. Would this trend (also implied by data on the GREs and GMATs) explain the recent trends in test scores?

If we accept stereotypes, women are supposed to perform above average on the verbal sections of examinations and below average on the quantitative. The trends, however, suggest precisely the reverse. In fact, fragmentary data from 1978-79 to 1981-82 evidence declining GRE-V scores and advancing GRE-Q scores for women. But as the same trends hold for men, gender cannot be a critical variable here. On the GMAT verbal and quantitative sub-test scores for the years 1977-1982, women's performance basically was unchanged (+.03 S.D.U. for the GMAT/Verbal and +.02 S.D.U. for the GMAT/Quantitative). It is hard to attribute stability in academic performance to gender.

What may be happening, though, as women raise their educational aspirations, is that more women of average ability are taking these examinations, or, at the least, that there is greater variance in the abilities of female examinees. As Turnbull's marginalia on an early draft of this paper noted, "here. . .the conventional wisdom [concerning numbers of test-takers] applies." If that were true, though, the Standard Deviations for women would have risen. Unfortunately, the public data on this issue are fragmentary. Nonetheless, even the fragmentary data suggest that precisely the opposite has occurred. The Standard Deviations for women on the LSATs have remained fairly stable since 1975-76; those on the GMATs have fallen dramatically since 1977-78; and those on the GRE/V have declined slightly since 1978-79. Only on the GRE/Q has the Standard Deviation for women risen.

15. Explaining the Changes, Round 3: Citizenship and Native Language

One of the most noticeable characteristics of GRE and GMAT test-taking populations is a significant percentage of non-U.S. citizens (as

previously noted, neither the LSAT nor the MCAT reports identify citizenship or native language*). For the GMATs, that proportion has hovered consistently around 20% since 1977 (data for earlier years are unavailable). For the GREs, the proportion of non-U.S. citizens rose from 7.5% in 1975/76 to 13.3% in 1981/82. This trend becomes more significant in light of a parallel increase in the percentage of GRE test-takers who say that English is not their native language: from 6.0% in 1975/76 to 10.2% in 1981/82 (among U.S. citizens, however, the percentage of those who felt more comfortable in a language other than English remained stable and low--about 2%).

Graduate and professional education in the United States has a worldwide reputation for quality. And the systems of higher education in many other countries do not provide the same opportunities for advanced study as we do (whether the American Ph.D. in some fields is regarded overseas with the same reverence we accord it here is another question--one that lies beyond the domain of this discussion). So there is nothing surprising about a significant number of foreign nationals applying to our graduate and professional schools, and taking the requisite qualifying examinations.

Is this proportion of foreign student test-takers on the GRE/General examinations and the GMAT sufficient to influence mean scores and Standard Deviations? Yes. Does it explain the changes we observe of the 18 year period, 1964-1982, let alone the shorter-term (1977-1972) for which we have full information on test-takers? No.

Where there has been a shift in the ratio of non-U.S. citizens, we have some clues as to how those test-takers will influence scores. The GRE/General examinations offer such a case. And the GMAT data confirm the trends evident in the GREs. Table H illustrates the comparative performance of "domestic" and "foreign" students on these examinations --the only two for which we can disaggregate scores. Leaving aside the way in which the GREs define "domestic" test-takers (the definition is a dubious one), and leaving aside the various methods of disaggregation, we can clearly observe that foreign nationals drag down the mean scores on the <u>verbal</u> sections of these examinations and prop up the mean scores on the <u>quantitative</u> sections.

Without foreign nationals in the picture, for example, the GRE/Q would have <u>declined</u> by (.15) of a Standard Deviation Unit instead of remaining unchanged between 1964-1982. Without foreign nationals in the picture, the GRE/V would have declined by (.38) of an S.D.U. instead of (.48) over the same period.

But what we do <u>not</u> know, in both cases, is the percentage of those foreign nationals who graduated from U.S. colleges and universities <u>and/or</u> who were native speakers of English (i.e. held citizenship in Canada, the U.K., Ireland, Australia, New Zealand, and nations in both

---

* Perhaps with good reason, as only 0.3% of all law degrees awarded in 1980-81 and only 0.8% of all medical degrees--as opposed to 12.8% of the Ph.Ds.--went to non-resident aliens. (24)

Africa and the West Indies where English was the colonial language, and hence the language of the schools and colleges).

To illustrate one part of this point: for 1981-82, the GMATs indicate that 21% of the test-takers were non-U.S. citizens and 21% did not speak English as a native language. But 7.5% of the test-takers were citizens of English-speaking nations other than the U.S. or nations where the colonial language was English. What that may mean for the GMATs is that we can reduce the effects of the performance of foreign nationals on the Verbal sub-score by roughly one-third (7.5% over 21%). If the same pattern holds for the GREs (and we do not know because the GRE does not cover this information), then the effects might be diluted in a parallel manner. Of course, this is all hypothesis: to prove the case one needs a statistical analysis using primary data.

If commentators want to claim that the performance of foreign students accounts for most of the change in GRE scores we have witnessed in recent years (and, indeed, we have heard such claims), they cannot do so on the basis of the reported data. In order to support that claim, one needs to look at a critical universe of test-takers who neither (1) speak English as a native language, (2) reside in the United States, nor (3) graduated from a U.S. college or university. And yet none of the examinations reports performance for this critical group. The GMATs present cross-tabulations of scores by native language and by country of citizenship (the Graduate Management Admissions Council is to be commended for requesting that information), but do not indicate which of those students graduated from U.S. colleges. The GRE Summary distinguishes resident from non-resident aliens in gross numbers, but the data in Table H (drawn from other ETS studies) do not make the same distinction. One suspects that a resident alien at the time of test-taking is likely to be a graduate or soon-to-be-graduate of a U.S. college or university; hence, his/her score should not be included among the "foreign means."

I continue to use the GMATs as a gloss on this issue because, among all the tests we are examining, the GMATs provide the most detailed and comprehensive data. We can thus use the GMAT data to provide us with some insight as to the impact of foreign nationals on quantitative scores. The best way to elicit this insight is to perform an analysis by two variables: citizenship and undergraduate major. Table K breaks out the performance of U.S. citizens and foreign nationals on the 1981-82 GMAT for selected quantitatively-oriented undergraduate majors and selected non-quantitatively-oriented undergraduate majors. From this data, we can conclude that:

(a) U.S. citizens and foreign nationals who majored as undergraduates in quantitatively-oriented subjects perform equally as well on the quantitative sections of the GMAT; and

(b) Foreign nationals who majored in non-quantitatively based subjects as undergraduates outperform their U.S. counterparts on the quantitative section of the GMAT.

If the same phenomena hold for the GREs (and we cannot tell because the way the data are reported in GRE Summary does not allow us to engage in this analysis), and if the percentage of foreign nationals taking the GREs who majored in non-quantitatively based subjects has been rising, then that might account for part of the rise in the GRE-Q scores since 1975.

Can we determine whether the trends in performance on the GRE Subject Area Tests have been influenced by the participation of foreign nationals? Since, to the best of my knowledge, there is no available data on the percentage of GRE Subject Area test-takers who are foreign nationals, I tried an indirect route to answering the question. While the route does not bring us to the answer (indeed, there are strong arguments against taking it in the first place), it raises a critical issue and hence is worth recounting.

The National Research Council reports regularly on Doctoral degree recipients from U.S. universities, covering citizenship (resident and non-resident) by field (among other variables). (25) Here, for example, are the percentages of doctorates that were granted to foreign nationals in 1980 in selected fields for which there are GRE Subject Area Tests. The list is arranged in the order of net changes in test scores as reported in Table D:

Field	Percentage of Foreign National Doctorates, 1980	Change in Test Scores (in S.D.U.s) 1964-1982
Mathematics	27.1%	+.28
Physics	24.4	+.17
Economics	31.9	-.08
Chemistry	21.8	-.11
Engineering	46.3	-.22
Psychology	3.9	-.26
Education	8.2	-.28
History	6.3	-.70
English	5.2	-.72

It certainly is coincidental that the greater the drop in test scores, the lower the participation of foreign nationals--at least among doctorates. But that is just coincidence. After all, those who received the doctorate in 1980 began their graduate studies an average of 9.3 years before that time. And those who actually receive the doctorate are a very select sub-set of those who took the GRE examinations those many years previously. There are other reasons not to pay much attention to the coincidence (e.g. due to practices in the U.S. labor market, native graduate students in Engineering do not tend to seek the Ph.D., rather stop at the Master's level), but the exercise raises what may be the most critical variable of all in explaining trends in performance on all the examinations under consideration: undergraduate major.

16. Explaining the Changes, Round 4: Undergraduate Major

The basic demographic variables did not help us explain the trends in test scores of college graduates. And the last of those variables--citizenship and native language--while offering some potential, also proved unsatisfactory until we introduced a non-demographic variable, the nature of formal schooling as reflected in undergraduate major.

It stands to reason that people who take examinations for admission to graduate or professional school bring a tremendous amount of intellectual training to the testing room. Whether they are taking an examination covering knowledge and methods in a specific field such as Biology or History, or an examination that seeks to elicit facility in certain general modes of thought, e.g. inductive problem-solving on the GMATs or deductive reasoning on the LSATs, students bring anywhere from 15-20 years of formal schooling to the examination. While the effects of formal schooling are always cumulative, what scholars of rhetoric and propaganda call the "recency effect" is rather strong when it comes to performance on an examination. That is, whatever you have been studying or doing during the few years before you take an examination will have a much stronger bearing on your performance than earlier schooling or experience. For most test-takers, that "whatever" is the undergraduate major.

To be sure, there are other influences, some having to do with formal schooling, others having to do with work experience. For 23% of the GRE test-takers (and 14% of the LSAT test-takers) enrolled in graduate school in 1981-82, the graduate program might have great effect. For the 50% of the GMAT test-takers who have two or more years of work experience (the Graduate Schools of Business and Management encourage practical work experience among applicants for admission), the particular context and nature of that work might have considerable influence on test performance. And certainly all of us can cite sloppy intervening variables in our lives that might influence our performance on different measures of learned ability or specific field knowledge. But few of us expend as much time or effort on any aspect of our learning as we do on our undergraduate major. Certainly, that experience should result in observable effects on examinations such as these.

The analysis of test performance by undergraduate major is presented in Table F, first by year and then by major. Since a number of categories represent aggregated scores, I could not use Standard Deviation Units to indicate differences in a secondary data analysis such as this. Therefore, and with some reluctance, I chose to ask the question:

> "By what percent did the mean score of those who majored in ________ differ from the mean score of all test-takers who identified their undergraduate major?"

This question was asked with reference to 30 different categories of undergraduate major on the tests of general learned ability (GRE/V, GRE/Q, LSAT, and GMAT)*, and for the years for which the data was

---

*The MCAT is excluded from this analysis because it reports for only seven broad categories of undergraduate major.

available for all three tests (1977-1982). What can one conclude? And what hypotheses can one offer for further exploration?

> Conclusion #1: With the exception of Engineering majors, undergraduates who major in professional and occupational fields consistently underperform those who major in traditional arts and sciences fields on these examinations.

Table F covers fields that are common in the data reporting of the three examinations in question. Hence, by "professional fields," I am specifically referring to Business Administration (and its allied or sub-fields), Education, Social Work, and Journalism. U.S. citizen test-takers who identified their major in one of these professional fields accounted for approximately 30% of U.S. citizens who identified their major for any of the 30 fields listed.

In addition, one can turn to the GRE and LSAT data for such fields (not listed in Table F) as Health Administration, Pharmacy, Agriculture, Nutrition, etc. and find the same pattern of underperformance (Nursing and Architecture--both of which are programs that often run more than four years--are exceptions). In 1981-82, these other fields would have added approximately 8,000 test-takers to the professional/occupational category, and increased the overall representation of undergraduate professional/occupational field majors to 32% (Nursing and Architecture would add 9,800 test-takers, and bring the professional/occupational proportion up to 34% of those who identified undergraduate major).

Now this group comprises a substantial and growing portion of the test-takers. Why undergraduate professional/occupational majors have an almost exclusive purchase on the bottom of the performance barrel on these examinations is not hard to see. As recent ACT data on entering college freshmen have demonstrated, most of these majors (again, Nursing, Architecture and Engineering excepted) do not attract the best students (see Table L). But this "input" explanation, by itself, is insufficient. Driven by the requirements of specialized accrediting bodies, the curricula in many of these areas tend to be confined to very few fields, none of which require the exercise or development of the verbal skills necessary to perform well on examinations such as these. Nor do any of these professional/occupational "disciplines" have strong knowledge paradigms, structures that require the rigorous exercise of analysis and synthesis that is so often reflected on the tests.

I realize this is a sweeping statement, and one that is part of an old argument in American higher education. But the test data support this inference. One example should suffice. Let us compare the performance of Economics majors with that of Business majors on the GMAT, an examination that indicates whether a student is prepared to undertake graduate work in Business Administration. As Tables F 4-8 and F-13 show, Economics majors rank in the top performing groups on the GMAT while the various Business majors rank at the bottom. To be sure, nearly five times as many Business majors take the GMATs as do Economics majors (though Economics majors are more likely to take the examination

than Business majors). But the "conventional wisdom" about numbers and test scores is not be sufficient to explain the extreme differential in performance.

Economics is a discipline with a strong knowledge paradigm, emphasizes abstract models and theory, requires research, and seeks predictive knowledge. It is a basic discipline. We know that undergraduate Business majors (the GMAT categories include Accounting, Finance, Marketing, Management, Hotel Administration, Real Estate, and others) are required to take some Economics, though the core requirements are rather minimal. The accreditation guidelines of the American Assembly of Collegiate Schools of Business are not even specific in this regard. (26) At the same time, we assume that Business majors receive a good deal of training in basic quantitative skills and problem solving, both of which are heavily emphasized on the GMAT. But the contexts for that emphasis on the GMAT--as in the business world itself--are verbal, draw on theory and models, require a broad knowledge of the world and political and cultural forces in American society, and challenge students to demonstrate prowess in the kind of reasoning in which the researcher engages. None of these contexts are developed in specialized programs in Business Administration (or allied fields). And when departments follow accreditation guidelines and students take a minimum of one-half of their credits in Business Administration, it is no wonder that Business majors do not perform well on an examination used in the admissions process to graduate schools of Business.

> Hypothesis : The greater the proportion of test-takers whose undergraduate majors were in professional/occupational fields, the lower the test scores.

It would be reassuring to unlock the mystery of test-score trends solely with reference to undergraduate majors, but we do not have the information to do so in a secondary analysis. Hence, a hypothesis.

While it may be the case that test-takers with professional or occupational degrees underperform others, the effect on mean scores over time would not be significant unless the proportion of test-takers from these fields: (1) is substantial to begin with, and (2) has risen significantly during the period under consideration.

We know that the proportion of bachelor's degrees awarded in professional and occupational fields has increased from 51% in 1971 to 64% in 1982, a 25.5% increase in eleven years. (27) And we know that the proportion of test-takers from the professional and occupational fields (exclusive of Engineering) has increased from roughly 29% in 1977 to 34% in 1982, a 17.2% increase in five years. Common sense will allow us to extrapolate from both these trends.

Even if arts and sciences majors are twice as likely to take these examinations as others, we can say that, since 1971 (and at a minimum), there has probably been an increase in the proportion of test-takers who hold professional/occupational degrees equivalent to the increase in

the proportion of the degrees awarded in those fields, namely 25.5%.

However, there is reason to believe that the increase in the proportion of test-takers whose undergraduate majors were in professional or occupational fields has been greater than 25.5% since 1971. The fastest growing examination in terms of candidate volume from 1971 to 1982 was the GMAT (146% increase), and GMAT test-takers are more than twice as likely to hold professional/occupational degrees as GRE test-takers.

This is a difficult case, and one for the statisticians to confront. That is why I have advanced it as a hypothesis. Regression analyses are necessary to isolate the effects of a variable such as undergraduate major over a period of time during which tests were reweighted, rules were changed and disclosure was added to the complications. And the demonstrable effects would have to be statistically significant. Better still would be research with primary data that could demonstrate that entering freshmen <u>of similar ability</u> who pursue professional/vocational v. arts & sciences curricula evidence significantly different degrees of achievement on these examinations.

> <u>Conclusion #2</u>: Students with undergraduate majors in science, mathematics and Engineering perform better than all others on these examinations.

