RESEARCH REPORT

Predicting Long-Term Success in Graduate School: A Collaborative Validity Study

Nancy W. Burton
Ming-mei Wang

April 2005

GRE Board Report No. 99-14R
ETS RR-05-03

Predicting Long-Term Success in Graduate School:
A Collaborative Validity Study

Nancy W. Burton and Ming-mei Wang

ETS, Princeton, NJ

GRE Board Research Report No. 99-14R

ETS RR-05-03

April 2005

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

Educational Testing Service, Princeton, NJ 08541

****

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in Graduate Record Examinations Board reports do no necessarily represent official Graduate Record Examinations Board position or policy.

The Graduate Record Examinations and ETS are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

****

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

Educational Testing Service
Princeton, NJ 08541

Abstract

GRE verbal and quantitative scores and undergraduate grade point average were evaluated as predictors of multiple measures of long-term graduate school success. The measures of success were cumulative graduate grade point average and faculty ratings on three student characteristics: mastery of the discipline, professional productivity, and communication skill. Seven graduate institutions and 21 graduate departments in biology, chemistry, education, English, and psychology collaborated in order to identify measures of valued outcomes, develop reports useful to individual departments and graduate schools, and initiate a database for future studies. Results are reported for all departments combined and by discipline and, where sample sizes permitted, for master's and doctoral degree students, men and women, U.S. citizens and noncitizens, domestic ethnic groups, and test takers who took the GRE computer-based test and those who took the paper-and-pencil version of the test. The results indicate that the combination of GRE scores and undergraduate grade point average strongly predicts cumulative graduate grade point average and faculty ratings. These results hold in each discipline and appear to hold in the small subgroups.

Key words: Predictive validity, GRE scores, measures of long-term graduate success, faculty ratings of graduate students, undergraduate grade point average, cumulative graduate grade point average, GRE verbal scores, GRE quantitative scores

Acknowledgments

The complex activities and analyses resulting in this report could not have been carried out without the dedication and support of a large number of people. We particularly want to mention the unique and major contributions made by the graduate deans and college deans and their associates, department chairs and faculty, data managers, and most importantly, the study liaison at each participating university and department. We are thankful for the insight and problem-solving flexibility of our colleagues at ETS. Our study team consisted of Linda DeLauro, the principal liaison with university participants; Laura Jerry, designer and creator of the validity database and the tables and graphics for departmental reports; Ting Lu, our statistical data analyst; and Fred Cline, who developed the final report tables and graphs. Special thanks are due to reviewers Brent Bridgeman, Linda Crocker, Rick Morgan, Liora Schmelkin, and Steve Sireci and to Dan Eignor for his own review and synthesis of their comments. Finally, we want to thank the GRE Board for funding the study and ETS's Research group for providing technical resources and steady support.

The opinions expressed in this report are those of the authors and do not necessarily reflect those of ETS.

Table of Contents

List of Tables

List of Figures

The current study is part of a long tradition of research on the predictive validity of the GRE®. Prior to 1975, most criterion-related validity information came from locally conducted institutional studies (for example, Lannholm 1960, 1968, 1972; Lannholm & Schrader, 1951) and from studies conducted by ETS in cooperation with graduate institutions, as summarized by Willingham (1974). Wilson (1979, 1986) conducted a series of cooperative validity studies with some 130 participating graduate departments to provide general predictive validity information as well as validity information for special subgroups of graduate students. Starting in 1978–79, despite technical problems caused by small department sizes, highly correlated admission measures, a restricted range of talent among enrolled graduate students, and very limited variation in graduate grades, the GRE Validity Study Service provided free studies to participating institutions. In the mid-1980s, the GRE Board supported the introduction of improved empirical Bayes statistical methods in the Validity Study Service (Braun & Jones, 1985; Schneider & Briel, 1990).

By the early 1990s, however, a moratorium was placed on the GRE Validity Study Service because the improved empirical Bayes methods could not completely overcome the technical problems mentioned above. During the years of this moratorium, further progress was made in areas relevant to GRE. Longford (1991) proposed statistical improvements in empirical Bayes methodology to control negative regression weights. Other statistical methods were developed (Ramist, Lewis, & McCamley, 1990; Ramist, Lewis, & McCamley-Jenkins, 1994), and meta-analysis validity generalization methods in wide use in employment studies were used to summarize GRE predictive validity studies (Kuncel, Hezlett, & Ones, 2001). These statistical methods show promise of being adaptable to GRE's needs. The current validity study is a type of meta-analysis, combining results from collaborating institutions and departments, although it uses different methods than those used by Kuncel et al. The Lewis and Ramist procedures (Ramist et al., 1990, 1994) for correcting for multivariate restriction of range were used in the in this study.

