



GRE[®]

GRADUATE RECORD EXAMINATIONS[®]

**What Is The Value
of the
Graduate Record Examinations?**

July 2004

*Listening.
Learning.
Leading.*

Copyright © 2004 by Educational Testing Service. All rights reserved.
EDUCATIONAL TESTING SERVICE, ETS, the ETS logos, GRADUATE RECORD
EXAMINATIONS, and GRE are registered trademarks of Educational Testing Service.

Background

Questions and criticisms of the limitations of standardized testing are a perennial theme in the educational and psychological literature, particularly with respect to fairness and predictive validity. Fairness concerns focus particularly on differences in scores between men and women, and between racial and ethnic groups. Concerns about predictive validity are centered on the claim that test scores do not correlate highly with grades and other measures of success in graduate school (Glanz, 1997; Sternberg & Williams, 1997). Given the importance of GRE test scores in graduate admissions and financial aid decisions, institutions and individuals who use GRE test scores are properly concerned with assessing the validity of these criticisms.

Important Points

There are some important, though often overlooked, points to remember about the value of the GRE. They include the following¹.

1. GRE tests are intended to measure only a portion of the individual characteristics that are important for graduate study: reasoning skills, critical thinking, and the ability to communicate effectively in writing in the General Test, and discipline-specific content knowledge through the Subject Tests.
2. The General Test measures specific reasoning skills that are developed over a long period of time.
3. Standardized test scores and prior grades measure overlapping but different aspects of educational attainment, and they have complementary strengths. Each also has specific weaknesses as an indicator of likely student success. Unlike prior grades, GRE scores reflect performance on tasks that are common to all applicants. The complementarity between GRE scores and other elements of the application means that GRE scores should always be used in conjunction with those other elements.
4. The power of the GRE General Test and Subject Tests to predict first-year grades in graduate education is greater than has been represented in some recent criticisms. In fact, GRE scores enable important, practical improvements in admissions decision-making.
5. There are many different criteria for success in graduate education and a great disparity of views of how to measure it. Thus, there is no single, universally accepted way to gauge the utility of GRE scores. The prediction of graduate school grades is a common yardstick, but grades fail to capture all that is important about graduate study.

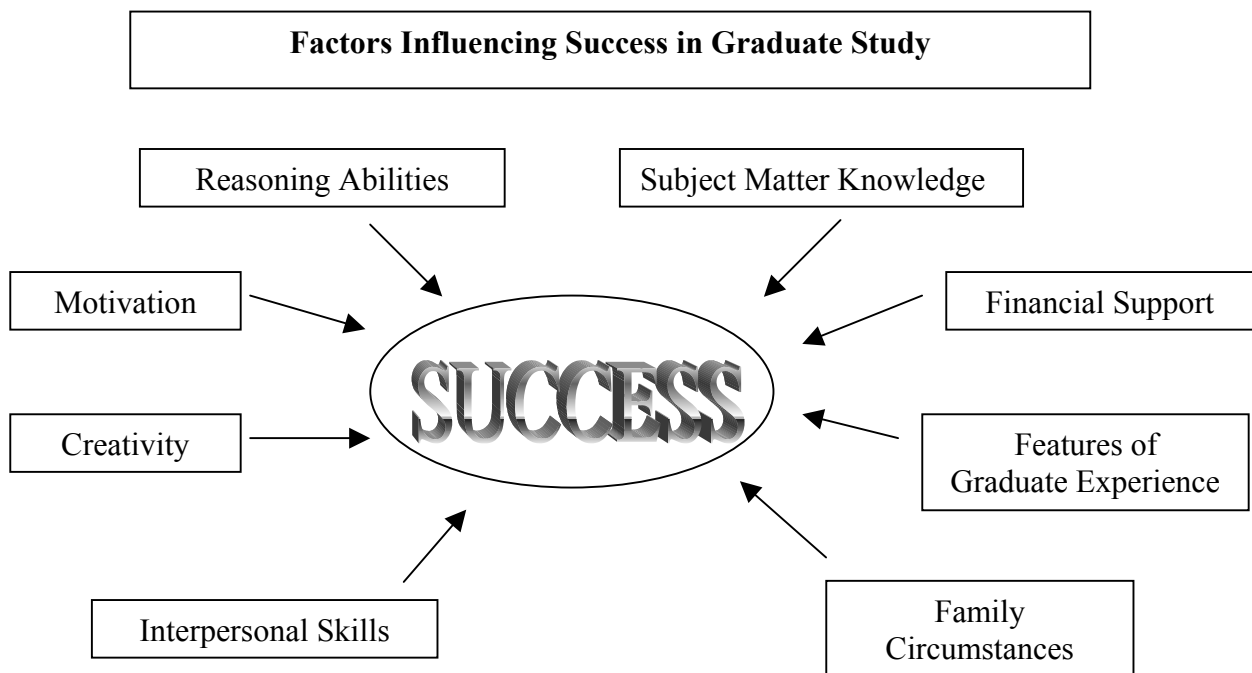
¹ The following points have frequently been made in GRE publications such as the *GRE Guide to the Use of Scores*.