These students perform better not only on quantitatively-based examinations (as we would expect), but also, comparatively speaking, on examinations that rely heavily on verbal abilities. The LSAT is a case in point: all the science majors outperform students with undergraduate majors not only in professional fields but also in fields that conventional wisdom holds would be good preparation for the law, e.g. Political Science and History. And, in this group of majors, only Engineers and Computer Scientists perform below average on the GRE/Verbal. In fact, Biology and Physics majors outperform Psychology, Art and Music majors on the GRE/V.

How do we explain? Our first cut sounds like an "input approach." Undergraduate science and math curricula generally attract more gifted students who are willing to undergo the considerable rigor that those disciplines demand, and who probably have higher academic aspirations (hence are more likely to take at least one of these examinations). That the students may be more gifted is hinted at by the GRE Subject Area Tests, where the long-term trends in scores in the sciences and mathematics are far more positive than those in other fields.

But when one looks at the list of majors that consistently outperform all others across these examinations (see Table F-13), one notices some non-scientific fields, e.g. Philosophy and Economics, and hence cannot be wholly comfortable with the "input" explanation. Considering the addition of those fields, Conclusion #2 should be modified.

> <u>Conclusion #3</u>: Students who major in a field characterized by formal thought, structural relationships, abstract models, symbolic languages, and deductive reasoning consistently outperform others on these examinations.

Mathematics, Economics, Philosophy, Chemistry, Engineering--all of these require the use of symbolic languages, all are characterized by structural relationships that proceed according to rules, and all require students to exercise the powers of <u>deductive</u> reasoning. To be sure, that is but a very general account of some of the key common elements of the knowledge paradigms of those disciplines, and no doubt there will be some quibbling. But both the LSAT and GMATs, in particular, require students to exercise the modes of thought implied by those paradigms.

One can verify this conclusion by breaking up some of the 30 categories of majors in order to isolate others--such as Music and Linguistics--that evidence similar knowledge paradigms. Indeed, Music majors, whose work in composition and theory is very structural, symbolic, deductive, etc., hold up the scores in the Fine Arts categories. And if one could disaggregate the "Foreign Language" category across all three tests, I think it is reasonable to speculate that those who major in the more highly inflected and "synthetic" languages (such as German, Russian, or Latin) would perform equally as well as the Mathematics, Economics, Philosophy, etc. majors--and for similar reasons.

To be sure, Humanities majors (particularly those in English and Foreign Languages) perform well above the mean on all these examinations (except the GRE/Q). One can explain, in part, with reference to the nature of the tests. For example, the GMATs emphasize "analysis of situations" questions that call for the close reading of text and the sensitivity to nuance to which English, Foreign Language, History and Philosophy majors are accustomed--one reason that they perform comparatively well on the GMATs.

But that is only a half-explanation. After all, when one looks back at the GRE Subject Area Tests (Table D), it appears that, as Eldon Park of Graduate Record Office at ETS observed, the virtual bottom has fallen out of some of the disciplines (28), and I would include the Humanities among them (though the case concerning the Social Sciences is far more severe). What is going on?

> <u>Conclusion #4</u>: Even as the numbers of their majors decline, the Humanities disciplines are witnessing their best students going on to professional school, not graduate school. But we cannot reach the same conclusion concerning the Social Sciences.

As Table F-9 shows, the percentage of test-takers on the LSATs, GMATs, and GRE/Generals from the various undergraduate Humanities fields has remained relatively stable in recent years. At the same time, however, Table A demonstrates a precipitous decline in the number of candidates taking the GRE Subject Area Tests in English and History, even while performance on these tests plummeted. We can thus infer that a greater percentage of Humanities majors are moving into professional schools following college.

Consider the evidence: the test scores on the LSAT are up, those on the GMAT are relatively stable, and those on the GRE/Verbal and Subject Area

(French, History, English) are down. While the percentage of Humanities majors among the LSAT and GMAT test-taking population is not sufficient to determine trends in scores, no doubt the shift has contributed to those trends. It may be worth adding that even though they comprise only 3% of the first-time examinees on the MCATs, Humanities majors have outperformed all others (including Biological Sciences majors) on the Biology sub-test since 1980 (and outperformed all others on the Reading sub-test since the new MCATs were introduced in 1977).

Test results for majors in the Social Sciences are much more variable and do not allow for a similar interpretation. Social Science majors comprised 24.7% of those who identified their major in 1981-82 (down from 27.6% in 1977-78). As in the case of English and History, the steepest declines in test participation occurred in subjects for which performance on the GRE Subject Area Tests fell most dramatically during even that short recent period:

GRE Subject Area Test	Decline in Participation, 1977-1982	Change in Test Performance, 1975-1982 (S.D.U.)
Economics	5.7%	+.15
Psychology	11.8%	+.01
Political Science	34.2%	-.13
Sociology	44.8%	-.20

Our hypothesis concerning the movement of the better English, History and other humanities majors to graduate work in the professions does not bear fruit here. Neither Political Science nor Sociology majors perform comparatively well on either the LSAT or the GMAT. One might conclude that the better students who major in those disciplines are not participating at all in these examination programs.

17. General Conclusions

What should we make of all this? The standardized test scores of college graduates (and soon-to-be-graduates) have declined. Even in the most recent period covered by this analysis, 1976-1982, scores generally continued to decline, though with modest slopes and with noted exceptions in professional and some quantitatively oriented fields.

It bears repeating that in no place in this paper do I mean to imply that trends in scores on standardized tests should be the principal indicator of quality in American higher education. It is rhetorically discomforting to repeat that statement too many times.

Nor should anyone take this analysis to be definitive concerning the causes of change in test performance over time. Student perceptions of the necessity of graduate or professional education in relation to their understanding of the labor market, for example, may be significant in determining the composition of the test-taking population (and that, in turn, may influence our judgment of the aggregate measures of performance). Such hypotheses concerning external influences are

numerous, but the data are sparse, and this paper does not pretend to tread in territory where the roadsigns are few. What we have here, instead, are hypotheses grounded in data internal to the testing process, that seem to make sense, but that should be put to the test of statistical analysis by others who are better qualified to do so and who have access to the primary data.

My general conclusion may sound like a begging of the question: we need that research, and with particular attention to two phenomena that have been stressed in this analysis:

The first of these covers the three periods of rate and direction of change in the scores, along with explanations for the "turning points," explanations related principally to matters of test content, content weighting, administration and score reporting. This phenomenon has nothing to do with the test-taking population.

The second covers change in the test-taking population in terms of the undergraduate major and academic experience of test-takers, in combination with controls for ability based on SAT or ACT scores. Of all the factors that influence performance on these examinations, these appear to be the most persuasive. None of the other variables we examined--numbers of test-takers, age, race, gender, citizenship, native language--seems to mean as much until it is combined with undergraduate major; and controls for ability at the point of college entrance seem very important to our understanding of the impact of college major on student performance. There are, of course, statistical techniques for exploring these hypotheses, and I urge the very capable research personnel at the testing services to follow through, using as much of the primary data as they can clean up.

But there is a great challenge in designing the research on the impact of undergraduate major: it has to fit the context of those three periods of change. For example, if we isolate the period of sharply declining test scores (1964-1970), we may find that undergraduate major has far less of a correlation with performance than demographic variables. The results may be completely different within the more recent period (1976-1982) of relative stability/modest decline in test scores.

It is this more recent period of relative stability/modest decline for which we have considerable data on the test-taking population, and is perhaps as good as any period to begin monitoring the future. But for us to monitor well, we need better information and cleaner data.

18. A Message to the Testing Services: Gathering Consistent Data

Throughout this exploration, we have been frustrated by data that has been either inconsistently gathered or inconsistently reported. At the outset, I pointed out that the testing services have five constituencies, but that they serve only one when they report data. It does no violation to the basic responsibilities of the testing services to their primary clients if they also gather and report much more accurate and consistent data for use by the others.

There are at least two more compelling reasons for the principal clients of the testing services to be more "public-spirited" in this matter, and to go to the additional effort and expense that the gathering and accurate reporting of more detailed data entails.

- o To mitigate the effects of the mis-use of tests. As previously pointed out, the lay public and policy makers--let alone higher education administrators--are going to use test scores in highly symbolic ways and will make academic judgments and develop academic policies on the basis of their perceptions. They will take all of these actions whether we like it or not. But the nature of their perceptions and judgments will result in gross misuse of the scores if the data are inconsistent and incomplete.

- o To stimulate improvement in higher education. That is, if the testing programs can gather data that will help faculty and administrators think more carefully about the factors that influence student academic achievement and growth and about the most appropriate and productive ways of assessing that achievement and growth, everyone benefits.

But the first task is one for the historians. My experience with the LSAT data indicates that it is possible to go back into the existing files and clean them in ways so that information can be reported:

(a) only for those who actually took a particular test in a given testing year (and for those who took it more than once, to report only the most recent score within that year); and

(b) for the same universe of test-takers on each variable.

In this way, the testing services could reconstruct their public statistical histories to reflect the performance of real people who took the examinations, and could provide at least a modicum of consistency for future analyses.

But cleaning up past data is not as important as providing full and accurate data for the future. The existing information is extensive, but in many ways intractable. The suggestions offered below are designed for the public statistical histories that will be analyzed at the turn of the next century.

The first group of suggestions concerns the administration of information:

1) The administrators of the GREs, GMATs, LSATs, and MCATs should require that all students fill out their respective background information questionnaires as a condition of taking the examination.

By "all" students, I mean all: pre-registrants, walk-ins, Monday test-takers, etc. And at all administrations: domestic, foreign and special. We cannot afford to work with incomplete data.

We know that some students are always suspicious of questionnaires and their potential uses (and abuses), and that some will always resist--even if we protect their anonymity and sign a statement of assurance that the information will be used only for research purposes. But we should insist.

Now I admit that this is a very hard line, and one that might prove counterproductive. We have a choice between honest and complete data from an incomplete sample or frivolous and contaminated data from a complete sample. I am an optimist, and prefer to take the risk. Besides, we're collecting contaminated data now (e.g. self-reported grades), and don't seem to be too worried about it.

2) No irregularities should be included in any data base drawn from the tests and/or background questionnaires.

If a student cancels a test score within five days of an examination (as allowed), the background questionnaire should be cancelled as well. If a pre-registrant fails to show up for a particular test administration, the background questionnaire should not be included in the data files. And if a student walks out of an examination and capitulates on the spot, the background information collected by questionnaire should likewise not be included. Unfortunately, all these irregularities (and others) turn up in the existing data; and the irregularities do no great service even to the primary clients of that data.

The second group of suggestions concerns the content of the information gathered and reported. The idea is to anticipate interpretive issues and demographic trends, not to wait until they appear and then be caught in uncomfortable webs of speculation.

3) All four examination programs should ask the same questions in the same way about citizenship, native language, and country in which the examinee's baccalaureate institution is located.

Questions about citizenship should distinguish between resident and non-resident status. Questions about language should determine not only native language, but also primary language of the examinee's parents (this is particularly important for U.S. minorities), and the language of instruction in the schools and colleges attended. For non-resident aliens, it may also be helpful to understand what kind of postsecondary institution they attended (technical institute, regional university, national university, agricultural college, etc.)

4) All four examination programs should ask the same questions in the same way about the type and characteristics of undergraduate institutions attended by U.S. citizens and resident aliens who graduated (or expect to graduate) from U.S. colleges.

In his study of institutional diversity in higher education,

Robert Birnbaum provides some options for institutional classification that may be more helpful than the old Carnegie typology (or even its latest refined edition). (29) The information one could garner by asking these questions is important in light of the changing enrollment mix in American higher education. Our interpretation of test performance by undergraduate major, for example, would have been enhanced by empirical evidence of where the examinees received their baccalaureate degrees (at least by institutional type).

5) All four examination programs should ask the same questions in the same way concerning the work experience of examinees.

Only the GMATs ask any question about work experience; and confine that question to the simple notion of extent. In considering performance on tests of general learned abilities, however, it may be beneficial to know more about the type of job(s), setting(s) and responsibilities. At present, approximately 20% of the test-taking population (exclusive of those who take the MCATs) is over the age of 30. That percentage is bound to increase, and intervening variables will become even more important to the analysis of trends.

6) The four examination programs should jointly develop and adopt a comprehensive list of undergraduate majors for use on their background information questionnaires.

The GRE list, with the addition of discriminations among Business fields (Accounting, Finance, Marketing, Management, etc.), and trimmed of some existing discriminations that may be a bit too fine (e.g. Social Psychology, Parasitology, Slavic Studies) would probably do the job. When reporting, data on mean scores, Standard Deviations, and candidate volume should be listed by both individual major and aggregate field (e.g. French/ Modern Languages; Finance/Business Administration; Zoology/Biosciences). If this information is important now, it may become more important in the future if our colleges either consolidate programs or move toward further proliferation.

7) All four examination programs should add a question to their background information questionnaires that will help determine why an individual is taking a particular examination.

The commentaries on test score analyses frequently include statements such as, "students take these examinations for many different reasons." But as long as we fail to ask the question, those statements are wholly speculative. Among the options are graduate school admissions requirements (or recommendations), self-assessment, participation in local program evaluations at colleges, requirements for the undergraduate degree, requirements in graduate school programs, requests of employers, student perceptions of the labor market etc. I am sure that the examination programs can come up with a productive question here.

8) All four examination programs should refine their definitions of "age" in reporting data in order to provide accurate information on the educational careers of examinees.

Here I am indebted to Nancy W. Burton of ETS who, in an internal proposal to restructure existing data on the GREs, pointed to the critical distinction between age at testing and age at baccalaureate. (30) To those two points we should add "age at graduate degree" for those who already hold graduate degrees.

The third set of suggestions concerns longitudinal research on academic performance and the impact of schooling in the United States. Ideally, we should key off the High School and Beyond sample (Class of 1980) for which college transcripts are currently being coded. The sample has been drawn, and the base-line data are there (including high school transcripts, SAT/ACT scores, and other performance data). The analysis of college transcripts in fine detail (31) will be one of the most important breakthroughs in our understanding of the college curriculum as experienced by students, and will better enable us to assess performance on one or more of the major examinations used in graduate and professional school admissions. How much of the original HSB sample will remain by the time we get to the examinees is almost beside the point if the intent is to establish a true longitudinal study with standardized performance measures; but to insure a robust sample, we could supplement High School and Beyond with a college cohort selected primarily to reflect the distribution of undergraduate majors in different types of institutions.

Better still would be a parallel undertaking. It is possible to identify a sample test-taking population at the point of matriculation in college (not high school graduation), which could be traced through various undergraduate experiences (academic and otherwise), employment and family life, performance on these examinations and performance in graduate or professional school. A convincing sample of test-takers (one which reflects the demographic, ability, and enrollment-mix characteristics of all entering college freshmen, full- and part-time, and in all types of institutions) would allow us to control for all those intervening or unknown variables to which we currently turn as blind excuses for trends that we cannot seem to explain otherwise. Since we would be able to ask our experimental group to sit for a number of examinations, we would also be able to test the validity of grouping the GRE/General examinations, GMAT and LSAT together as tests of "general learned ability"

19. A Message to the Commentators: a Plea Against Excuses

Any observer of discussions of test scores over the past two decades cannot help but noticing a very strange phenomenon. Whenever it is apparent that scores on standardized tests have declined, the testing services, examination governing boards, college administrators and (sometimes) the media rush to out-do each other with explanations so intense and abstruse that they appear as excuses. Whenever scores rise, however, no such explanations are offered.

The products and services of the testing industry may not be ideal, but are of generally high quality. Certainly we have been flooded with

enough studies of reliability and validity to convince us that the products and services are worth what we pay for them. Paradoxically, our excuse-mongering for performance casts unnecessary doubts on the quality of the tests. Whenever the testing services, governing boards, deans, admission officers, etc. engage in excuse-mongering, they seem to be saying, "Don't take the product seriously!" What strange behavior!

An analogy from the business world may be appropriate. The behavior of the groups that control and operate the tests is like that of a CEO who has to report a quarterly loss to stockholders, and starts reciting arcane accounting jargon about write-offs and deferred maintenance, let alone less arcane observations about strikes or energy costs or special R&D investments. Yet if that same CEO presents a quarterly report evidencing an outstanding gain in net earnings, it is often left to individual stockholders (let alone securities analysts or SEC investigators) to determine how much of the gain was caused by extraordinary items or unusual circumstances. None of these explanations (offered or hidden) have anything to do with the quality of the basic products or services sold by the company, nor do they have much to do with operating income.