Recent events have focused increased attention on studies of predictive validity. In response to legislation in California and Washington, graduate, professional, and undergraduate institutions have become concerned about or have dropped affirmative action in admissions. This has placed pressure on tests and other admission measures that show lower performance by minority applicants. Some graduate faculty have published criticisms of the GRE General and Subject Tests (Georgi, as quoted in *"How Not to Pick a Physicist*," 1996; Goldberg & Alliger, 1992; Morrison & Morrison, 1995; Sternberg & Williams, 1997). Several of the criticisms are based on single validity studies with poor results, or on conflicting results from different validity studies.

Graduate institutions need reliable and up-to-date validity research to guide them in choosing which information to use when selecting graduate students. In addition, there are many important questions about graduate admission, such as fair treatment of minority groups, the effectiveness of the GRE Subject Tests, and the relationship of admission variables to long-term success in the field, that can very seldom be answered in a single department. Thus useful validity research information for graduate schools should include summaries of multiple studies, especially summaries of interpretable collections of disciplines and institutions, in order to answer questions of general interest, and provide more stable results than can be provided by studies done in individual departments.

Meta-analysis or validity generalization studies (Glass, 1976; Hedges & Olkin, 1985; Hunter & Schmidt, 1990) are useful methods of providing summaries of many independent studies. Individuals doing meta-analyses collect studies in the published literature, adjust the data to make them more comparable, and provide summaries that can be evaluated for statistical significance. Kuncel et al. (2001) report a major meta-analysis of approximately 50 years of published GRE validity studies, from the late 1940s to the late 1990s. Meta-analyses are limited, however, by what the original researchers chose to study and the data they chose to publish. Much of the art of meta-analysis involves developing plausible estimates for data not reported, such as standard deviations, correlations, and reliabilities of critical variables. In this study, we used a common design to collect comparable data from all participants, and hence were able to calculate the crucial statistics for all departments. Meta-analyses are further limited by the types of departments and institutions that choose to do studies and publish them. The sample of studies available in the literature may not represent some disciplines or types of institutions well. Even if statistical tests reveal no significant differences among disciplines or institutions, graduate deans and graduate faculty may pay little attention to results in which their discipline or type of institution is not represented.

Advantages of This Study

This study was developed to collect new data on the predictive validity of the GRE. It presents the first predictive validity data for the GRE administered in a computer-adaptive mode, introduced in the 1993–94 school year. The study collected multiple measures of graduate school outcomes to provide more comprehensive information about what GRE scores and undergraduate grade point average are able to predict. Graduate deans and faculty were invited to collaborate to assure that the outcome measures developed would be important to a variety of graduate institutions and disciplines. The collaborators also evaluated the efficiency and usefulness of data collection and quality assurance, analyses and reports of individual department results, analyses and reports summarized overall and by discipline, and a database designed for the accumulation of future studies. A total of 21 departments in biology, chemistry, education, English, and psychology from seven different graduate institutions participated. Institutions submitted analyzable data on 1,700 students who entered either a master's or a doctoral degree program in 1995–96, 1996–97, or 1997–98.

This study was necessarily small to encourage active collaboration and to allow the procedures to be modified based on user evaluations. The study was intended to describe admissions in the graduate community. At this early stage of understanding admission to graduate education, hypothesis generation is our goal; hypothesis testing can follow when we begin to believe we understand the system. Because we are attempting to capture a national picture of admission to graduate education, we consider results for small departments and small groups of students to be just as important as those for large groups (though certainly less reliable). Since our purpose is hypothesis generation and our sample sizes are small, we do no statistical tests in this report.

Outcome measures. A major advantage of this collaborative study is that we collected comparable information on a number of important outcomes of graduate school. It is often remarked that first year grades do not represent the most important goals of graduate school (Sternberg & Williams, 1997; Yee, 2003). Yet, test publications have frequently stated that the GRE is meant to predict performance in the first year of graduate school. Presumably, this cautious statement is based on the fact that the vast majority of validity studies use first-year graduate grades as a convenient proxy for success in graduate school. However, it does not make sense that the skills that lead to success in the first year of graduate school would differ radically from the skills that lead to ultimate success. In any case, a measure that predicts first-year grades but is unrelated to later success would not be a desirable admission measure. This study was designed to collect information on a broader definition of success in graduate school. There is evidence, summarized most recently in Kuncel et al. (2001), that GRE scores and undergraduate grades predict a number of long-term outcomes of graduate school. This is a long-standing but seldom-studied finding on which this study will provide further evidence.