Moreover, there are some serious technical problems associated with their use in studies of predictive validity.

6. Although institutions are the major users of test scores, test takers may also use them for their personal educational decisions.

Each of these points is discussed in detail below.

1. **GRE tests are intended to measure only a portion of the individual characteristics that are important for graduate study: reasoning skills, critical thinking, and the ability to communicate effectively in writing in the General Test, and discipline-specific content knowledge through the Subject Tests.**



The *Guide to the Use of GRE Scores* (Graduate Record Examinations Board [GREB], 2004) is very specific about the purpose of GRE test scores vis à vis other available information:

GRE test scores can be used by admissions or fellowship panels to supplement undergraduate records and other qualifications for graduate study. The scores provide common measures for comparing the qualifications of applicants and aid in the evaluation of grades and recommendations. [p. 5]

As illustrated above, many factors other than developed abilities in reasoning and knowledge of subject matter may influence successful graduate study, degree completion, and eventual career achievements. Some of these factors are attributes of the student, though others are attributes of the student's environment, including the graduate institution.

Currently, it is feasible to measure only a subset of these attributes in a fair and sufficiently accurate manner. Moreover, these factors are not likely to be equally important for success, or to be equally weighted in admissions decisions. Factors such as the ability to finance graduate education may affect completion of the degree program, but they are not generally part of the evaluation of applicants. Other institution related factors, such as faculty teaching and mentoring, and the match between student interests and the type of graduate program, are also important to success in graduate study (see Shavelson, Short, Muthen, Li, & Muthen, 1987 for a pertinent example from graduate schools of management). Of the factors shown on page 3, the GRE General Test and Subject Tests are intended to measure only reasoning, critical thinking, and subject matter knowledge.

Research on what is (and what is not) measured by the GRE tests and other standardized assessments is voluminous (Briel et al., 1993; Kuncel, Hezlett, and Ones, 2001). The important role of alternative assessments and indicators of achievement is clearly recognized in the *Guide to the Use of GRE Scores* (GREB, 2004). Guideline 1 is as follows:

Use Multiple Criteria. Regardless of the decision to be made, multiple sources of information should be used to ensure fairness and balance the limitations of any single measure of knowledge, skills, or abilities. These sources may include undergraduate grade point averages, letters of recommendation, personal statements, samples of academic work, and professional experience related to proposed graduate study. GRE scores should not be used exclusively.

The use of multiple criteria is particularly important when using GRE scores to assess the abilities of educationally disadvantaged students, students whose primary language is not English, and students who are returning to school after an extended absence. Score users are urged to become familiar with factors affecting score interpretation for these groups as discussed in [the *GRE Guide to the Use of Scores*]. [p. 6]

GRE tests add value by adding breadth to the profile of information about applicant abilities. When used in conjunction with other sources of information, GRE scores can contribute to greater flexibility in admissions and assist in the selection of a more diverse student body. For example, a candidate with outstanding recommendations from an unknown undergraduate institution may be more likely to gain admission when GRE scores are available to confirm his or her qualifications.