If the test scores decline, we blame everything but the quality of student learning. If the scores go up, any commentary is complimentary to schools, colleges, and students. This contradictory behavior does no great service to students--let alone anyone else associated with the education enterprise. In the course of this paper, we found that a number of the standard nostrums used in the excuse process do not have "face validity" (or, at the least, found no evidence to suggest their validity). But we also pointed out that it is important to understand why performance on these measures changes over time.

It appears that what many commentators fear--and why they are defensive concerning declining scores--is misuse of test results. This is a just fear, and I share it. But it is not an excuse for excuses. As the British philosopher, J. L. Austin, has observed, an excuse involves a situation "where someone is said to have done something which is bad, wrong, inept, unwelcome, or . . . untoward. Thereupon he or someone on his behalf will try to defend his conduct or get him out of it." (32) This situation, normally a province of ethics and/or the law, is somewhat out of place in the realm of academic performance. And "explanations" for "inept" behavior which often are used in the realm of academic performance are dominated by conditions, exemptions and qualifications that sound too much like excuses (at least in their linguistic properties). The problem is that we use these "explanations" as if they were statements of cause--which they are not.

Instead of excuses, the justified fear of misuse of test results should be a spur to the search for the factors within our control that are most likely to influence student learning. An explanation of test scores that focuses on something other than teaching, curriculum and learning is an excuse that will distract the energies and efforts of those who make educational policy, and will probably turn those energies and

efforts away from the forces that make a real difference in the professional lives of college faculty and the learning of college students.

So I conclude with a plea for both candor and focus: the scores and trends do not explain everything in the world of higher education. None of them help us measure the development of leadership, artistic talent, organizational skills, and personal values, for example, that we expect colleges to advance in different students and upon which the national culture and polity depend. At the same time, the measures are proven ones, and the results are important indicators of student learning. As Alexander Astin remarked in discussing a draft of this paper, the remarkably simple but sensible notion that "you learn what you study" seems to escape people when they talk about test scores. I hope it is an equally simple but sensible notion to propose that if we use these indicators to raise the level of discussion about the college curriculum, we will increase their role in the learning and development process. In this way, the examinations can become formative tools as well as summative measures. Let us not turn our attention away from that challenge, for student learning is the bottom line of our business.

NOTES

1. (Page 1) Adelman, Clifford. "The Major Seventh: Standards as a Leading Tone in Higher Education," in J. R. Warren (ed.), Meeting the New Demand for Standards. San Francisco: Jossey-Bass, 1983, pp. 39-54.

2. (Page 2) Resnick, Lauren and Daniel Resnick. "Standards, Curriculum, and Performance: an Historical and Comparative Perspective." Commissioned paper for the National Commission on Excellence in Education, 1982. ED #227-104, p.1.

3. (Page 4) Linda Cole, Statistical Coordinator for the College Board Achievement Tests at ETS, has performed such studies on Math: Level 1, Math: Level 2, Biology, Chemistry, Physics, American History and English Composition, using the 1980 scale on which to measure the effects of rescalings since 1972-73. The author has examined a handwritten summary, in table form, of the results of these studies.

4. (Page 7) Jerilee Grandy. Profiles of Prospective Humanities Majors: 1975-1983. Princeton, N.J.: Educational Testing Service, 1984. Final Report # OP-20119-83 (National Endowment for the Humanities), p. 70.

5. (Page 8) The figures are extrapolated from the following:

Digest of Education Statistics, 1983-84. Washington, D.C.: National Center for Education Statistics, 1984, Tables 81, 86, and 114, pp. 95, 100, and 132.

The Condition of Education, 1982. Washington, D.C.: National Center for Education Statistics, 1983, Table 4.5, p. 134.

6. (Page 8) William Turnbull, personal correspondance with the author.

7. (Page 10) The substance of this paragraph is drawn from both personal correspondance and conversations between Jonathan Warren and the author.

8. (Page 10) Committee of Examiners for the Political Science Test. GRE Examinations, 1982-1984: a Description of the Political Science Test. Princeton, N.J.: Educational Testing Service, 1982, p. 7.

9. (Page 10) ibid., p. 8.

10. (Page 11) W. Ann Reynolds, personal correspondance with author.

11. (Page 11) Wigdor, Alexandra K. and Wendell R. Garner (eds.). Ability Testing: Uses, Consequences and Controversies. Washington, D.C.: National Academy Press, 2 vols, 1982. I, p. 56.

12. (Page 11) Anne Anastasi, "Abilities and the Measurement of Achievement." In William B. Schraeder (ed.), Measuring Achievement: Progress Over a Decade. San Francisco: Jossey-Bass, 1980, p. 8.

13. (Page 12) Nancy Burton, comments to the author on a previous draft of this paper.

14. (Page 12) Committee of Examiners for the Economics Test. GRE Examinations, 1982-1984: a Description of the Economics Test. Princeton, N.J.: Educational Testing Service, 1982, p.7.

15. (Page 12) The literature on the predictive validity of these examinations is large and complex. See Richard O. Fortna, Annotated Bibliography of the Graduate Record Examinations. Princeton, N.J.: Educational Testing Service, 1980. For a convincing synthesis of the arguments supporting the predictive validity of some of these tests, see Robert L. Linn, "Admissions Testing on Trial," American Psychologist, March, 1982, pp. 279-291. The arguments, evidence, and controversies surrounding predictive validity lie beyond the scope and purpose of this paper.

16. (Page 16) Bowen, Howard R. Investment in Learning. San Francisco: Jossey-Bass, 1977, p. 98.

17. (Page 18) It is surprising how sparse the literature is on this issue. In a 1961-65 study, Astin used the National Merit Scholarship Qualifying Test (the PSAT) as a control for 669 students who took the now-discontinued GRE general area tests (in Humanities, Social Sciences, and Natural Sciences) as college seniors. Astin's objectives were to measure the impact of the selectivity of institutional environment on GRE scores. What he found was that, as soon as the controls were applied, the correlations were negative. At the same time, though, stepwise linear regression analyses demonstrated that of all student characteristics on entrance to college, "the most important single determinant" of the GRE scores was "academic ability as measured during high school" (by the NMSQT). Alexander W. Astin, "Undergraduate Achievement and Institutional Excellence." Science, vol. 161, no. 3842 (1968), pp.661-668.

18. (Page 18) William Turnbull, "Project TRACE: a Prospectus." Unpublished paper, n.d.

19. (Page 19) When the GRE/Undergraduate Assessment Program came on stream as a systematic assessment service in 1976-77, it did not reintroduce the local scoring option for the GRE Subject Area Tests. But it did offer individual students participating in the UAP the option of placing their scores in the GRE "history file." The GRE/UAP program, offered independently of the GRE Testing Program, does not seem to have had any noticeable effects on candidate volume on the Subject Area Tests for the years 1976-77 through 1978-79. See Undergraduate Assessment Program Council, UAP Guide. Princeton, N.J.: Educational Testing Service, 1976, p. 18.

20. (Page 21) Grandy, p. 72.

21. (Page 22) Wigdor and Garner, I, p. 189.

22. (Page 23) The New York Times, May 7, 1981.

23. (Page 24) Turnbull, personal correspondance with author.

24. (Page 27) Digest of Education Statistics, 1983-84, Table 104, p.126.

25. (Page 29) Syverson, Peter D. Summary Report: 1980 Doctorate Recipients from U.S. Universities. Washington, D.C.: National Academy Press, 1981. Table 2, pp. 30-31.

26. (Page 32) In fact, the Guidelines for Curriculum state only that "40 to 60 percent of the coursework in the baccalaureate program shall be devoted to studies in business administration and economics," but that "the major portion . . . shall be in business administration." The only economics course listed as an example is "principles of economics."

27. (Page 32) Baker, Curtis O. Earned Degrees Conferred: an Examination of Recent Trends. Washington, D.C.: National Center for Education Statistics, 1981. Table 7, p. 27.

28. (Page 34) Eldon Park, personal communication with author.

29. (Page 39) Birnbaum, Robert. Diversity in American Higher Education. New York: Institute for Higher Education, Teachers College, Columbia University, 1982. Final report of Grant #G-81-0058 (National Institute of Education).

30. (Page 40) Nancy W. Burton. Trends in the Performance and Participation of Potential Graduate School Applicants. Princeton, N.J.: Educational Testing Service, 1982. GREB 82-5, p. 12.

31. (Page 40) The only analysis of college transcripts currently in the literature is that by Robert Blackburn, et. al., Changing Practices in Undergraduate Education. Berkeley, Calif.: Carnegie Council on Policy Studies in Higher Education, 1976. Blackburn and his colleagues surveyed a stratified sample of 271 institutions and analyzed transcripts at ten in order to demonstrate the distinctions between formal degree requirements and the actual student experience of the curriculum.

32. (Page 41) J. L. Austin, "A Plea for Excuses." In V. C. Chappel (ed.), Ordinary Language. Englewood Cliffs, N.J.: Prentice-Hall, 1964, p.42.

BIBLIOGRAPHY

Adelman, Clifford. "The Major Seventh: Standards as a Leading Tone in Higher Education." In J. R. Warren (ed.), Meeting the New Demand for Standards. San Francisco: Jossey-Bass, 1983, pp. 39-54.

Altman, Robert A. and Paul W. Holland. A Summary of Data Collected from Graduate Record Examinations Test-Takers During 1975-76. Princeton, N.J.: Educational Testing Service, 1977.

Altman, Robert A. A Summary of Data Collected from Graduate Record Examinations Test-Takers During 1976-77. Princeton, N.J.: Educational Testing Service, 1977.

American College Testing Program. College Student Profiles: Norms for the ACT Assessment. Iowa City, Iowa: ACT, 1983.

Anderson, Charles J. Student Quality in the Humanities: Opinions of Senior Academic Officials. Washington, D.C.: American Council on Education, 1984 (Higher Education Panel Report Number 30).

Astin, Alexander, W. "Undergraduate Achievement and Institutional Excellence." Science, vol. 161, no. 3842 (1968), pp. 661-668.

Atelsek, Frank J. Student Quality in the Sciences and Engineering: Opinions of Senior Academic Officials. Washington, D.C.: American Council on Education, 1984 (Higher Education Panel Report Number 58).

Baird, Leonard. "Biographical and Educational Correlates of Graduate and Progressional School Admissions Test Scores." Educational and Psychological Measurement, vol. 36 (1976), pp. 415-520.

Baker, Curtis O. Earned Degrees Conferred: an Examination of Recent Trends. Washington, D.C.: National Center for Education Statistics, 1981.

Birnbaum, Robert. Diversity in American Higher Education. New York: Institute for Higher Education, Teachers College, Columbia University, 1982. (Final Report for NIE Grant #G-81-0058).

Blackburn, Robert et al. Changing Practices in Undergraduate Education. Berkeley, Calif.: Carnegie Council for Policy Studies in Higher Education, 1976.

Bowen, Howard R. Investment in Learning. San Francisco: Jossey-Bass, 1977.

Buros, Oscar K., "Fifty Years in Testing: Some Reminiscences, Criticisms, and Suggestions." Educational Researcher, Vol. 6, no. 7 (1977), pp. 9-15.

Burton, Nancy W. Trends in the Performance and Participation of Potential Graduate School Applicants. Princeton, N.J.: Educational Testing Services, 1982. GREB 82-5.

Conrad, Linda, Donald Trismen and Ruth Miller. Graduate Record Examinations Technical Manual. Princeton, N.J.: Educational Testing Service, 1977.

Cronbach, Lee J. "Five Decades of Public Controversy Over Mental Testing." American Psychologist, January 1975, pp. 1-14.

Ferguson, Richard L and E. James Maxey. Trends in the Academic Performance of High School and College Students. Iowa City, Ia.: American College Testing Program, 1976 (ACT Research Report No. 70).

Fortna, Richard O. Annotated Bibliography of the Graduate Record Examinations. Princeton, N.J.: Educational Testing Service, 1980.

Frederickson, Norman and William C. Ward. Development of Measures for the Study of Creativity. Princeton, N.J.: Educational Testing Service, 1975. GREB #72-2P

Goodison, Marlene B. A Summary of Data Collected from Graduate Record Examinations Test-Takers During 1980-81. Princeton, N.J.: Educational Testing Service, 1982.

Goodison, Marlene B. A Summary of Data Collected from Graduate Record Examinations Test-Takers During 1981-1982. Princeton, N.J.: Educational Testing Service, 1983.

Graduate Management Admissions Council. The Official Guide to the Graduate Management Admissions Test. Princeton, N.J.: Educational Testing Service, 1982.

Graduate Record Examinations Board. Graduate Programs and Admissions Manual. Princeton, N.J.: Educational Testing Service, 1972, 1976, and 1980.

Graduate Record Examinations Board. Guide to the Use of the Graduate Record Examinations. Princeton, N.J.: Educational Testing Service, 1981.

Grandy, Jerilee. Profiles of Prospective Humanities Majors: 1975-1983. Princeton, N.J.: Educational Testing Service, 1984. (Final Report for NEH Grant #OP-20119-83).

Hartnett, Rodney T. and John A. Centra. "The Effects of Academic Departments on Student Learning." Journal of Higher Education, vol. 45, no. 5 (1977), pp. 491-507.

Kraft, I. "Admissions to Public Professional Schools and Administrative Openness." Journal of Legal Education, vol. 29 (1977), pp. 52-61.

Levine, Murray. "The Academic Achievement Test: Its Historical Context and Social Functions," American Psychologist, March 1976, pp. 228-238.

Linn, Robert L. "Admissions Testing on Trial," American Psychologist, March, 1982, pp. 279-291.

Manning, Winton H. "Social Trends and the Future of Testing in Schools and Colleges." Unpublished paper, Fall 1983.

Maxfield, Betty D. Science, Engineering, and Humanities Doctorates in the United States, 1981 Profile. Washington, D.C.: National Academy Press, 1982.

Messick, Samuel. "The Standard Problem: Meaning and Values in Measurement and Evaluation," American Psychologist, Oct. 1975, pp. 955-966.

Momeni, Jamshid. Adult Participation in Education: Past Trends and Some Projections for the 1980s. Washington, D.C.: National Institute for Work and Learning, 1980.

Munday, Leo A. and Jeanne C. Davis. Variety of Accomplishment After College: Perspectives on the Meaning of Academic Talent. Iowa City, Iowa: American College Testing Program, 1974 (ACT Research Report No. 62).

Munday, Leo A. Toward a Social Audit of Colleges: an Examination of College Student Outcomes in Terms of Admissions Information. Iowa City, Iowa: American College Testing Program, 1976 (ACT Research Report No. 75).

Oltman, Philip K. Content Representativeness of the GRE Advanced Tests in Chemistry, Computer Science and Education. Princeton, N.J.: Educational Testing Service, 1982 (GREB 81-12p).

Resnick, Lauren and Daniel Resnick. "Standards, Curriculum and Performance: an Historical and Comparative Perspective." ED #227-104, 1982.

Rock, Donald and Charles Werts. An Analysis of Time-Related Score Increments and/or Decrements for GRE Repeaters Across Ability and Sex Groups. Princeton, N.J.: Educational Testing Service, 1980. GREB 77-9R.

Schrader, William B. (ed.) Measuring Achievement: Progress Over a Decade. San Francisco, Jossey-Bass, Inc., 1980.

Schrader, William B. (ed.) Admissions Testing and the Public Interest. San Francisco: Jossey-Bass, Inc., 1981.

Solmon, Lewis C. Schooling and Subsequent Success: Influence of Ability Background, and Formal Education. Iowa City, Iowa: American College Testing Program, 1973 (ACT Research Report No. 57).

Syverson, Peter D. Summary Report: 1980 Doctorate Recipients from U.S. Universities. Washington, D.C.: National Academy Press, 1981.

Turnbull, William. "Education's Report Card: What Will It Say About Colleges?" Unpublished paper, Feb. 1984.

Turnbull, William. "Testing Scores in Perspective," Educational Record,

vol. 59, no. 4 (1978), pp. 291-296.