Outcome measures for this study were developed based on the research literature and on interviews with GRE users about their most important goals for graduate students (Walpole, Burton, Kanyi, & Jackenthal, 2002). We collected data on cumulative graduate grade point average and faculty ratings. Faculty rated students on their professional knowledge, ability to apply that knowledge, and ability to learn independently (mastery of the discipline); their judgment in choosing professional issues and their creativity and persistence in solving the issues (professional productivity); and their ability to communicate what they have learned (communications skills). This expanded outcome information is important because it allows users to evaluate admission measures against a variety of goals considered important for graduate students.

Institutions and disciplines studied. Another advantage of this study is that the sample of institutions and disciplines covers the breadth of the graduate community. Institutions from master's, doctoral, and research Carnegie classifications represent a variety of missions, from regional professionally oriented master's degree programs, to programs primarily focused on teaching, to research programs that recruit nationally and internationally for top doctoral students. (Participating institutions and departments are listed in Appendix A.)

The disciplines sampled were

- Biology
- Chemistry
- Education
- English
- Psychology

These disciplines were chosen because they enroll large numbers of students and require a wide variety of skills and knowledge.[1] The academic areas were limited to make it possible to summarize validity results within discipline. A relatively small sample of departments is dictated by the need for close collaboration among researchers and participants. The sample is intended to initiate a broadly representative and cumulative database, which would allow a variety of analyses and summaries. A representative database is critical because it determines whether the graduate community will believe that summary results adequately represent their students and what they study.

Common reporting subgroups. A third advantage of this study is that we were able to collect a set of background questions that allowed us to combine data and report results for subgroups including:

- Women and men
- African American, Asian American, Hispanic American, and White students
- Citizens and noncitizens
- Master's and doctoral degree students
- Test takers who took the computer-based test and those who took the paper-and-pencil version of the test

In addition to the overall effectiveness of the admission process, score users and prospective students are concerned that the process be equally valid and fair for all prospective students, particularly for groups that are relatively new to graduate education, or those who have been traditionally underrepresented in graduate school. A study with a common design is a first step in being able to answer these questions, starting a database that can eventually give dependable answers to questions involving small groups.

Predicting success in graduate school. This study evaluated the most common objective measures used to predict graduate school success at admission: GRE verbal and quantitative scores and undergraduate grade point average. A study like this could be used to evaluate possible new admission measures, but we felt that it was important to develop broader outcome measures first. Focusing on outcomes is a good way to start a dialogue among participating institutions about the goals of graduate education, and clarity about goals is the best start for a consideration of new admission measures. It may also be necessary to develop new outcome measures to serve as criteria for evaluating new admission measures. Willingham's (1985) study of undergraduates found that grades and test scores are the only measures necessary for predicting academic outcomes. It was only when broader outcomes such as leadership and accomplishment were evaluated that alternative admission measures were required to achieve good predictions. We suspect that this will also be true in graduate school, where broad outcomes are even more important than they are for undergraduates.

The results of this study are reported in two parts. After a brief discussion of the methods used in the research, we will present the most important results of the study. These are the correlations between admission measures—GRE verbal and quantitative scores and undergraduate grade point average—and outcomes of graduate school. These correlations allow the reader to evaluate how strongly GRE scores and undergraduate grades predict important long-term measures of success in graduate school. The second part of the results presents more detailed analyses. Because the most data were available for cumulative graduate grade point average, this detailed analysis will focus on that one outcome measure. The detailed analysis will present regression equations that can be used to check individual department results, and can also be used in admission by departments that have not yet done an individual prediction study. The detailed analysis will also examine the effectiveness and fairness of GRE scores and undergraduate grade point average for use in admitting selected subgroups of students.

The primary audience of this report is current and potential users of GRE scores who are concerned about the validity of decisions made using the GRE and undergraduate grades. The text of the report is written for this academic audience and makes minimal assumptions about knowledge of measurement or statistics. For those interested, a few study details and statistical issues are discussed in endnotes or table footnotes.

Methods

Measures

Admission measures. The admission measures, or *predictors*, studied were GRE verbal and quantitative scores and undergraduate grade point average. These measures were taken, when possible, from institutional records. For example, some institutions reported using the best GRE score from any administration of the GRE taken by an applicant; we used those scores when possible, since the purpose of a validity study is to validate the actual admission decisions made at an institution. Students in participating departments were also sought on the official GRE files at ETS. When institutions did not provide GRE scores, undergraduate grade point average, or background information about students, the relevant information was taken from the GRE files. Although the institutional files were, in general, the preferred source for information, we preferred to use the students' self-report of race/ethnic group and citizenship when available, since this is information that each student should know better than anybody else. Finally, in the analysis comparing applicants who took the computer-adaptive GRE to those who took the paper-and-pencil test, we did not use the institution-supplied scores, since they did not have information on mode of delivery.