2. The General Test measures specific reasoning skills that are developed over a long period of time.

The General Test is designed to measure reasoning with verbal and quantitative materials, as well as the ability to make and to critique arguments in concise, well-organized essays. The test is designed to be independent of particular courses of study, instead representing knowledge, skills, and abilities that are developed across many courses and that reflect exposure to demanding courses of study. Of course, the knowledge, abilities, and skills

examined in the General Test are not immutable. At the same time, because skills measured in the General Test are developed over a long period of time, research has shown that short-term study will not alter scores greatly for most people (Powers, 1985). It is important that test takers be thoroughly familiar with the test – its contents and procedures – before the actual testing day, in order to avoid receiving lower scores than they might otherwise have obtained.

The *Guide to the Use of GRE Scores* (GREB, 2004) describes the test content as follows:

- The analytical writing section tests critical thinking and analytical writing skills. It assesses the ability to articulate and support complex ideas, analyze an argument, and sustain a focused and coherent discussion. It does not assess specific knowledge, and there is no single best way to respond.
- The verbal section tests the ability to analyze and evaluate written material and synthesize information obtained from it, to analyze relationships among component parts of sentences, and to recognize relationships between words and concepts. In each test edition, there is a balance among the passages across three different subject matter areas: humanities, social sciences, and natural sciences.
- The quantitative section tests basic mathematical skills and understanding of elementary mathematical concepts, as well as the ability to reason quantitatively and to solve problems in a quantitative setting. There is a balance among the questions requiring arithmetic, algebra, geometry, and data analysis.

The long-term reasoning component measured by the GRE General Test makes it a test of *developed skills*. Both words are significant here. The *skills* are reasoning skills of the type most important to success in graduate study. The *developed* nature of these skills comes from a lifetime as a reader, as one who thinks through quantitative problems, and as one who is accustomed to communicating and critiquing arguments in writing. The GRE General Test measures, in part, an orientation to thought and reflection as expressed in a lifetime of learning.

3. **Standardized test scores and prior grades measure overlapping but different aspects of educational attainment, and they have complementary strengths. Each also has specific weaknesses as an indicator. Unlike prior grades, GRE scores reflect performance on tasks that are common to all applicants. The complementarity between GRE scores and other elements of the application means that GRE scores should always be used in conjunction with those other elements.**

Standardized test scores provide information that is highly comparable across examinees, and that references a well documented set of knowledge and skills. Such scores do not, however, provide information about other important characteristics such as motivation and persistence. Furthermore, they do not typically provide a good vehicle for understanding such complexities as changes in level of attainment over time, or obstacles overcome to reach a particular level of attainment.

Grades and letters of recommendation, in contrast, are less comparable across examinees because they are based on different requirements for different individuals. Grades represent different combinations of courses and are based on grading standards unique to each institution, department, and professor. Letters of recommendation may provide valuable information about an individual, but typically offer little systematic basis for comparing one applicant to another. Despite these weaknesses, grades and letters of recommendation probably provide better information than standardized tests about such traits as persistence, motivation, and ability to overcome obstacles, all of which may be of considerable importance to success in graduate study. This is why it is important to combine grades, letters of recommendation, and test scores to develop a holistic understanding of an applicant's abilities and of achievement in relation to abilities.

The complementarity of test scores and grades is made clear by the results of validity studies conducted by the GRE Program in collaboration with various universities. These studies show that the combination of GRE scores and undergraduate GPA predicts first-year grades more effectively than any single piece of information (Schneider & Briel, 1990).

These various sources of information represent a system of checks and balances in decision-making. Fairness is enhanced by using multiple measures because systematic over- or underprediction will be decreased as additional measures are used. For example, students coming from schools or fields that employ strict grading practices will be less disadvantaged when tests are used along with grades. Conversely, test scores may represent peak performance capacity, while grades and letters of recommendation can offer testimony to an individual's ability to sustain performance over longer periods of time.

4. The power of the GRE General Test and Subject Tests to predict first-year grades in graduate education is greater than has been represented in some recent criticisms. In fact, the use of GRE scores enables important, practical improvements in admissions decision-making.