Undergraduate Assessment Program Council. Undergraduate Assessment Program Guide. Princeton, N.J.: Educational Testing Service, 1976.

Wigdor, Alexandra K. and Wendell R. Garner (eds.). Ability Testing: Uses, Consequences and Controversies. 2 vols. Washington, D.C.: National Academy Press, 1982.

Wild, Cheryl L. A Summary of Data Collected from Graduate Record Examination Test-Takers During 1977-78. Princeton, N.J.: Educational Testing Service, 1979.

Wild, Cheryl L. A Summary of Data Collected from Graduate Record Examination Test-Takers During 1978-79. Princeton, N.J.: Educational Testing Service, 1980.

Wild, Cheryl L. A Summary of Data Collected from Graduate Record Examination Test-Takers During 1979-80. Princeton, N.J.: Educational Testing Service, 1981.

Wild, Cheryl L. Summary of Research on Restructuring the Graduate Record Examinations Aptitude Test. Princeton, N.J.: Educational Testing Service, 1979.

Wilson, Kenneth M. The Validation of GRE Scores as Predictors of First Year Performance in Graduate Study. Princeton, N.J.: Educational Testing Service, 1979. GREB 75-8R.

TABLE A: MEAN SCORES, STANDARD DEVIATIONS, AND NUMBER OF TEST TAKERS, 1964-1982

Test Title		1964-65	1965-66	1966-67	1967-68	1968-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76	1976-77	1977-78	1978-79	1979-80	1980-81	1981-82
GMAT	Mean	N.A.	485	486	485	481	474	466	462	465	463	461	463	462	465	463	462	467	468
	S.D.	N.A.	96	98	99	103	103	105	106	105	107	108	107	107	104	105	104	104	104
	N	N.A.	40,153	43,652	57,567	67,267	82,031	85,044	81,552	90,100	101,296	123,261	125,415	139,363	159,039	18[illegible],923	205,676	212,513	200,631
LSAT	Mean	510	511	514	516	516	518	519	521	522	527	520	530*	533*	528*	536*	539*	544*	553*
	S.D.	N.A.	N.A.	N.A.	N.A.	102	102	103	102	104	104	105	109*	107*	110*	107*	108*	110*	110*
	N	39,162	44,776	46,752	49,897	59,444	76,315	104,408	119,391	120,198	130,726	127,908	117,618*	117,468*	111,555*	98,307*	94,856*	100,793*	99,928*
MCAT/Read.	Mean	The MCAT was changed in both format and scale in 1977. Reporting is on a calendar year basis.												7.98	8.01	7.90	7.73	7.50	7.74
	S.D.													2.41	2.49	2.28	2.53	2.52	2.53
	N	N.A.	19,705	22,288	26,539	28,880	33,869	45,324	51,696	54,835	58,218	57,522	53,599	56,579	51,791	48,075	49,646	48,203	47,597
MCAT/Quan	Mean	The MCAT was changed in both format and scale in 1977. Reporting is on a calendar year basis.												7.99	7.91	7.78	7.56	7.45	7.45
	S.D.													2.54	2.53	2.36	2.52	2.46	2.47
GRE/Ver	Mean	530	520	519	520	515	503	497	494	497	492	493	492	490	484	476	474	473	469
	S.D.	124	125	125	124	124	123	125	126	125	126	125	127	129	128	130	131	128	130
	N	93,792	123,960	151,134	182,432	206,113	265,359	293,600	293,506	290,104	301,070	298,335	299,292	287,715	286,383	282,482	272,281	262,855	256,381
GRE/Quan	Mean	533	528	528	527	524	516	512	508	512	509	508	510	514	518	517	522	523	533
	S.D.	137	133	134	135	132	132	134	136	135	137	137	138	139	135	135	136	136	137
Biology (GRE AREA)	Mean	617	610	613	614	613	603	603	606	619	624	623	627	629	622	621	619	617	616
	S.D.	117	115	114	114	112	111	114	115	110	110	110	112	113	113	117	115	115	114
	N	5,228	6,597	7,831	9,467	10,032	12,895	13,844	13,951	14,815	17,280	19,351	20,360	20,845	20,842	18,795	16,693	15,002	14,185
Biology (MCAT)	Mean	A separate Biology test was not offered on the MCAT until 1977. Reporting is by calendar year.												7.87	7.90	7.85	8.03	8.13	8.23
	S.D.													2.39	2.39	2.31	2.57	2.51	2.55
	N													56,579	51,791	48,075	49,646	48,203	47,597
Chemistry (GRE AREA)	Mean	628	618	615	617	613	613	618	624	630	634	629	627	630	624	623	618	615	616
	S.D.	114	110	104	104	104	113	117	124	114	115	105	107	109	108	104	105	103	105
	N	3,783	3,919	4,139	4,781	4,791	5,411	5,350	4,833	4,535	4,648	4,936	5,058	5,268	5,671	5,725	5,422	4,928	4,940
Chemistry (MCAT)	Mean	A separate Chemistry test was not offered on the MCAT until 1977. Reporting is by calendar year.												7.82	7.75	8.06	7.92	7.94	7.99
	S.D.													2.63	2.52	2.30	2.49	2.52	2.48
	N													56,579	51,791	48,075	49,646	48,203	47,597

* Figures for the LSAT from 1964/5 through 1974/5 were taken from data generated by the Educational Testing Service and duplicated on a sheet entitled, "LSAT Score Statistics by Year." The figures for the LSAT from 1975/6 through 1981/2 were derived from data analysis performed under contract to the National Institute of Education by the Law School Admissions Services. As discussed in the text of this paper, there are significant differences between the two data sets, and hence it is difficult to compare pre-1975 data with post-1975 data.

TABLE A-1

TABLE A-2

Test Title		1964-65	1965-66	1966-67	1967-68	1968-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76	1976-77	1977-78	1978-79	1979-80	1980-81	1981-82
Physics (GRE AREA)	Mean	623	623	618	624	624	625	636	643	654	657	655	654	656	652	648	650	648	647
	S.D.	134	136	134	141	140	142	145	146	145	149	147	141	144	146	145	146	147	145
	N	4,002	3,984	4,353	4,795	4,368	4,701	4,122	3,670	3,279	3,250	3,114	3,063	3,024	3,263	3,302	3,255	3,227	3,322
Biology (GRE AREA)	Mean	Fewer than 1,000 Test-takers, therefore not included			603	587	573	572	569	587	581	581	580	581	577	574	576	574	570
	S.D.				104	104	97	98	92	96	94	94	93	92	95	93	90	91	86
	N				1,148	1,157	1,514	1,654	1,732	1,887	1,990	2,152	2,520	2,688	2,894	2,991	3,077	3,047	3,359
Math, (GRE AREA)	Mean	651	660	653	655	652	641	650	656	661	677	680	693	691	692	679	696	695	696
	S.D.	152	159	155	154	149	154	163	159	158	159	161	162	157	160	155	160	163	159
	N	4,323	4,972	5,363	6,520	6,532	7,371	7,351	6,428	5,634	5,101	4,555	4,089	3,731	3,736	3,342	3,424	3,109	3,287
Engineering (GRE AREA)	Mean	618	609	603	601	591	586	587	594	593	591	583	594	592	594	592	590	590	593
	S.D.	108	106	104	105	103	110	115	119	114	121	118	119	115	114	115	116	116	115
	N	6,587	6,840	7,540	8,614	7,812	9,632	9,186	6,889	6,292	5,793	5,908	5,725	6,480	7,535	8,036	7,747	7,469	7,482
Economics (GRE AREA)	Mean	623	613	619	621	614	600	581	581	588	594	596	597	605	609	601	603	613	614
	S.D.	127	115	116	118	112	111	108	115	110	111	110	113	113	109	110	106	108	108
	N	2,096	2,796	3,348	3,915	3,924	4,892	4,881	4,601	3,914	3,619	3,414	3,443	3,625	3,671	3,684	3,600	3,436	3,461
Political Science (GRE AREA)	Mean	553	546	537	524	519	501	492	479	492	489	475	473	472	471	464	456	461	461
	S.D.	96	95	91	92	90	90	90	91	85	88	89	91	91	90	90	88	85	86
	N	2,720	3,835	4,412	5,101	5,710	6,466	6,108	5,467	4,897	4,426	4,104	3,567	3,536	3,411	3,183	2,792	2,512	2,243
Sociology (GRE AREA)	Mean	546	552	535	539	518	496	469	459	474	472	462	457	451	450	436	438	427	433
	S.D.	122	119	120	120	115	118	123	126	121	120	122	119	119	113	117	109	109	106
	N	1,591	2,318	2,931	3,989	4,428	5,895	6,641	6,164	4,981	4,333	3,737	2,956	3,282	2,875	2,617	2,268	1,967	1,588
Psychology (GRE AREA)	Mean	556	552	553	547	543	532	530	528	529	530	527	531	538	529	530	534	538	532
	S.D.	91	91	93	93	89	91	92	92	92	95	95	93	95	97	97	98	97	97
	N	5,633	7,[illegible]	9,246	11,161	12,732	16,202	17,749	19,613	19,437	19,337	19,100	18,270	17,285	17,270	16,515	15,656	14,802	15,237
Education (GRE AREA)	Mean	481	474	476	478	477	462	457	446	459	452	455	454	453	452	451	449	455	456
	S.D.	86	87	90	87	84	88	92	95	93	96	93	93	93	91	89	90	90	89
	N	8,204	9,426	12,060	15,066	17,268	23,518	27,429	25,386	19,870	17,205	15,457	13,180	12,288	12,310	10,827	9,169	7,460	5,798

TABLE A-3

Test Title		1964-65	1965-66	1966-67	1967-68	1968-69	1969-70	1970-71	1971-72	1972-73
History (GRE AREA)	Mean	564	561	557	554	548	535	528	527	525
	S.D.	81	87	82	82	82	84	83	85	83
	N	4,746	7,076	7,923	9,213	9,599	11,519	11,026	9,506	7,276
Literature in English (GRE AREA)	Mean	591	588	582	572	569	556	546	544	545
	S.D.	95	94	91	91	89	90	91	96	96
	N	6,533	8,845	10,390	12,511	13,477	15,016	14,553	12,810	10,909
French (GRE AREA)	Mean	569	572	562	565	560	549	539	539	531
	S.D.	96	92	92	93	94	93	91	91	85
	N	1,102	1,535	1,850	2,217	2,486	2,431	2,497	2,038	1,793
Music (GRE AREA)	Mean	Not Offered	Fewer than 1,000 Test-Takers	518	517	511	491	481	483	494
	S.D.			103	100	98	99	98	100	95
	N			1,207	1,550	1,856	2,479	2,640	2,681	2,670

Test Title		1973-74	1974-75	1975-76	1976-77	1977-78	1978-79	1979-80	1980-81	1981-1982
History (GRE AREA)	Mean	518	513	518	510	509	505	508	502	507
	S.D.	83	83	83	82	82	82	80	77	78
	N	6,232	5,309	4,623	4,276	3,980	3,843	3,223	2,738	2,458
Literature in English (GRE AREA)	Mean	547	541	539	532	530	525	521	520	521
	S.D.	99	100	101	101	102	105	99	100	100
	N	9,821	8,351	7,035	6,948	6,420	5,948	5,557	4,923	4,085
French (GRE AREA)	Mean	527	522	519	515	514	507	Fewer than 1,000 takers, therefore not included.		
	S.D.	88	91	90	94	90	90			
	N	1,642	1,435	1,169	1,029	994	1,028			
Music (GRE AREA)	Mean	499	505	498	492	502	495	500	489	494
	S.D.	95	91	95	95	97	92	94	91	90
	N	2,783	2,831	2,865	3,339	3,597	3,496	3,156	2,769	1,954

TABLE B

Selected Background Data on GRE, LSAT, GMAT, and MCAT Test Takers, 1975-1982

(All figures are rounded percentages of the Universe used)

Universe (see code below):	19-24 AGE			Non-U.S. Citizen		English as a Second		FEMALE				WHITE				BLACK				HISPANIC				Attend Graduate School		2+Yrs Work Exper.	Coll. Grad.
	2	3	2	1	3	1	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3	1	3	2	3
	GRE	LSAT	GMAT*	GRE	GMAT	GRE	GMAT	GRE	LSAT	GMAT	MCAT	GRE	LSAT	GMAT	MCAT	GRE	LSAT	GMAT	MCAT	GRE	LSAT	GMAT	MCAT	GRE	LSAT	GMAT	MCAT
1975-1976	61	63	NA	8	NA	6	NA	NA	29	NA	27	87	**	NA	NA	7	**	NA	NA	2	**	NA	NA	22	11	NA	28
1976-1977	60	62	NA	8	NA	6	NA	NA	32	NA	27	87	**	NA	80	7	**	NA	9	3	**	NA	5	23	12	NA	34
1977-1978	57	62	52	9	20	6	19	NA	34	30	30	86	**	82	82	7	**	6	8	3	**	1	4	20	12	50	33
1978-1979	55	61	53	10	21	8	19	55	35	34	32	86	**	82	81	7	**	6	8	3	**	1	6	21	13	50	31
1979-1980	53	58	52	11	21	8	20	56	37	36	34	86	83	82	80	7	7	6	8	3	4	1	6	22	15	52	33
1980-1981	52	57	51	12	20	9	20	57	39	37	36	86	84	82	79	7	7	6	8	[illegible]	4	1	8	22	15	53	31
1981-1982	52	57	++	13	21	10	21	57	39	38	37	86	84	83	78	7	7	6	8	3	4	1	7	23	14	50	33

Universe
- #1 = All first time test takers only
- #2 = All U.S. Citizens only
- #3 = All test takers (percentages based on respondents only)

* 19-25 for the GMAT
++ impossible to determine for U.S. Citizens
** The percentage of non-respondents in these years was so high that the residual data should dot be used. According to the LSAS analysis 47.3% of the test-takers did not respond to the question in 1975-6; 40.3% in 1976-7; 42.6% in 1977-8; and 39.2% in 1978-9. The percentage of non-respondents dropped to virtually zero in 1969-80 when the position of the question on the background information form was changed, i.e. the question became part of the test-registration information as opposed to the information to be provided to law schools.

TABLE C

DEGREE OF CHANGE IN SCORES, 1964-1982 (UNLESS OTHERWISE INDICATED)

Test	Scale[1] of Reported Scores Max.	Min.	Approximate Number Questions	CHANGE in STANDARD DEVIATION UNITS: Standard Deviation =S.D. in Base Year	Standard Deviation = Mean S.D. over 18 Years	Term
GMAT (1965-82)	800	200	150	-.18	-.16	Small decline
LSAT (1968-74)[2]	800	200		+.04	N.A.	No Change
LSAT (1975-82)[2]	800	200	115	+.21	N.A.	Moderate increase
MCAT (1977-82)	15	1	325[3]			
Reading			60	-.09	N.A.	Small decline
Quantitative			60	-.21	N.A.	Moderate decline
Biology			55	+.15	N.A.	Small increase
Chemistry			70	+.06	N.A.	No change
GRE: Verbal	850[4]	210[4]	---[5]	-.49	-.48	Large decline
Quantitative	845[4]	200[4]	---[5]	.00	.00	No Change
Biology	990	260	210	-.01	-.01	No Change
Chemistry	990	440	150	-.11	-.11	Small decline
Physics	990	370	100	+.18	+.17	Small increase
Geology[6]	910	300	200	-.32	-.35	Moderate decline
Mathematics	990	420	66	+.30	+.28	Moderate increase
Engineering	990	320	150	-.23	-.22	Moderate decline
Economics	990	400	160	-.07	-.08	No Change
Polit. Sci.	850	250	170	-.96	-1.02	Extreme decline
Psychology	940	270	200	-.26	-.26	Moderate decline
Sociology	990	210	200	-.93	-.96	Extreme decline
Education	810	220	200	-.29	-.28	Moderate decline
History	870	330	190	-.70	-.70	Large decline
English Lit.	810	250	230	-.74	-.72	Large decline
French[7]	810	290	190	-.65	-.68	Large decline
Music[8]	820	270	200	-.23	-.25	Moderate decline

NOTES:
1. Scales in use since 1976 (with exceptions as noted)
2. LSAT data is broken into two periods throughout this paper, corresponding to the two different data sets upon which we drew. See the discussion in the text. Computations for the first period use 1968-69 as the base year, as Standard Deviations for prior years are unavailable.
3. There are 6 sections of the MCATs (we use only 4 here). Some of the questions are experimental and not counted in scoring; others are scored in more than one section
4. Average of 7 forms of the GRE General Examination used between 1973 and 1977.
5. The number of questions changed in 1976 with the introduction of the GRE/Analytic.
6. Base year is 1967-8. Prior to then, there were fewer than 1,000 candidates annually
7. Terminal year is 1978-9. Since then, there have been less than 1,000 candidates annually.
8. Base year is 1966-7. Prior to then, there were fewer than 1,000 candidates annually

TABLE D

Summary of Test Score Changes by Standard Deviation Units*

Range	Descriptive Term and Tests	Change
+.40 and above	Large Increases: NONE	
+.20 to +.39	Moderate Increases:	
	Mathematics (GRE Area Test)	+.28
	LSAT (1975-82)	+.21
+.10 to +.19	Small Increases:	
	Physics (GRE)	+.17
	Biology (MCAT Sub-test; 1977-1982)	+.15
-.09 to +.09	NO CHANGE:	
	Chemistry (MCAT Sub-test; 1977-1982)	+.07
	LSAT (1968-74)	+.04
	GRE Quantitative	.00
	Biology (GRE)	-.01
	Economics (GRE)	-.08
-.10 to -.19	Small Decline:	
	Reading (MCAT Sub-test, 1977-1982)	-.10
	Chemistry (GRE)	-.11
	GMAT	-.16
-.20 to -.39	Moderate Decline:	
	MCAT Quantitative (1977-1982)	-.22
	Engineering (GRE)	-.22
	Music (GRE)	-.25
	Psychology (GRE)	-.26
	Education (GRE)	-.28
	Geology (GRE)	-.35
-.40 to -.74	Large Decline:	
	GRE Verbal	-.48
	French (GRE)	-.68
	History (GRE)	-.70
	English Lit. (GRE)	-.72
-.75 and below	Extreme Decline:	
	Sociology	-.96
	Political Science	-1.02

*All test-takers, including non-U.S. citizens.