The basic analysis of the predictor measures involved using the *multiple regression* statistical technique to find the combination of admission measures that best predicts an outcome measure in a department. The multiple regression analysis provides a *regression weight* for each predictor. When a predictor score for a given applicant is multiplied by its regression weight and added to the other predictor scores for that student (also multiplied by their weights), the resulting number is a predicted outcome; for example, a predicted cumulative graduate grade point average for that student. This predicted cumulative graduate grade point average can be thought of as a summary of all the information in GRE scores and undergraduate grade point average that is relevant to earning graduate grades. The equation developed on one year's entering students is frequently used to predict the future performance of applicants in subsequent years: It is a convenient way to summarize in one number the objective information about applicants.

Outcome measures. At the start of this research, we conducted telephone interviews with GRE score users. We spoke to deans in seven institutions and to faculty in six academic disciplines (the disciplines included in this report plus engineering). The interviewees were asked to discuss the qualities and skills of successful graduate students. The top five (adapted from Walpole et al., 2002, p. 14) are:

- Persistence, drive, motivation, enthusiasm, positive attitude
- Amount and quality of research or work experience
- Interpersonal skills/collegiality
- Writing/communication
- Personal and professional values and character, such as integrity, fairness, openness, honesty, trustworthiness, consistency

On the basis of these discussions with members of the graduate community, and a review of the literature on faculty ratings, we developed several measures to be used as outcomes or criteria of graduate school success in this study. We asked faculty to rate three characteristics of each student—mastery of the discipline, professional productivity, and communication skills. We requested that two faculty members familiar with the student rate each student on each of these three characteristics.

Our definition of *mastery of the discipline* reflects an academic component, but goes beyond knowledge to include three other components: ability to apply that knowledge to new situations; ability to structure, analyze, and evaluate problems; and an independent ability to continue learning.

Our *professional productivity* faculty rating includes, among other things, the most highly valued quality, persistence. The complete definition is the extent to which the student shows good judgment in selecting professional problems to attack, and the practical abilities of planning, flexibility in overcoming obstacles, and determination in carrying problems to successful completion.

Our *communication skill* faculty rating combines both interpersonal skills and communication, and, in addition, basic standard English for nonnative speakers. Communication skill is defined as the ability to judge the needs of one's audience; a mastery of the language of the discipline; a mastery of standard English; and the ability to communicate and work cooperatively with others. All three faculty ratings use a six-point scale, ranging from 1 for unsatisfactory and to 6 for outstanding, with 0 for students the faculty member does not know well enough to rate (counted as missing data in the analysis). To increase the reliability of the ratings, departments were asked to have each student rated by two faculty members who knew them well.

Although the list of qualities and skills of successful graduate students developed for this study conspicuously lacks a mention of academic accomplishments, the interviewees appeared to assume that their students, admitted on the basis of past achievements, would continue to achieve in the future. Thus, cumulative graduate grade point average was added to the study as the primary measure of academic accomplishment in graduate school. It is reported on a scale ranging from 0 (failing) to 4 (A). (See Appendix B for the complete definitions of the outcome measures used.)

Finally, we collected information on the students' progress to degree including such important milestones as master's and doctoral common examinations and degree attainment. Originally intended to be used as an outcome measure, analysis results were inconsistent and difficult to interpret, and several problems in the data were revealed, so degree progress was removed from the final analysis. Descriptive information about the measure is reported in the results section, and suggestions for developing better measures of progress to degree are addressed in the discussion.

Data Collection and Checking

Participating departments submitted data for students who initially enrolled as master's degree candidates in the 1995–96, 1996–97, or 1997–98 school years, or who enrolled as doctoral degree candidates in the 1995–96 or 1996–97 school years. They submitted data on demographic characteristics, admission measures, grades, graduate school milestones, and faculty ratings. Data were checked for plausibility and missing values, and matched to GRE score files containing test scores and background questionnaire responses.