The power of the GRE General Test and Subject Tests to predict first-year grades in graduate education is high enough to increase users' ability to identify potentially successful students and to significantly decrease the number of decision errors that are made. Rosenthal & Rubin (1982) demonstrated that there is a widespread tendency to underestimate the importance of correlations of the size typically found between GRE scores and graduate school grade point averages.

The usefulness of the GRE tests is generally underestimated because of the effects of selection and the lack of reliability of the measures that are used to gauge success in graduate education. First-year graduate school grades are generally very high, and typically they show little variation either within or across programs and institutions. The lack of variability in grades creates a restriction of range that artificially limits the size of the correlation that can be obtained.

Restriction of range issues have an analogue in the GRE test scores themselves, due to the problem of selection bias. Recipients of lower GRE scores are less likely to be accepted into a graduate program, and may also be less likely to apply. Many recipients of low GRE test scores, then, will never have the opportunity to earn first-year grades. To the extent such people would have been less successful in graduate school, GRE test scores served the intended purpose of making this known in advance of a decision to admit. Ironically, it is a successful prediction that cannot be demonstrated statistically.

In response to these issues, a recent meta-analysis of GRE predictive validity by Kuncel, Hezlett, & Ones (2001) applied corrections for range restriction and the unreliability of grade point average as a criterion measure. When a correction for the restriction of range is applied, the correlation between GRE scores and first-year graduate grades becomes substantial. Applying such corrections to the data given in GREB (1999) also produced higher estimates of validity. For GRE verbal, quantitative, and analytical scores in combination with undergraduate grade point average, corrections increased multiple correlations to between .55 and .60 when predicting first year graduate grade point averages in various academic disciplines (D. Powers, personal communication).² With respect to the GRE Subject Tests, the corresponding correlations were also higher with corrections, but much more variable across tests, ranging from .24 to .58.

The combined predictive power of GRE scores and undergraduate grades is also increased by using them together rather than separately. The median correlation between undergraduate grade point average and first year average in graduate school across a variety of disciplines is .33. When GRE General Test scores and scores from the relevant subject test are also included, the median multiple correlation increases to .50. Further details about these correlations are given in GREB (2004, p. 22).

5. **There are many different criteria for success in graduate education and a great disparity of views on how to measure it. Thus, there is no single, universally accepted way to gauge the utility of GRE scores. Grades are frequently used, but they fail to capture all that is important about graduate study. Moreover, there are some serious technical problems associated with their use in studies of predictive validity.**

Grade point averages are often used as the criterion of success in predictive validity studies because there is general agreement that grades in graduate school have some important educational meaning and, unlike other theoretically desirable criteria, they are readily available for study. However, first-year graduate school grades are generally very high, and typically they show very little variation either within or across programs and

² This information describes multiple correlations based on optimally weighted combinations of predictors, not on a simple sum of scores from the separate GRE General Test measures. The program specifically advises against summing scores, as the same total score can result from very different profiles of performance on the three parts of the General Test (and therefore have a very different meaning).

institutions. Grades in subsequent years may exhibit even less variability (Sternberg & Williams, 1997). Moreover, "grade inflation" exacerbates this problem (Archibold, 1998). The lack of variability of grades, which has long been documented (Wilson, 1979), creates a restriction of range that artificially limits the size of the correlations that can be obtained, thus necessitating statistical corrections such as those discussed in point 4.

Of course, first year grades are not the only possible measure of success in graduate school, and a strong first year classroom performance is not an end in itself. Kuncel and his colleagues also summarized the relationship between GRE scores and several alternative criteria, such as faculty ratings of students and student performance on comprehensive exams. The correlations between the verbal and quantitative measures for predicting faculty ratings and comprehensive exam scores were moderately large. In general, the Subject Tests were better predictors of performance in a field than were the General Test measures.

Completion of one's degree program is generally regarded as a compelling criterion of success in graduate school. However, the ability of the GRE tests to predict degree completion or professional success is limited by the influence of numerous factors that are entirely unrelated to a student's academic abilities. Hence, the relationship between GRE scores and such distant criteria of success will often, of necessity, be unimpressive. If the purpose of the application packet is to demonstrate a candidate's fitness to undertake graduate study in a given field, then the correlation with first-year grades makes clear that GRE scores are a valuable part of the application packet.