SUMMARY OF TEST SCORE CHANGES BY STANDARD DEVIATION UNITS

TABLE D-2

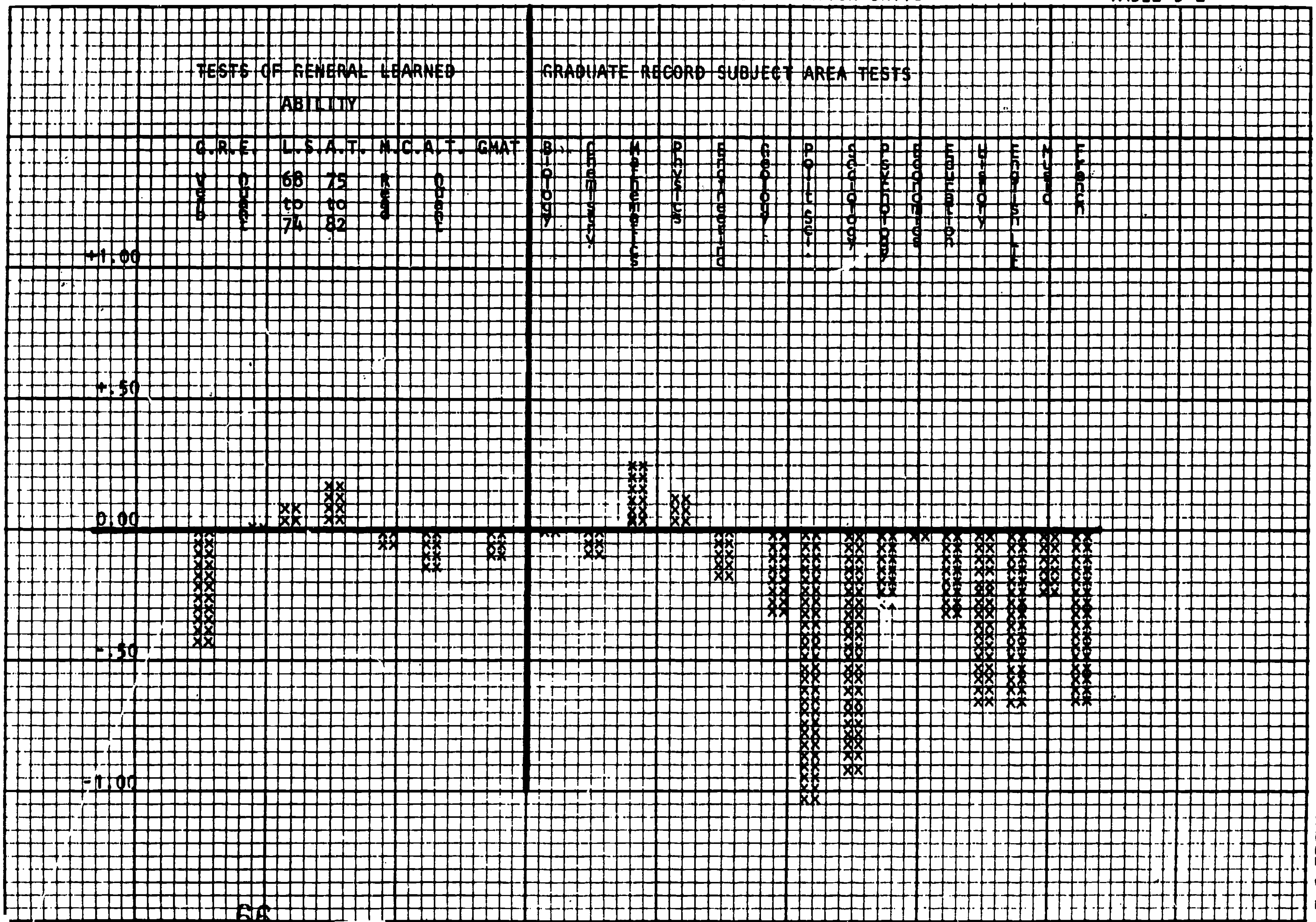


TABLE D-2

Table E TURNING POINTS IN GRADUATE RECORD TEST SCORE TRENDS, 1964-1982 **Table E**

Abbreviations:

M = Mean Score
S.D. = Standard Deviation
SDU = Change in terms of Stand. Dev. Units from previous reference year
N = Number of test-takers

Test		Base Years: 1964/5	Trend 1965-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76	1976-77	1977-78	Trend 1978-1981	Terminal Year: 1981/2	Ave. S.D.	Net Pt. Diff.	S.D.U. Change
GRE/V	M	530						492						469		-61	
	S.D.	124	Down					126					Stable	130	128		
	SDU	---						(.31)						(.18)			(.48)
	N	93.8k						301.1k						256.4k			
GRE/Q	M	533				508								533		0	
	S.D.	137	Stable			136							Up	137	137		
	SDU	---				(.18)								.18			.00
	N	93.8k				293.5k								256.4k			
Biol.	M	617			603					627				616		-1	
	S.D.	117	Down		114					112			Stable	114	116		
	SDU	---			(.12)					.21				(.10)			(.01)
	N	6.2k			12.9k					20.4k				14.1k			
Chem.	M	628		613				634						616		-12	
	S.D.	114	Down	113				115					Stable	105	111		
	SDU	---		(.13)				.19						(.16)			(.11)
	N	3.8k		4.8k				4.6k						4.9k			
Math.	M	651		641						693				696		+45	
	S.D.	152	Down	154						162			Up	159	158		
	SDU	---		(.07)						.34				.02			.28
	N	4.3k		7.4k						4.1k				3.3k			
Physics	M	623									656			647		+24	
	S.D.	134	Up								144		Stable	145	143		
	SDU	---									.25			(.06)			.17
	N	4.0k									3.0k			3.3k			
Engin.	M	618			587									593		-25	
	S.D.	108	Down		115								Stable	115	112		
	SDU	---			(.29)									.05			(.22)
	N	6.6k			9.2k									7.5k			

TABLE E-1

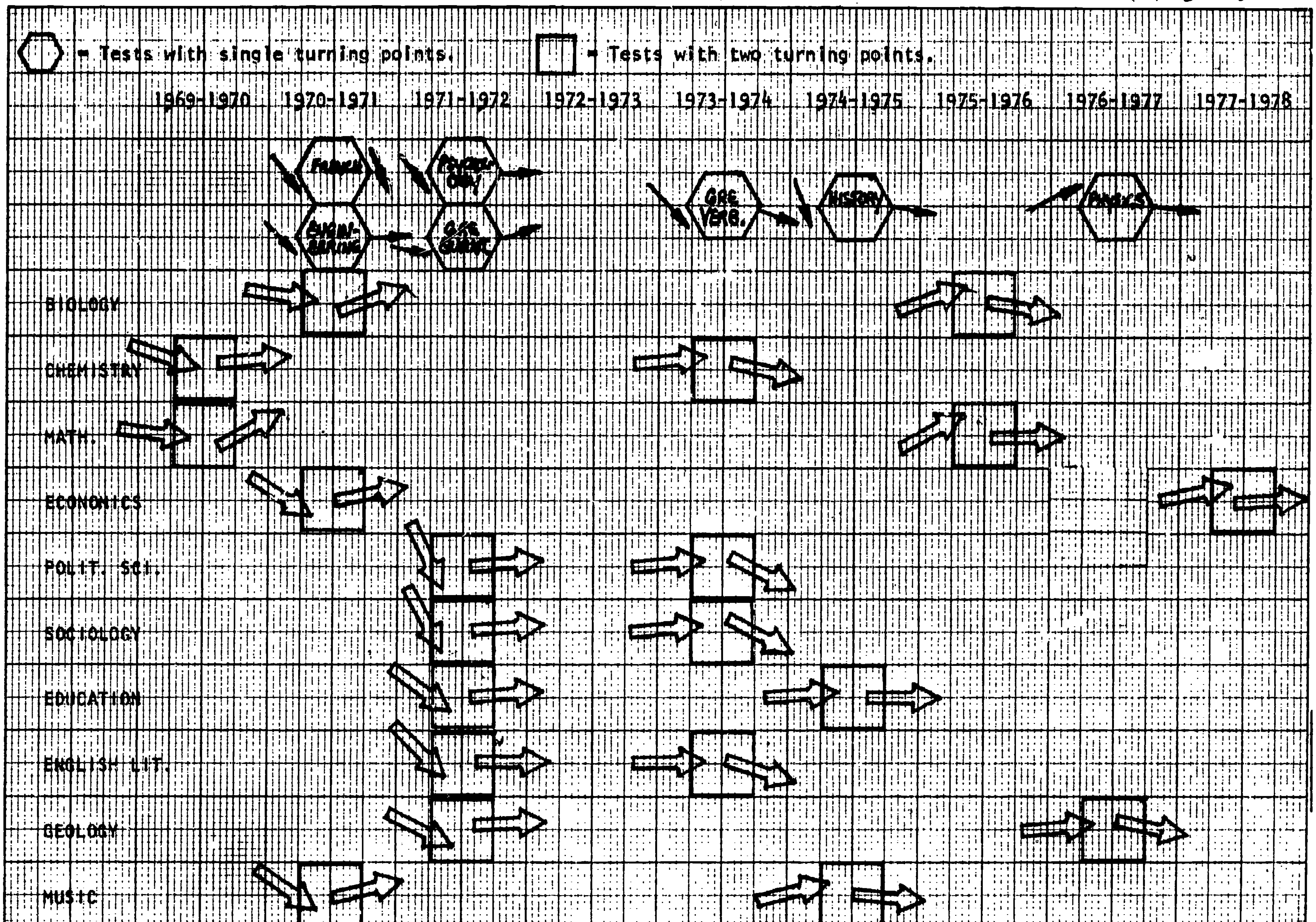
Test		Base Year: 1964/5	Trend 1965-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76	1976-77	1977-78	Trend 1978-81	Terminal Year 1981/2	Ave. S.D.	Net Pt. Diff.	S.D.U. Change
Econ.	M	623			581							609		614		-9	
	S.D.	127	Down		108							109	Up	108	112		
	SDU	---			(.33)							.26		.05			(.08)
	N	8.1k			4.9k							3.7k		3.5k			
Pol. Sci.	M	553				479		489						461		-92	
	S.D.	96	Down			91		88					Down	86	90		
	SDU	---				(.77)		.11						(.32)			(1.02)
	N	2.7k				5.5k		4.4k						2.2k			
Sociology	M	546				459		472						433		-113	
	S.D.	122	Down			126		120					Down	106	122		
	SDU	---				(.71)		.10						(.33)			(.96)
	N	1.6k				6.2k		4.3k						1.6k			
Psych.	M	556				528								532		-24	
	S.D.	91	Down			92							Stable	97	94		
	SDU	---				(.31)								.04			(.26)
	N	5.6k				19.6k								15.2k			
Education	M	481				446			455					456		-25	
	S.D.	86	Stable			95			93				Stable	89	90		
	SDU	---				(.41)			.09					.01			(.28)
	N	6.2k				23.4k			15.5k					5.6k			
History	M	564							513					507		-57	
	S.D.	81	Down						83				Stable	78	82		
	SDU	---							(.63)					(.07)			(.70)
	N	4.7k							5.3k					2.5k			
English Lit.	M	591				544		547						521		-70	
	S.D.	95	Down			96		99					Stable	100	97		
	SDU	---				(.49)		.03						(.26)			(.72)
	N	6.5k				12.8k		9.8k						4.9k			

TABLE E-2

Test		Base Year:	Trend 1965-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76	1976-77	1977-78	Trend 1978-1981	Terminal Year: 1981/2	Ave. S.D.	Net Pt. Diff.	S.D.U. Change
Biology		1967/8															
	M	603	Not			569					581			570		-33	
	S.D.	104	Appli-			92					92		Stable	86	94		
	SDU	---	cable			(.33)					.13			(.12)			(.35)
	N	1.1k				11.7k					2.7k			3.4k			
French		1964/5															
	M	569	Down		539								507	1978/9		-62	
	S.D.	96			91								90	is	91		
	SDU	---			(.31)								(.35)	term.			(.68)
	N	1.1k			2.5k								1.0k	year			
Music		1966/7															
	M	518	Not		481				505					494		-24	
	S.D.	103	Appli-		98				91				Stable	90	96		
	SDU	---	cable		(.36)				.24					(.12)			(.25)
	N	1.2k			2.6k				2.8k					2.0k			

TABLE E-4

TURNING POINT YEARS IN GRADUATE RECORD EXAMINATIONS SCORES (change measured in Standard Deviation Units), 1964-1982



Arrows to left of boxes represent trend prior to turning point year; arrows to right represent subsequent trend.

TEST PERFORMANCE BY UNDERGRADUATE MAJOR, 1977-1982

This set of tables demonstrates the percentages by which the mean score of test-takers who majored as undergraduates in specific fields differed from the mean scores of all test-takers who indicated their undergraduate major on the background questionnaires administered by the LSAT, GMAT and GRE General examinations.

Since the background characteristics (including undergraduate major) of GMAT test-takers are not available prior to 1977, I can present only five (5) recent years of data in these tables.

The data is presented in two ways: (1) by year (pages F-4 to F-8)and (2) by major, within general curricular groupings (pages F9 to F 12).

Here is a guide to reading the tables:

The dimensions of the universe of test-takers on which this analysis focuses are indicated at the top of the page for each year (F-4 to F-8).

Total Test-Takers: LSAT, as reported in the NIE-sponsored analysis, and used with the permission of the Law School Admissions Council;
GMAT, as reported in the comprehensive tables provided to the author, copyright by the Graduate Management Admissions Council and used with its permission;
GRE, as reported on a set of tables provided by ETS and entitled, "GRE Candidate Volume, Means, and Standard Deviations," covering the years, 1964-1983. The total number of test-takers on these tables differs from that reported in the annual GRE Summary.

Subject Universe: The objective was to winnow out non-U.S. citizens, people who had scores reported in a given year but who did not take the test that particular year, etc.

LSAT: tends to be the full number reported in "Total Test-Takers" (the LSAT does not ask a question about citizenship);
GMAT: those who responded to the background question on citizenship by indicating that they were U.S. citiznes;
GRE: all first time test-takers. Mean scores for U.S. citizens only would be slightly higher on the GRE/V and slightly lower on the GRE/Q. It is possible to disaggregate scores for U.S. citizens x first-time test-takers for three out of the five years reported.