A second round of data checking occurred after initial analyses were completed. Observed cumulative graduate grade point averages were plotted against cumulative graduate grade point average as predicted by the equation combining all three predictors (GRE verbal and quantitative scores and undergraduate grade point average). Unusual data points were checked against the student's full record and, where necessary, against institutional records. Forty-one students were removed from the initial cumulative graduate grade point average analysis data set of 1,351 students, and analyses were rerun with the edited data. Students were removed, for example, because they were international students whose undergraduate grades had been converted in a way that led to an implausible predicted cumulative graduate grade point average. Others were removed because they had only attended for a term or two, and their observed grades were very low, often because of unresolved incompletes.

Analysis Strategies

Relating design and analysis to purpose. The unifying theme of our design and analysis is that the process of graduate program selection is probably best viewed, and best evaluated, from a slightly more general perspective than that of the individual department or graduate institution. There are many reasons for this viewpoint. The most practical is that many graduate programs are too small to supply stable results. Another important element is that the process by which students select graduate schools occurs well before an application is submitted. The students do not consider all programs (they self-select), and their undergraduate mentors suggest programs to pursue and to ignore. Thus only part of the total selection process occurs in graduate admission offices or faculty selection committees.

Graduate education is a highly interactive national system. Institutions and departments have many links. Graduate faculty come from various training institutions; students come from more or less widely scattered undergraduate institutions; the students are also linked through undergraduate professors to still another collection of institutions. This geographical matrix is overlain by links created by disciplinary schools of thought, professional organizations, and even consulting circuits. That is, individual graduate departments are best understood as part of a national or international professional community. This is particularly important in studying the process by which students select institutions and institutions select students. This study uses a national context to organize results and to make individual department results as comparable as possible.

For example, the collaborating institutions were sought out to cover a broad range of the graduate community. Although it would be ridiculous to speak of seven institutions as representative, they can act as the basis for a database that could eventually represent a national system of graduate education. We focused on outcome measures in order to find goals that are common across the graduate community. We focused on a limited number of disciplines because users told us they would find summaries for their own discipline meaningful. Finally, we used statistical techniques to make results more comparable across institutions and disciplines. These are discussed in the following sections on analyses.

Within-department analyses and summaries by discipline. Within each department, the analysis data set for each outcome consists of those students with complete data on all three predictors and the outcome measure. The minimum sample size for analysis was defined as 9 students with complete data. The small samples were allowed so that participating departments would get a report based in part on their own data; even so, two of the original 21 participating departments had too little complete data for analysis. All possible combinations of the three predictors were used to compute prediction equations. Because we used the same set of students to compute each equation, the results from different equations are comparable.

Correlation coefficients are reported uncorrected and corrected for restriction of range on all predictors. Measures used in student selection become restricted in range. For example, very few students are admitted with undergraduate grade point averages below 2.5. Restriction in range lowers correlation coefficients, so grade point average will look like a poorer predictor of graduate school outcomes than it really is. Those students with low grades who were not admitted would have tended to earn low grades in graduate school; the missing data would have supported the validity of undergraduate grade point average for selection decisions. The correction for restriction in range estimates what the correlations would be if the relationship found in a single education department, for example, were applied to all GRE takers who sent scores to education departments.[2] Since the correction is applied to each department, it creates a common statistical population across all departments within a general disciplinary area. The reference populations used in the correction were the GRE test-taking population for the 1994–95 testing year in each of four different general areas—natural sciences for biology and chemistry departments, social sciences for psychology, arts and humanities for English, and education for education. These corrections put all correlation coefficients within a general area on a comparable basis, and the total cross-disciplinary summary of coefficients also combines areas that are roughly comparable, because each area was adjusted to its national GRE population.

Summaries of correlations are averages of the individual department coefficients corrected for multivariate restriction of range and weighted by the number of students in the department. This method of summary will help compensate for the unstable results that are likely to occur in small departments, since their results, multiplied by a small number of students, will have little influence on the weighted average.

Regression analysis maximizes the correlation between predictors and criterion, and may be inordinately influenced by unusual data points. When sample sizes are small, inflated correlations become likely. Small samples occur frequently in our subgroup analysis and so the subgroup tables include correlations corrected for shrinkage.[3] The shrinkage adjustment did not seem conceptually compatible with our correction for restriction of range, so we adjusted uncorrected correlations only. These may help the reader estimate how much the correlations have been affected by small samples.

Results are discussed when they are considered to be of notable size, using arbitrary criteria such as those proposed by Cohen (1977) for the behavioral sciences. We follow Cohen's convention of classifying correlations between .1 and .3 as *small*; between .3 and .5 as *medium* or *moderate*; and .5 and higher as *large* or *strong*.