6. Although institutions are the major users of test scores, test takers may also use them for certain decisions.

Tests may provide information that is used by students in such decisions as which schools to apply to and in what disciplines. It is often of interest to students to know, for example, how their scores compare with those of other applicants and with those of enrolled students at particular institutions. This information may serve as an indication of the likelihood of gaining admission and of the probability of success once admitted. However, there has been little systematic study of how students actually use test score information, or of the effects of this use. On the one hand, applicant self selection on the basis of test score information might serve merely to "homogenize" the student body rather than to promote diversity. On the other hand, it might serve to facilitate good matches between applicants and graduate departments.

In any case, just as with the use of test scores by institutions, students should be made aware that their test scores are only one indicator of the likelihood of successful graduate study. Test scores should never be considered in isolation from other factors important to success in graduate study.

Conclusion

The above considerations suggest the following concluding points:

- GRE test scores are one factor among many that should be taken into account in admissions and fellowship decisions.
- GRE test scores improve the predictive capacity of other information such as undergraduate grade point averages, letters of recommendation, application essays, and other writing submissions. Each of these elements of the standard application packet contains valuable information, but each measures different traits of the applicant and does so with different operational strengths and weaknesses.
- The particular strengths of GRE test scores as a part of the application packet lie in their focus on developed reasoning skills, their comparability across all applicants, and (of special relevance for returning students) their recentness.
- As graduate school applicant populations become more and more diverse in culture, educational background, and recentness of undergraduate training, GRE test scores are likely to become an increasingly important component of the application packet.

Specific guidelines for the use of GRE scores can only be developed based on the experience of a given department with its pool of applicants. Among the factors that would come into play in looking at applications are selectivity of the department and past experience of students in the program. The use of GRE scores will differ depending on whether the admissions process is one of selecting just a few individuals from a large pool of qualified candidates, or if a department seeks solely to determine whether a given candidate has the ability to complete the course of study. For this reason, and because the skill areas most important to success differ from discipline to discipline, it is important for admissions committees in each department to develop a sense of how GRE scores and other elements of the applicant's file relate to success in their program. GRE staff at the Educational Testing Service will be pleased to work with any department to organize and analyze their data on the relationship between GRE scores and success in the department's graduate programs.

References

- Archibold, R. C. (1998, February 18). Just because the grades are up, are Princeton students smarter? *New York Times*, pp. A1, B8.
- Briel, J. B., O'Neill, K. A., & Scheuneman, J. D. (Eds.) (1993). *GRE Technical Manual: Test Development, Score Interpretation, and Research for the Graduate Record Examinations Program*. Princeton, NJ: Educational Testing Service.
- Durso, R. (1997, May 6). *GRE Test Use and Score Interpretation: Current Concerns and Possibilities for the Future*. Presented at the Annual Meeting of the National Association of Graduate Admissions Professionals, Orlando, FL.
- Glanz, J. (1996, November 1). How not to pick a physicist? *Science*, 274 (5288), 710-712.
- Graduate Record Examinations Board (2004). *2004-2005 Guide to the Use of the GRE*. Princeton, NJ: Educational Testing Service.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127, 162-181.
- Powers, D.E. (1985). Effects of coaching on GRE Aptitude Test scores. *Journal of Educational Measurement*, 22, 121-136.
- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.
- Schneider L.M., & Briel, J.B. (1990). Validity of the GRE: 1988-89 Summary Report. Princeton, NJ: ETS.
- Shavelson, R., Short, L., Muthen, B., Li, W., & Muthen, L. (1987). *Relationship between applicant characteristics, MBA program attributes, and student performance*. McLean, VA: Graduate Management Admission Council.
- Sternberg, R., & Williams, W. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? *American Psychologist*, 52, 630-641.
- Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement*, 30, 1-21.
- Wilson, K. M. (1979). *The validation of GRE scores as predictors of first-year performance in graduate study: Report of the GRE Cooperative Validity Studies Project* (GRE Board Research Report GREB No. 75-8R). Princeton, NJ: Educational Testing Service.