Respondents: The number of individuals in the subject universe who identified their undergraduate major field.

% Repondents/Universe: Self-evident. Why the LSAT percentage is significantly lower than the others is a mystery.

Mean and Standard Deviation: The Mean and Standard Deviation are indicated for the subject universe only, and thus differ from the Mean and Standard Deviation for all test-takers.

Analysis by Major. No two of these examinations ask the question, "What is your undergraduate major?" the same way. However, it was possible to create 30 standard categories, with the criterion that to be included in the analysis, each undergraduate major category had to account for 0.5% or more of the respondents.

Depending on what categories were used on the background questionnaires for the individual examinations, aggregation was often necessary. For example, the LSAT lists one category for "Fine Arts" and another for "Music." The GMAT uses only "Fine Arts" (and one assumes that a Music major will check that box). The GRE reports both for individual majors in the category of "Fine Arts" and for the entire category. To render the data comparable across the three tests, I aggregated the individual majors employed on the LSAT questionnaire, and used the GRE reporting for the entire category of "Arts."

Aggregating such data for purposes of comparability across the tests was necessary in the following cases:

LSATs: Fine Arts/Music, Foreign Languages, Other Humanities, Other Social Sciences, Other Sciences, Engineering, and Other Business.
GMATs: None
GREs : Foreign Languages, Other Humanities, Other Social Sciences, and Other Sciences.

What specific majors are included in the "Other" categories? The references below are to the categories used on the background questionnaires for the LSATs and GREs:

"Other Humanities":	LSATs:	Religion, Archaeology, Other Humanities
	GREs :	Religion, Archaeology, Other Humanities, Comparative Literature, Art History, Linguistics
"Other Social Sci.":	LSATs:	Geography, American Civilization, Other Social Sciences
	GREs :	Geography, American Studies, Other Social Sciences, Communication, International Relations, Social Psychology, Urban Development
"Other Sciences" :	LSATs:	Geology, Astronomy, Other Sciences
	GREs :	Geology, Astronomy, Other Physical Science, Applied Mathematics, Statistics, Oceanography

I am sure there will be quarrels with my categories and aggregations here. And I confess to being uneasy with a few of them. For example, I have aggregated "American Civilization" and "American Studies" majors with the "Other Social Science" category, even though the test-performance of American Civ majors is far superior to that of the other majors in the category. Likewise, Music majors tend to outperform other "Fine Arts" majors on the examinations, but I nonetheless lumped all Fine Arts majors together to create a cell of significant magnitude. Too, some might quarrel with what I have left out. For example, the GMATs have a category for "Statistics" (which I aggregated

with "Other Sciences" on the GREs). But the N in that category on the GMATs was consistently so small that I did not think it was worth aggregating.

ABBREVIATIONS:

N.Q. = Does not qualify for inclusion because the number of test-takers with that particular major was less than 0.5% of the subject universe in that year.

--- = Test data do not report separately for the major field indicated.

N.I. = Not indicated for the Subject Universe.

Test Performance by Undergraduate Major — **Table F-4**

Year: 1977-1978

Test	LSAT	GMAT	VER	GRE	QUAN	TOT. N
Total Test-Takers	111,555	169,908		286,383		567,846
Subject Universe	111,555	132,385		221,745		465,685
Respondents	85,422	131,778		217,977		435,177
% Responds./Universe	76.6%	99.5%		98.3%		93.4%
Mean (Universe)	538	473	494		517	
Stand. Deviation	110	103	N.I.		N.I.	

% by which Mean for Major is Above (Below) the Mean for the Universe

Major	LSAT	GMAT	GRE VER	GRE QUAN	402,286	*% of Respond. 92.4%
English	5.4%	6.8%	14.4%	(6.0%)	21,873	5.0%
Philosophy	8.9	11.2	18.4	5.4	4,255	1.0
Arts/Music	1.1	0.6	1.4	(7.7)	11,530	2.6
Foreign Langs.	5.0	4.0	10.5	(3.7)	8,925	2.1
Other Humanities	5.8	4.9	9.4	(2.3)	8,839	2.0
History	2.2	3.4	8.9	(5.4)	21,293	4.9
Economics	9.7	6.3	1.8	12.8	17,990	4.1
Government	3.2	6.6				
Political Science	(2.0)	1.5	3.8	(4.3)	30,460	7.0
Psychology	2.4	0.0	3.8	(1.7)	30,039	6.9
Sociology	(5.8)	(4.9)	(3.8)	(12.4)	12,370	2.8
Anthropology	7.6	---	15.2	0.0	2,719	0.6
Other Social Sci.	0.2	0.8	(0.2)	(5.0)	19,753	4.5
Biol/Bioscience	4.3	2.3	3.9	9.5	29,341	6.7
Chemistry	8.0	8.5	3.8	21.5	7,701	1.8
Mathematics	13.8	12.3	3.0	29.8	9,511	2.2
Physics	N.Q.	N.Q.	8.9	34.0	3,563	0.8
Other Science	5.9	2.1	2.9	17.4	7,911	1.8
Computer Science	---	6.8	1.2	24.6	2,907	0.7
Engineering	8.7	9.5	(6.5)	27.3	25,277	5.8
Accounting	3.5	(2.3)				
Finance	3.0	(2.1)				
Marketing	---	(7.8)				
Business Admin.	(3.7)	---	(11.1)	(2.1)	82,599	19.0
Management/Ind. Man.	(1.7)	(7.8)				
Other Business	(0.4)	(6.1)				
Education	(7.1)	(6.1)	(12.3)	(15.3)	34,128	7.8
Journalism	2.6	---	5.3	(7.5)	2,852	0.7
Social Work	(8.7)	---	(9.7)	(17.0)	3,404	0.8
Speech	(1.9)	---	(5.9)	(12.8)	3,046	0.7

*Due to founding, percentages for individual majors may not add up to the total indicated.

Test Performance by Undergraduate Major

Table F-5

Year: 1978-1979

Test	LSAT	GMAT	VER	GRE	QUAN	TOT. N
Total Test-Takers	98,307	187,039		282,482		567,828
Subject Universe	98,307	144,089		218,682		461,076
Respondents	74,343	143,354		213,843		431,540
% Responds./Universe	75.6%	99.5%		97.8%		93.6%
Mean (Universe)	536	475	489		515	
Stand. Deviation	107	104	123		131	

% by which Mean for Major is Above (Below) the Mean for the Universe

Major	LSAT	GMAT	VER	QUAN	TOT. N	*% of Respond.
					378,699	88.1%
English	4.7%	6.9%	15.5%	(5.6%)	20,266	4.7%
Philosophy	8.6	9.9	19.4	5.2	3,858	0.9
Arts/Music	0.1	(0.2)	2.2	(7.2)	11,154	2.6
Foreign Langs.	5.6	4.6	7.4	(3.9)	9,232	2.1
Other Humanities	6.2	4.6	9.2	(3.5)	8,507	2.0
History	1.9	2.9	10.6	(5.0)	19,240	4.5
Economics	9.5	5.9	3.7	13.6	18,025	4.2
Government	3.2	5.3				
Political Science	(1.9)	1.1	4.9	(3.5)	28,532	6.6
Psychology	2.1	(0.2)	4.3	(1.9)	22,779	5.3
Sociology	(6.7)	(5.3)	(3.3)	(12.6)	11,089	2.6
Anthropology	6.3	---	6.3	(4.3)	3,304	0.8
Other Social Sci.	(0.2)	0.6	(1.2)	(4.7)	20,212	4.7
Biol/Bioscience	4.1	2.5	0.6	7.8	30,650	7.1
Chemistry	7.8	8.4	5.1	21.9	7,930	1.8
Mathematics	12.9	14.1	4.1	29.7	8,353	1.9
Physics	N.Q.	N.Q.	11.5	34.2	3,294	0.8
Other Science	4.7	1.3	3.7	16.9	8,479	2.0
Computer Science	---	7.4	1.8	26.8	2,969	0.7
Engineering	8.9	9.7	(4.3)	27.0	26,137	6.1
Accounting	4.3	(2.5)				
Finance	3.4	(1.7)				
Marketing	---	(7.8)				
Business Admin.	(3.2)	---	(9.4)	(1.7)	71,883	16.7
Management/Ind. Man.	(2.4)	(7.4)				
Other Business	(0.5)	(5.5)				
Education	(6.7)	(6.3)	(11.2)	(14.6)	33,179	7.7
Journalism	2.1	---	7.0	(6.4)	2,572	0.6
Social Work	(7.2)	---	(9.0)	(17.1)	3,346	0.8
Speech	(0.9)	---	(9.4)	(10.5)	3,709	0.9

*Due to rounding, percentages for individual majors may not add up to the total indicated.

Test Performance by Undergraduate Major **Table F-6**

Year: 1979-1980

Test	LSAT	GMAT	VER	GRE	QUAN	TOT. N
Total Test-Takers	94,583	209,739		272,281		576,603
Subject Universe	94,583	156,548		210,749		461,880
Respondents	74,365	155,380		207,713		436,918
% Responds./Universe	78.6%	99.3%		98.3%		94.6%
Mean (Universe)	539	473	488		516	
Stand. Deviation	108	102	123		131	

Major	% by which Mean for Major is Above (Below) the Mean for the Universe: LSAT	GMAT	GRE VER	GRE QUAN	TOT. N	*% of Respond.
Major					387,766	88.8%
English	4.5%	6.8%	15.0%	(5.6%)	20,405	4.7%
Philosophy	7.2	10.8	19.5	5.0	3,734	0.9
Arts/Music	0.2	0.0	1.6	(7.4)	11,397	2.6
Foreign Langs.	4.3	4.2	10.7	(3.3)	8,114	1.9
Other Humanities	5.8	4.2	9.2	(5.0)	8,921	2.0
History	1.7	4.0	20.5	(5.0)	17,935	4.1
Economics	8.0	6.3	0.8	13.0	18,589	4.3
Government	2.4	4.4 }				
Political Science	(2.8)	1.1 }	3.9	(4.5)	27,701	6.3
Psychology	0.7	1.1	3.3	(3.1)	27,792	6.4
Sociology	(7.2)	(4.4)	(4.3)	(13.0)	10,790	2.5
Anthropology	3.5	—	16.0	(0.4)	2,319	0.5
Other Social Sci.	(1.9)	0.8	0.0	(5.4)	19,946	4.6
Biol/Bioscience	3.5	3.0	4.7	9.7	27,078	6.2
Chemistry	6.7	8.0	2.9	20.5	7,582	1.7
Mathematics	12.6	14.0	3.3	22.6	8,182	1.9
Physics	N.Q.	N.Q.	9.0	32.6	3,429	0.8
Other Science	2.8	0.8	2.9	17.1	9,353	2.1
Computer Science	—	5.5	(1.0)	25.0	3,696	0.8
Engineering	7.8	10.6	(6.6)	27.3	28,763	6.6
Accounting	3.2	(1.9) }				
Finance	2.3	(0.8) }				
Marketing	—	(7.6) }				
Business Admin.	(4.6)	— }	(10.5)	(1.2)	79,088	18.1
Management/Ind. Man.	(3.9)	(7.8) }				
Other Business	(0.9)	(5.1) }				
Education	(8.0)	(6.3)	(11.7)	(15.1)	33,947	7.8
Journalism	0.6	—	5.1	(8.3)	2,761	0.6
Social Work	(9.5)	—	(10.0)	(18.6)	3,574	0.8
Speech	(3.9)	—	(5.5)	(13.0)	2,670	0.6

*Due to rounding, percentages for individual majors may not add up to the total indicated.

Test Performance by Undergraduate Major   Table F-7

Year: 1980-1981

Test	LSAT	GMAT	VER	GRE	QUAN	TOT. N
Total Test-Takers	100,793	214,555		262,855		578,203
Subject Universe	100,793	156,591		203,131		460,515
Respondents	85,057	155,030		198,768		438,855
% Responds./Universe	84.4%	99.0%		97.9%		95.3%
Mean (Universe)	544	478	486		520	
Stand. Deviation	110	99	122		132	

% by which Mean for Major is Above (Below) the Mean for the Universe

Major	LSAT	GMAT	VER	QUAN	386,592	*% of Respond. 88.3%
English	5.1%	6.7%	14.6%	(6.2%)	20,071	4.6%
Philosophy	8.8	11.4	17.9	3.3	3,915	0.9
Arts/Music	(0.4)	(0.8)	1.4	(8.3)	11,008	2.5
Foreign Langs.	4.7	4.2	9.5	(3.3)	7,836	1.8
Other Humanities	5.5	3.8	8.0	(4.8)	8,940	2.0
History	2.9	4.0	10.3	(5.6)	17,504	4.0
Economics	9.7	6.9	0.6	12.7	19,150	4.4
Government	3.5	4.0 }				
Political Science	(2.0)	1.0 }	4.1	(4.0)	29,164	6.6
Psychology	0.2	0.8	3.3	(3.7)	27,692	6.3
Sociology	(7.2)	(5.2)	(3.3)	(13.1)	8,297	1.9
Anthropology	3.5	—	14.8	(2.3)	2,334	0.5
Other Social Sci.	(0.9)	0.2	(0.2)	(6.2)	20,576	4.7
Biol/Bioscience	3.9	3.3	4.7	8.8	24,654	5.6
Chemistry	8.1	7.9	3.5	20.2	7,433	1.7
Mathematics	12.7	13.0	3.1	27.9	7,525	1.7
Physics	N.Q.	N.Q.	7.8	31.3	3,441	0.8
Other Science	3.2	0.4	3.5	16.0	10,222	2.4
Computer Science	—	5.0	(0.6)	24.0	4,244	1.0
Engineering	8.3	9.8	(7.6)	26.2	30,405	6.9
Accounting	2.9	(1.7) }				
Finance	3.1	(0.4) }				
Marketing	—	(7.7) }				
Business Admin.	(4.2)	— }	(9.7)	(1.0)	82,610	18.8
Management/Ind. Man.	(3.9)	(7.5) }				
Other Business	(1.3)	(4.8) }				
Education	(8.3)	(6.1)	(10.7)	(15.4)	31,351	7.1
Journalism	0.9	—	4.3	(9.0)	3,088	0.7
Social Work	(8.5)	—	(9.5)	(19.2)	3,629	0.8
Speech	(3.1)	—	(6.0)	(14.4)	2,533	0.6

*Due to rounding, percentages for individual majors may not add up to the total indicated.

Test Performance by Undergraduate Major — **Table F-8**

Year: 1981-1982

Test	LSAT	GMAT	VER	GRE	QUAN	TOT. N
Total Test-Takers	99,928	203,304		256,381		558,613
Subject Universe	99,928	139,964		180,798		420,690
Respondents	85,198	138,946		174,624		398,768
% Responds./Universe	85.3%	99.3%		99.6%		94.8%
Mean (Universe)	553	482	482		525	
Stand. Deviation	110	97	123		134	

% by which Mean for Major is Above (Below) the Mean for the Universe

Major	LSAT	GMAT	VER	QUAN	TOT. N	*% of Respond.
					353,680	88.8%
English	5.6%	4.1%	14.5%	(5.7%)	17,757	4.4%
Philosophy	8.7	11.0	17.6	4.6	3,410	0.9
Arts/Music	(0.5)	(1.2)	1.7	(8.4)	9,670	2.4
Foreign Langs.	5.7	3.3	7.9	(4.2)	7,068	1.8
Other Humanities	4.7	1.8	7.3	(5.0)	8,341	2.1
History	2.9	4.6	10.8	(5.5)	15,123	3.8
Economics	9.6	7.3	0.8	12.4	17,562	4.4
Government	3.3	4.6				
Political Science	(1.6)	0.6	3.5	(5.0)	27,337	6.9
Psychology	0.9	0.8	3.1	(4.0)	24,885	6.2
Sociology	(7.0)	(5.0)	(5.0)	(15.0)	8,693	2.2
Anthropology	4.0	—	16.4	(1.7)	1,863	0.5
Other Social Sci.	(0.9)	0.3	(0.4)	(7.2)	20,048	5.0
Biol/Bioscience	4.0	3.3	5.4	8.0	22,820	5.7
Chemistry	7.6	7.5	2.1	18.3	6,867	1.7
Mathematics	12.8	13.3	2.7	26.3	6,564	1.6
Physics	N.Q.	N.Q.	6.6	29.5	3,183	0.8
Other Science	2.8	0.8	3.5	14.5	9,154	2.3
Computer Science	—	5.4	(1.5)	22.9	5,035	1.3
Engineering	8.0	10.0	(7.3)	25.1	29,718	7.5
Accounting	3.4	(1.5)				
Finance	3.4	(0.8)				
Marketing	—	(8.1)				
Business Admin.	(4.5)	—	(9.1)	(2.3)	77,679	19.5
Management/Ind. Man.	(5.4)	(7.7)				
Other Business	(0.9)	(5.0)				
Education	(8.7)	(4.2)	(10.4)	(15.8)	22,978	5.8
Journalism	0.7	—	5.7	(8.6)	2,767	0.7
Social Work	(10.1)	—	(9.1)	(20.8)	2,999	0.8
Speech	(2.7)	—	(6.0)	(14.3)	2,159	0.5

*Due to rounding, percentages for individual majors may not add up to the total indicated.