Pooled department analyses and summaries. In order to develop regression equations and correlation coefficients on a larger and more stable sample, we also performed a combined-department analysis for each discipline. Initial interviews with users indicated that they would be willing to accept discipline-level results. They are less interested in summaries for broader groups such as social sciences or natural sciences or for the total group. Data for departments in a given discipline (biology, chemistry, education, English, or psychology) were pooled to compute common regression weights. The analysis assumes homoscedasticity and common regression coefficients, but allows the regression constants to differ among departments. The differences in regression constants reflect possible differences in the quality of the students enrolled and/or differences in grading standards from department to department. In this analysis, ordinary least squares estimates of the common regression weights were obtained based on pooled within-department variances and covariances for each discipline. The resulting weights (and constants) were evaluated as alternative prediction equations, and compared to the results based on the analyses for each individual department. The alternative equations provide additional information to departments whose results are unreliable because of small samples and may be informative to departments that have not been able to do an individual predictive validity study.

Results

This results section is separated into two parts. The first part, "Predicting Long-Term Outcomes of Graduate School," focuses on an overall evaluation of how well GRE scores and undergraduate grade point average predict several broad measures of success in graduate school. The second section, "Detailed Results," reports more tentative and detailed analyses, including specific prediction equations that might be used by graduate departments that did not participate in this study. The second section also includes a first look at how well GRE scores and undergraduate grade point average predict success in graduate school for women, ethnic minority students, noncitizens, master's versus doctoral degree students, and applicants who took a computer-administered GRE versus those who took a paper administration.

Results Section I: Predicting Long-Term Outcomes of Graduate School

This section covers the most important results from this study, the information on a variety of long-term outcomes of graduate school. The most basic research question is how well do GRE scores and undergraduate grade point average predict the following long-term graduate school outcomes:

- Cumulative graduate grade point average
- Faculty rating of mastery of the discipline
- Faculty rating of professional productivity

- Faculty rating of communication skills

We report correlations to summarize results for all departments combined, and for each academic discipline separately.

Results for all disciplines combined. Table 1 presents the average single and multiple correlations for the above four graduate school outcomes (or criteria) summarized over all participating departments. Three different combinations of predictors are displayed. First, Table 1 shows the correlation for the combination of the scores for the GRE verbal and quantitative and undergraduate grade point average that best predicts each graduate school outcome or criterion. The criteria are presented in order by size of correlation. Then, to facilitate comparison, the multiple correlation for GRE verbal and quantitative scores alone and the single correlation for undergraduate grade point average alone are shown in the same order. To create a reasonable summary over different disciplines and institutions with very different missions and students, we corrected each correlation for restriction of range. Both uncorrected and corrected correlations are included in Table 1.

Table 1

Average Correlations for Four Graduate School Outcome for All Departments: Combinations of GRE Verbal and Quantitative Scores and Undergraduate Grade Point Average

	Numbers		*V, Q, U*		*V, Q*		*U*	
Criterion	*Depts.*	*Students*	R^C	*R*	R^C	*R*	r^C	*r*
Mastery of discipline (FR)	11	352	0.55	0.40	0.52	0.37	0.21	0.13
Professional productivity (FR)	10	319	0.53	0.38	0.46	0.30	0.25	0.16
Communication skill (FR)	11	339	0.50	0.39	0.46	0.35	0.23	0.16
CGPA	19	1,303	0.49	0.40	0.40	0.33	0.32	0.24

Note. V = GRE verbal; Q = GRE quantitative; U = undergraduate grade point average; CGPA = cumulative graduate grade point average; FR = faculty rating; *R*=multiple correlation; R^C=multiple correlation corrected for multivariate restriction in range; *r* =correlation of one predictor with the criterion; r^c=correlation of one predictor with the criterion, corrected for multivariate restriction in range. Average correlations weighted by number of students in each department.

Table 1 also gives the number of departments and the number of students whose results are summarized for each outcome. This information indicates how generalizable the results for an individual outcome are likely to be, and also how comparable the correlations for any pair of outcomes are likely to be. Note that the outcome with the most data is cumulative graduate grade point average, available for 19 departments and 1,303 students. The faculty rating criteria have, at most, 11 departments and 352 students. The correlations for the three different faculty ratings can be compared with each other, since the number of institutions and students for all three are comparable in size and based on nearly the same individuals. The correlation for cumulative graduate grade point average is only roughly comparable to those for faculty ratings.