General Area: Humanities

Table indicates the percentage by which the Mean for the major is above (below) the Mean for the universe.

Major & Test	1977-78	1978-79	1979-80	1980-81	1981-82
English:					
LSAT	5.4%	4.7%	4.5%	5.1%	5.6%
GMAT	6.8	6.9	6.8	6.7	4.1
GRE/Verbal	14.4	15.5	15.0	14.6	14.5
GRE/Quant.	(6.0)	(5.6)	(5.6)	(6.2)	(5.7)
% of Respondents:	5.0%	4.7%	4.7%	4.6%	4.4%
Philosophy:					
LSAT	8.9%	8.6%	7.2%	8.8%	8.7%
GMAT	11.2	9.9	10.8	11.4	11.0
GRE/Verbal	18.4	19.4	19.5	17.9	17.6
GRE/Quant.	5.4	5.2	5.0	3.3	4.6
% of Respondents:	1.0%	0.9%	0.9%	0.9%	0.9%
Foreign Languages:					
LSAT	5.0%	5.6%	4.3%	4.7%	5.7%
GMAT	4.0	4.6	4.2	4.1	3.3
GRE/Verbal	10.5	7.4	10.7	9.5	7.9
GRE/Quant.	(3.7)	(3.9)	(3.3)	(3.3)	(4.2)
% of Respondents:	2.1%	2.1%	1.9%	1.8%	1.8%
History:					
LSAT	2.2%	1.9%	1.7%	2.9%	2.9%
GMAT	3.4	2.9	4.0	4.0	4.6
GRE/Verbal	8.9	10.6	10.5	10.3	10.8
GRE/Quant.	(5.4)	(5.0)	(5.0)	(5.6)	(5.5)
% of Respondents:	4.9%	4.5%	4.1%	4.0%	3.8%
Other Humanities:					
LSAT	5.8%	6.2%	5.8%	5.5%	4.7%
GMAT	4.9	4.6	4.2	3.7	1.8
GRE/Verbal	9.4	9.2	9.2	8.0	7.3
GRE/Quant.	(2.3)	(3.5)	(5.0)	(4.8)	(5.0)
% of Respondents:	2.0%	2.0%	2.0%	2.0%	2.1%

General Area: Social Sciences

Table indicates the percentage by which the Mean for the major is above (below) the Mean for the universe.

Major & Test	1977-78	1978-79	1979-80	1980-81	1981-82
Economics:					
LSAT	9.7%	9.5%	8.0%	9.7%	9.6%
GMAT	6.3	5.9	6.3	6.9	7.3
GRE/Verbal	1.8	3.7	0.8	0.6	0.8
GRE/Quant.	12.8	13.6	13.0	12.7	12.4
% of Respondents:	4.1%	4.2%	4.3%	4.4%	4.4%
Psychology:					
LSAT	2.4%	2.1%	0.7%	0.2%	0.9%
GMAT	0.0	(0.2)	1.1	0.8	0.8
GRE/Verbal	3.8	4.3	3.3	3.3	3.1
GRE/Quant.	(1.7)	(1.9)	(3.1)	(3.7)	(4.0)
% of Respondents:	6.9%	5.3%	6.4%	6.3%	6.2%
Sociology:					
LSAT	(5.8%)	(6.7%)	(7.2%)	(7.2%)	(7.0%)
GMAT	(4.9)	(5.3)	(4.4)	(5.2)	(5.0)
GRE/Verbal	(3.8)	(3.3)	(4.3)	(3.3)	(5.0)
GRE/Quant.	(12.4)	(12.6)	(13.0)	(13.1)	(15.0)
% of Respondents:	2.8%	2.6%	2.5%	1.9%	2.2%
Political Science[1]:					
LSAT	(2.0%)	(1.9%)	(2.8%)	(2.0%)	(1.6%)
GMAT	1.5	1.1	1.1	1.0	0.6
GRE/Verbal	3.8	4.9	3.9	4.1	3.5
GRE/Quant.	(4.3)	(3.5)	(4.5)	(4.0)	(5.0)
% of Respondents[2]	7.0%	6.6%	6.3%	6.6%	6.9%
Other Social Science:					
LSAT	0.2%	(0.2%)	(1.9%)	(0.9%)	(0.9%)
GMAT	0.8	0.6	0.8	0.2	0.3
GRE/Verbal	(0.2)	(1.2)	0.0	(0.2)	(0.4)
GRE/Quant.	(5.0)	(4.7)	(5.4)	(6.2)	(7.2)
% of Respondents:	4.5%	4.7%	4.0%	4.7%	5.0%

NOTES: 1. Both the LSAT and GMAT distinguish between Political Science and Government. The GRE does not. The data here apply only to Political Science.

2. The percentage represents the total number of respondents indicating an undergraduate major in either Political Science or Government (even though the test data does not include Government Majors on the LSAT and GMAT).

General Area: Science and Mathematics

Table indicates the percentage by which the Mean for the major is above (below) the Mean for the universe.

Major & Test	1977-78	1978-79	1979-80	1980-81	1981-82
Biology/Biosci.:					
LSAT	4.3%	4.1%	3.5	3.9%	4.0%
GMAT	2.3	2.5	3.0	3.3	3.3
GRE/Verbal	3.9	0.6	4.7	4.7	5.4
GRE/Quant.	9.5	7.8	9.7	8.8	8.0
% of Respondents:	6.7%	7.1%	6.2%	5.6%	5.7%
Chemistry:					
LSAT	8.0%	7.8%	6.7%	8.1%	7.6%
GMAT	8.5	8.4	8.0	7.9	7.5
GRE/Verbal	3.8	5.1	2.9	3.5	2.1
GRE/Quant.	21.5	21.9	20.5	20.2	18.3
% of Respondents:	1.8%	1.8%	1.7%	1.7%	1.7%
Mathematics:					
LSAT	13.8%	12.9%	12.6%	12.7%	12.8%
GMAT	12.3	14.1	14.0	13.0	13.3
GRE/Verbal	3.0	4.1	3.3	3.1	2.7
GRE/Quant.	29.8	29.7	22.6	27.9	26.3
% of Respondents:	2.2%	1.9%	1.9%	1.7%	1.6%
Engineering[1]					
LSAT	8.7%	8.9%	7.8%	8.3%	8.0%
GMAT	9.5	9.7	10.6	9.8	10.0
GRE/Verbal	(6.5)	(4.3)	(6.6)	(7.6)	(7.3)
GRE/Quant.	27.3	27.0	27.3	26.2	25.1
% of Respondents:	5.8%	6.1%	6.6%	6.9%	7.5%
Other Sciences:					
LSAT	5.9%	4.7%	2.8%	3.2%	2.8%
GMAT	2.1	1.3	0.8	0.4	0.8
GRE/Verbal	2.9	3.7	2.9	3.5	3.5
GRE/Quant.	17.4	16.9	17.1	16.0	14.[illegible]
% of Respondents:	1.8%	2.0%	2.1%	2.4%	2.3%

NOTES: 1. One might have placed Engineering in the "Professional" category, but the patterns of test performance for Engineering majors were closer to those of Science and Math majors than to those of other undergraduate professional program majors.

General Area: Professional

Table indicates the percentage by which the Mean for the major is above (below) the Mean for the universe.

Major & Test	1977-78	1978-79	1979-80	1980-81	1981-82
Journalism:					
LSAT	2.6%	2.1%	0.6%	0.9%	0.7%
GMAT	---	---	---	---	---
GRE/Verbal	5.3	7.0	5.1	4.3	5.7
GRE/Quant.	(7.5)	(6.4)	(8.3)	(9.0)	(8.6)
% of Respondents:	0.7%	0.6%	0.6%	0.7%	0.7%
Social Work:					
LSAT	(8.7%)	(7.2%)	(9.5%)	(8.5%)	(10.1%)
GMAT	---	---	---	---	---
GRE/Verbal	(9.7)	(9.0)	(10.0)	(9.5)	(9.1)
GRE/Quant.	(15.3)	(14.6)	(15.1)	(15.4)	(15.8)
% of Respondents:	0.8%	0.8%	0.8%	0.8%	0.8%
Education					
LSAT	(7.1%)	(6.7%)	(8.0%)	(8.3%)	(8.7%)
GMAT	(6.1)	(6.1)	(6.3)	(6.3)	(4.2)
GRE/Verbal	(12.3)	(11.2)	(11.7)	(10.7)	(10.4)
GRE/Quant.	(15.3)	(14.6)	(15.1)	(15.4)	(15.8)
% of Respondents:	7.8%	7.7%	7.8%	7.1%	5.8%
Business					
LSAT	(3.7%)	(3.2%)	(4.6%)	(4.2%)	(4.5%)
GMAT[1]	(7.8)	(7.4)	(7.8)	(7.5)	(7.7)
GRE/Verbal[2]	(11.1)	(11.2)	(11.7)	(9.7)	(9.1)
GRE/Quant.[2]	(2.1)	(1.7)	(1.2)	(1.0)	(2.3)
% of Respondents:[3]	19.0%	16.7%	18.1%	18.8%	19.5%

NOTES 1. The GMAT does not report a separate category for Business Administration, rather disaggregates the field into eight (8) sub-categories. Of these, I have substituted "Management" for "Business Administration in this table, and on the grounds that (a) the term is often substituted for "Business Administration" as the title of the major in American colleges, and (b) after Accounting, the number of U.S. citizens indicating their major on the GMAT is higher for Management than for any of the other sub-categories of Business Administration.

2. The GRE aggregates all sub-categories of Business into a single field (the LSAT does not).

3. The percentage represents the total number of respondents indicating an undergraduate major in any field of Business covered in our tables.

TABLE F-13

Test Performance by Undergraduate Major, 1977-1982:

24 Majors: Rank by Average Mean Differential

Rank	LSAT[1]	GMAT[2]	GRE/V	GRE/Q
1.	Mathematics	Mathematics	Philosophy	Physics
2.	Economics	Philosophy	English	Mathematics
3.	Philosophy	Engineering	Anthropology	Engineering
4.	Engineering	Chemistry	History	Computer Science
5.	Chemistry	Economics	Foreign Langs.	Chemistry
6.	Oth.Humanities	English	Physics	Other Science
7.	Foreign Langs.	Computer Sci.	Oth.Humanities	Economics
8.	English	Foreign Langs.	Journalism	Biology
9.	Anthropology	History	Polit. Sci.	Philosophy
10.	Biology	Oth.Humanities	Biology	Anthropology
11.	Oth. Sciences	Biology	Psychology	Business
12.	History	Oth. Sciences	Chemistry	Psychology
13.	Psychology	Polit. Sci.	Other Science	Foreign Langs.
14.	Journalism	Psychology	Mathematics	Other Humanities
15.	Art & Music	Oth. Soc. Sci.	Art & Music	Polit. Sci.
16.	Oth. Soc. Sci.	Art & Music	Economics	Oth. Soc. Sci.
17.	Polit. Sci.	Sociology	Computer Sci.	History
18.	Speech	Education	Oth. Soc. Sci.	English
19.	Business	Business	Sociology	Art & Music
20.	Sociology		Engineering	Journalism
21.	Education		Speech	Speech
22.	Social Work		Social Work	Sociology
23.			Business	Education
24.			Education	Social Work

NOTES: 1. The LSAT does not report separately for Computer Science majors, and the number of test-takers with a Physics major was too small to include in the tables.

2. The GMAT does not report separately for Journalism, Social Work, Speech, or Anthropology majors, and the number of test-takers with a Physics major was too small to include in the tables.

TABLE G

TABLE G

Faculty v. Committee of Examiners: Perceptions of Test Content v. Specifications

Percents Allocated to Content Categories

CONTENT CATEGORIES	Respondents' Current Curriculum	Respondents' Ideal Curriculum	Respondents' Item Classification	Committee Specifications
COMPUTER SCIENCE				
Software Systems and Methodology	40.7*	33.1	33.0	35
Computer Organization and Logic	20.1	20.7	20.1	20
Theory	12.2*	17.7*	20.5	20
Computational Mathematics	15.5*	16.6*	23.4*	20
Special Topics	11.5*	11.9*	2.7*	5
EDUCATION				
Educational Goals	15.9	16.4	17.8*	15
Administration and Supervision of Schools	11.1*	12.3*	15.8	15
Curriculum Development and Organization	24.4*	21.1*	13.6	15
Teaching-Learning	35.4*	32.6*	33.0*	40
Measurement, Evaluation, and Research	13.4*	17.6*	19.9*	15
CHEMISTRY				
Analytical Chemistry	22.3*	22.5*	Indicated in Separate Analysis of Each Sub-Discipline	15
Inorganic Chemistry	20.3*	22.1*		25
Organic Chemistry	30.7	28.2		30
Physical Chemistry	26.6*	27.0*		30

*Differs from Committee Specifications, p<.05

Source: Philip K. Oltman, Content Representativeness of the Graduate Record Examinations Advanced Tests in Chemistry, Computer Science and Education. ETS: 1982. Tables 3, 8 and 9.

Domestic versus Foreign Student Performance: GREs and GMATs

	GRADUATE RECORD EXAMINATIONS					GRADUATE MANAGEMENT ADMISSIONS TEST				
	Per Cent Domestic[1] Test-Takers	VERBAL		QUANTITATIVE		Per Cent Domestic[3] Test-Takers	VERBAL		QUANTITATIVE	
Year		Mean	Domes[2] Mean	Mean	Domes[2] Mean		For.[4] Mean	Domes[5] Mean	For.[4] Mean	Domes[5] Mean
1972-3	-----	497	500	512	510	N.A.	N.A.	N.A.	N.A.	N.A.
1973-4	----	492	498	509	505	N.A.	N.A.	N.A.	N.A.	N.A.
1974-5	----	493	497	508	507	N.A.	N.A.	N.A.	N.A.	N.A.
1975-6	92.5%	492	498	510	507	N.A.	N.A.	N.A.	N.A.	N.A.
1976-7	91.3	490	495	515	509	N.A.	N.A.	N.A.	N.A.	N.A.
1977-8	91.1	484	491	518	512	80	18.9	27.8	26.5	27.2
1978-9	90.0	476	486	517	508	79	19.6	27.8	26.3	27.0
1979-80	89.3	474	484	522	512	79	20.2	27.8	27.1	26.8
1980-1	88.1	473	483	523	513	80	20.8	28.3	28.0	26.8
1981-2	86.7	469	483	533	519	79	21.0	28.5	28.5	27.2

Notes:

1. The percentage of U.S. citizens taking the GREs is based on responses from those test-takers who (a) were either taking the GRE for the first time or who had taken it previously but prior to the beginning of the testing year in question and (b) who filled out the background information questionnaire. In 1981-1982, for example, that universe was 180,798--or 70.5% of the total number of test-takers.

2. William Turnbull, former President and currently Distinguished Scholar-in-Residence at ETS, computed these mean scores--which are approximations based on results of regularly scheduled domestic administrations of the GRE.

3. The percentage of U.S. citizens taking the GMATs is based solely on respondents to the background information question about country of citizenship. In 1981-82, for example, some 12% of the GMAT test-takers did not respond to this question.

4. Mean scores for non-U.S. citizens taking the GMATs were determined by the author by disaggregation, i.e. by removing U.S. citizens and non-respondents from the total. The results are thus approximations.

5. Mean scores for respondents indicating U.S. citizenship are clearly identified in the tables provided to the author courtesy of the Graduate Management Admissions Council.