There are several notable points about these average correlations:

- When all three predictors are combined, the corrected correlations for all three faculty ratings are .5 or higher, correlations classified as large (Cohen, 1977).
- When all three predictors are combined, the corrected correlation for cumulative graduate grade point average rounds to .5.
- Correlations for the two GRE scores combined are nearly as high as those for all three predictors combined. Undergraduate grade point average does contribute to the prediction of all outcomes, but its greatest influence is on the prediction of cumulative graduate grade point average. The difference between the correlation for all predictors, R^C=.49, and GRE scores alone, R^C=.40, is .09. The unique contribution of undergraduate grade point average to the prediction is .09. It makes sense that undergraduate grade point average would contribute particularly well to the prediction of graduate grade point average, since they measure similar accomplishments in the same manner.
- The correlation of undergraduate grade point average alone is .32, so the GRE scores contribute .17 to the full correlation of .49.
- The correction for restriction of range has a substantial influence on most correlation coefficients. The median increase for a department is .11.

Table 2

Progress to Degree for Master's and Doctoral Degree Students: Numbers, Means, and (Standard Deviations) of GRE Scores and Undergraduate Grade Point Average

Progress to degree	N	V	Q	U
Master's degree students				
Total group				
Withdrew	103	421 (101)	422 (127)	2.95 (.58)
Not yet attained master's degree	34	483 (120)	471 (119)	3.10 (.54)
Attained master's degree	280	453 (133)	465 (127)	3.01 (.57)
Education depts.				
Withdrew	71	398 (90)	395 (123)	2.95 (.60)
Not yet attained master's degree	17	416 (90)	452 (114)	2.92 (.52)
Attained master's degree	164	386 (87)	418 (108)	2.89 (.51)
English depts.				
Withdrew	19	504 (100)	476 (117)	3.14 (.43)
Not yet attained master's degree	14	562 (115)	499 (131)	3.28 (.58)
Attained master's degree	78	603 (105)	557 (115)	3.39 (.52)
Biology, chemistry, and psychology depts.				
Withdrew	13	422 (101)	492 (123)	2.67 (.56)
Not yet attained master's degree	3	487 (76)	443 (85)	3.25 (.13)
Attained master's degree	38	433 (101)	481 (122)	2.71 (.48)
Doctoral degree students				
Total group				
Withdrew	54	573 (95)	672 (112)	3.31 (.36)
Not yet attained doctoral candidacy	95	564 (92)	652 (90)	3.43 (.38)
Attained doctoral candidacy or degree	238	587 (99)	659 (88)	3.47 (.40)
Education depts.				
Withdrew	1	---	---	---
Not yet attained doctoral candidacy				
Attained doctoral candidacy or degree	10	520 (106)	535 (96)	3.36 (45)
English depts.				
Withdrew	4	662 (90)	420 (137)	3.17 (.78)
Not yet attained doctoral candidacy	5	600 (64)	560 (91)	3.81 (.19)
Attained doctoral candidacy or degree	27	673 (80)	579 (88)	3.58 (.39)
Biology, chemistry, and psychology depts.				
Withdrew	49	567 (94)	697 (76)	3.32 (.32)
Not yet attained doctoral candidacy	90	562 (93)	658 (87)	3.41 (.38)
Attained doctoral candidacy or degree	201	578 (95)	676 (76)	3.46 (.40)

Note. V = GRE verbal; Q = GRE quantitative; U = undergraduate grade point average.

Progress to degree. Table 2 reports the stage of progress students had reached when we collected the data for our study. It displays the numbers of students in several progress categories, separated into those pursuing master's degrees and those pursuing doctoral degrees. Table 2 also displays the means and standard deviations of GRE scores and undergraduate grade point average by stage of progress. The data are given for all master's or doctoral degree students combined and then separated into three disciplinary groups: education, English, and science (combined biology, chemistry, and psychology departments). Master's degree students are separated into three progress stages: (a) those who withdrew or did not register after the second semester; (b) those who have not yet attained the master's degree, and (c) those who have. Doctoral degree students are also separated into three groups: (a) students who withdrew (defined as for master's degree students); (b) those who have not yet attained candidacy for the doctoral degree (this includes students registered for a doctoral degree who took a master's degree and left); and (c) those who have either attained doctoral candidacy or the doctoral degree. The last two groups were combined because of very large differences among departments in rate of progress after candidacy. This group probably contains students who have trouble producing a dissertation, students who are actively engaged in research with faculty, and students who have had to take jobs to support themselves, to name only a few possibilities.