*Chart provided courtesy of the Educational Testing Service

TABLE H, Page 2

GRE General Test Score Trends: Including and Excluding the Scores of Foreign Students, 1972-73 through 1981-82*

TABLE H-2

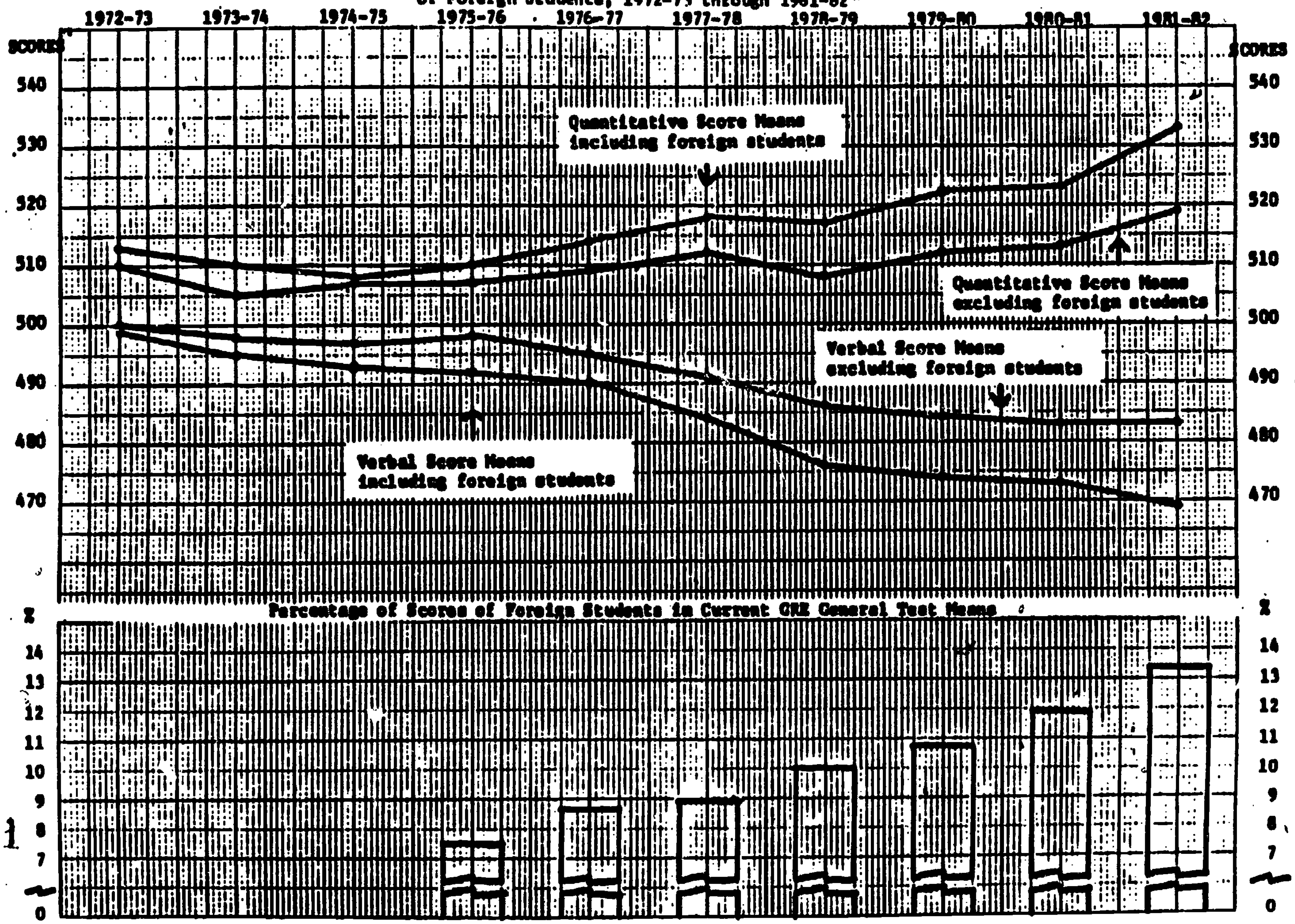


TABLE H-2

TABLE I

TABLE I

Changes in Number of Test Takers Following Inclusion of Local Program Assessment Uses of the GRE Subject Area Tests in the National Administrations in 1969

Test	1967-8 to 1968-9 % Change in N of Test Takers	1967-8 to 1968-9 Direction of Change in Mean Score	1968-9 to 1969-70 % Change in N of Test-Takers	1968-9 to 1969-70 Direction of Change in Mean Score	1969-70 to 1970-71 % Change in N of Test-Takers	1969-70 to 1970-71 Direction of Change in Mean Sco
Biology	6.0%	Stable	28.5%	Down	7.4%	Stable
Chemistry	0.2	Down	12.9	Stable	(1.1)	Up
Physics	(8.9)	Stable	7.6	Stable	(12.3)	Up
Geology	0.8	Down	30.9	Down	9.2	Stable
Mathematics	0.1	Down	12.8	Down	(0.3)	Up
Engineering	(9.3)	Down	23.3	Down	(4.6)	Stable
Economics	0.2	Down	24.7	Down	(0.2)	Down
Pol. Science	11.9	Down	13.2	Down	(5.5)	Down
Sociology	11.0	Down	33.1	Down	12.7	Down
Psychology	14.1	Down	27.3	Down	9.5	Stable
Education	14.6	Stable	36.2	Down	16.6	Down
History	4.2	Down	20.0	Down	(4.3)	Down
Eng. Lit.	7.7	Down	11.4	Down	(3.1)	Down
French	12.1	Down	(2.2)	Down	2.7	Down
Music	19.0	Down	33.6	Down	6.4	Down

Table J

Percentage of Graduate Departments with Viable Doctoral Programs* in Selected Fields that Either Required or Recommended the GRE Area Tests

Field	1971	1975	1979	Criteria for Inclusion: # of Ph.D.s Prior 3 Yrs.	or	# of Students	# of Departments Included, 1979
English	74%	83%	73%	10		35	120
French	73	76	76	5		15	49
History	76	75	71	10		24	113
Economics	78	75	69	10		30	99
Political Science	73	71	60	10		24	93
Sociology	68	71	62	15		25	102
Psychology	74	80	73	15		40	164
Chemistry	74	78	83	10		20	155
Physics	84	90	92	5		20	125
Mathematics	67	76	69	5		12	132

Source: Graduate Programs and Admissions Manuals: Princeton, N.J.: Educational Testing Service. Quadrennial publication.

*Percentages were determined by counting the number of "viable" doctoral programs and determining the number of those programs that either required or recommended the GRE Subject Area tests. The criteria for "viable" are key. There are no objective measures, and each field has to be looked at differently on the basis of the data in the Manuals.

TABLE J

TABLE K

U.S. Citizen/Foreign Student Mean Score x Selected Undergraduate Major:

GMAT Quantitative Examination, 1981-82

	U.S. Citizens Mean	Non-U.S. Citizens Mean*
Quantitative Based Diciplines		
Chemistry	31.32	31.09
Computer Science	31.34	31.47
Engineering	34.16	33.47
Mathematics	34.78	34.55
Physics	36.72	34.88
Economics	29.64	28.32
Accounting	27.83	27.78
Non-Quantitative Disciplines		
English	26.02	27.37
Foreign Languages	25.74	27.37
History	26.37	26.03
Psychology	25.40	26.03
Political Science	25.81	25.81
Sociology	23.49	24.83
Management	24.58	26.40

*Disaggregated, therefore approximate.

TABLE K U.S. /:::/ versus Foreign /XX/ Student Means on the GMAT Quantitative Examination Table K

by Selected Undergraduate Major

1981 - 1982

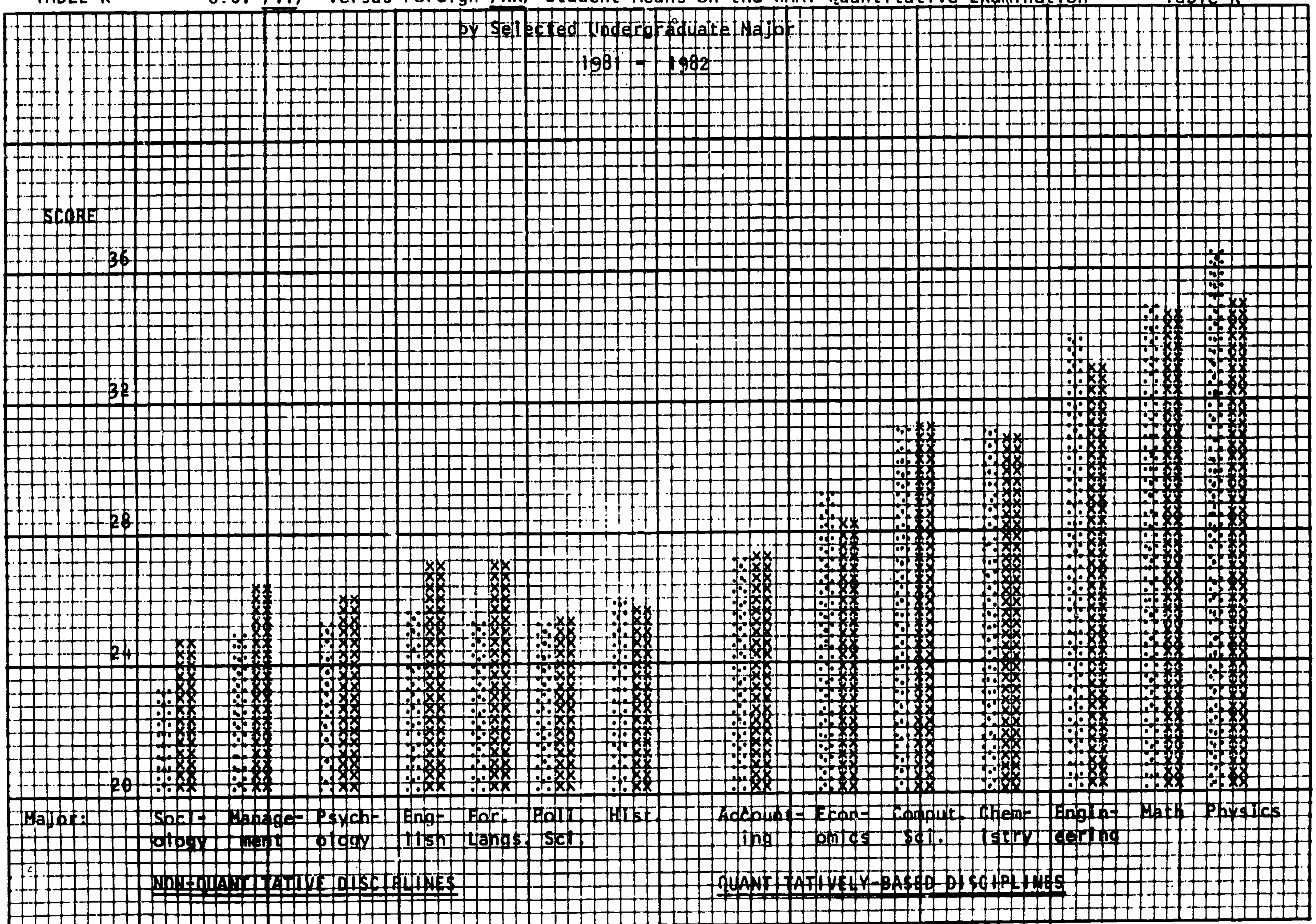


Table K-2

Table L

Table L

1982 ACT Composite Mean Scores for Entering Freshmen Planning to Major in Selected Disciplines*

Professional/Occupational Fields	Mean
Home Economics	15.9
Community Services (e.g. Social Work)	16.9
Industrial & Technical Fields	16.9
Education	17.8
Agriculture	18.2
Business Fields	18.4
Health Professions (e.g. Nursing)	19.3
Architecture	19.4
Engineering	21.7
Arts & Sciences Fields	
Social Sciences	20.6
Foreign Languages	20.7
Letters (e.g. English, History)	21.0
Biological Sciences	22.1
Mathematics	23.2
Physical Sciences	24.0

* American College Testing Program. College Student Profiles: Norms for the ACT Assessment. Iowa City, Iowa: ACT, 1983, pp. 64-83.

The "Conventional Wisdom" Hypothesis: Mean Scores and Candidate Volume on Four GRE Subject Area Tests, 1964-1982

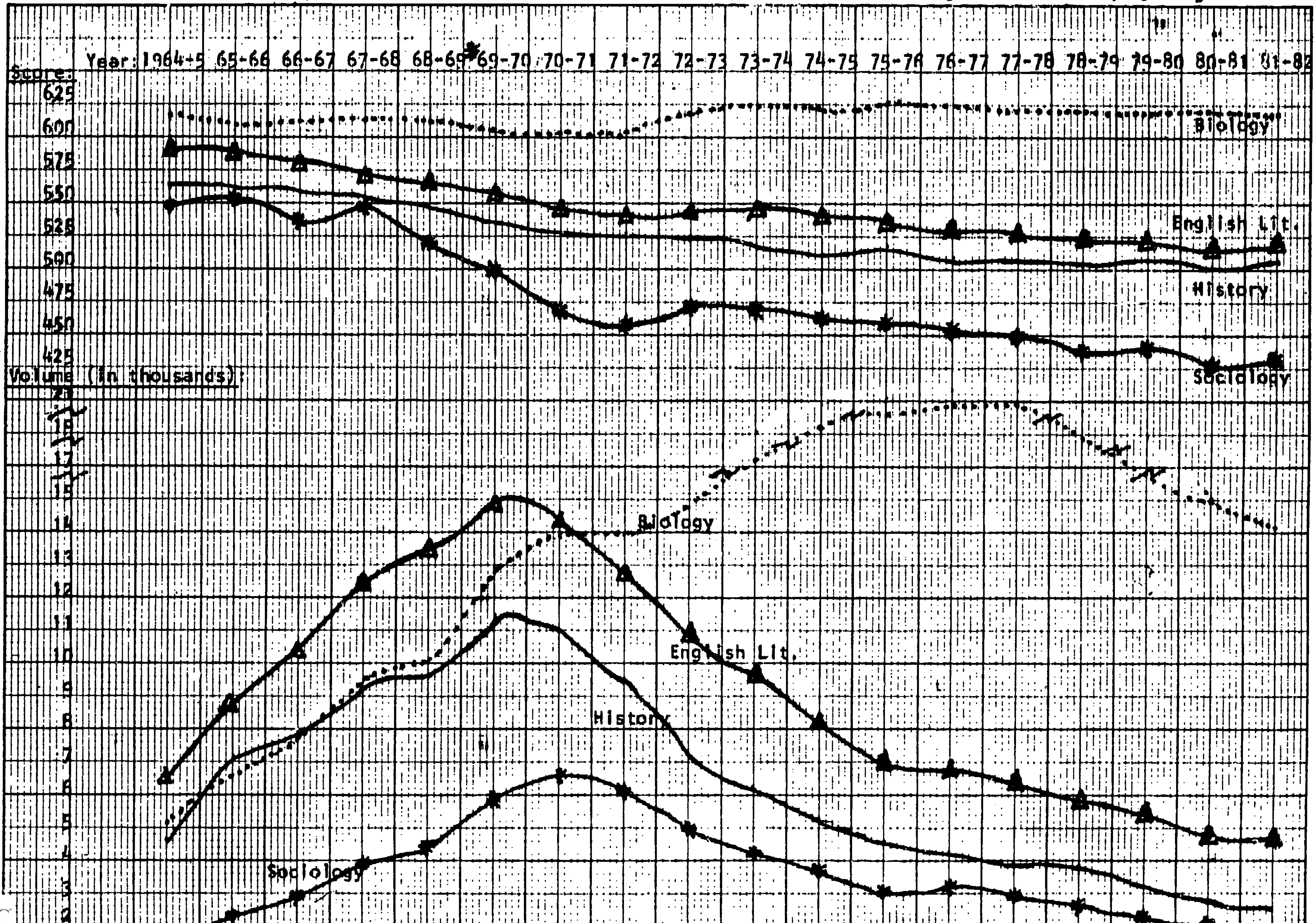


TABLE M

*For special factors influencing volume in 1969-70, see pp. 19-20 and Table I.