It can be seen that the main differentiation in predictor scores in the table is between master's students and doctoral degree students. When master's and doctoral degree students are considered separately, results are complicated and hard to interpret. For example, among master's degree students, those who withdrew tend to be the lowest scorers. Among doctoral degree students, students who withdrew have relatively good GRE scores and undergraduate grade point average. It is unclear why this happened. It may simply be an anomaly in our sample; alternatively, it may be that master's degree programs are more likely to give marginal students a chance. For another example, the quantitative scores in science areas are higher for withdrawing students than for either other group. Is this because the students who withdrew were able to transfer to more prestigious graduate programs or get good jobs without a degree? Recall that this study was done during the roaring '90s, when industry was competing strongly for students in technical areas. These complexities illustrate why it is difficult to find high correlations between predictors and degree progress. In the discussion, we make several suggestions about how to develop better progress measures.

Discipline-specific results. The information shown for all departments in Table 1 is displayed graphically in Figures 1 through 5. The numbers used to generate these graphs are recorded in Table C1. There are separate graphs for each discipline. The graphs display three correlations for each outcome measure: (a) the multiple correlation of all three predictors (GRE verbal and quantitative scores and undergraduate grade point average); (b) the multiple correlation for the GRE verbal and quantitative combined; and (c) the single correlation for undergraduate grade point average. The same students and departments are included in all three correlations, so the results are fully comparable. All correlations displayed in the figures are corrected for restriction of range. Uncorrected correlations are reported in Table C1.

Biology. Figure 1 shows the results for 145 students in five biology departments. The pattern of correlations for biology departments is similar to the overall pattern.

- All four outcomes are predicted equally well in biology departments and all are predicted strongly.
- Undergraduate grades make a relatively small contribution to the prediction of all outcomes.

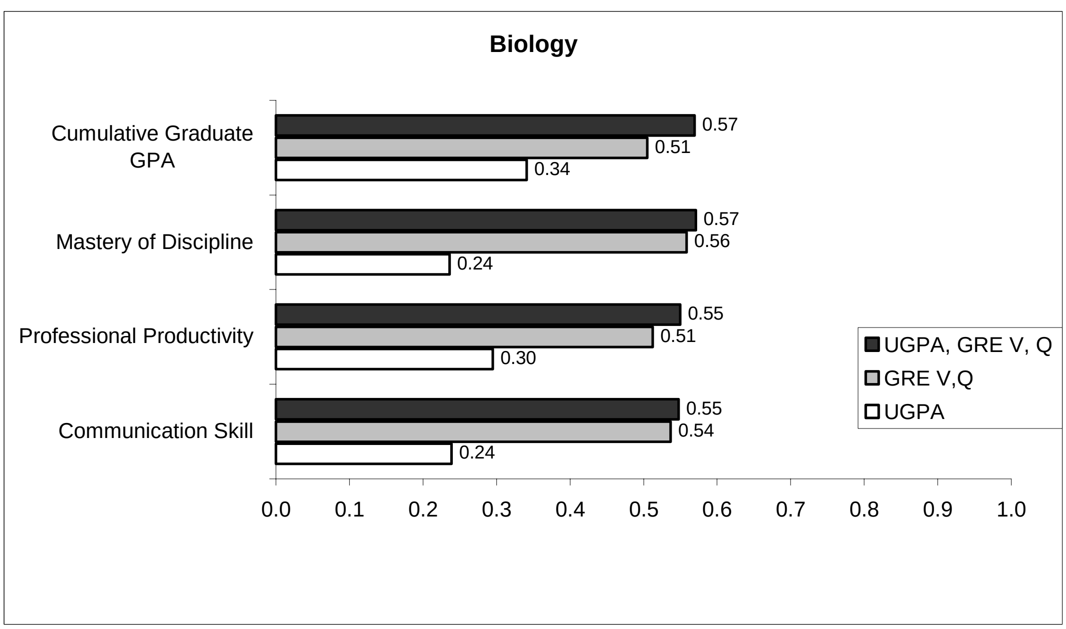

***Figure 1.* Average correlations for predictions of four graduate school outcomes in biology.**

Chemistry. Figure 2 shows the results for 134 students in two chemistry departments. The pattern of correlations for chemistry departments is very similar to the biology and overall results.

- Cumulative graduate grade point average and faculty ratings of mastery of the discipline and professional productivity are predicted best, and all three are predicted strongly.
- Communication skills are predicted moderately well.
- Undergraduate grade point average makes a stronger contribution in chemistry departments than it does in biology departments. In chemistry departments, both GRE scores and undergraduate grade point average contribute to the prediction of all graduate school outcomes. GRE scores are the primary predictor of mastery of the discipline; GRE scores and undergraduate grade point average share equally in predicting cumulative graduate grade point average and professional productivity; and undergraduate grade point average is the primary predictor of communication skills.

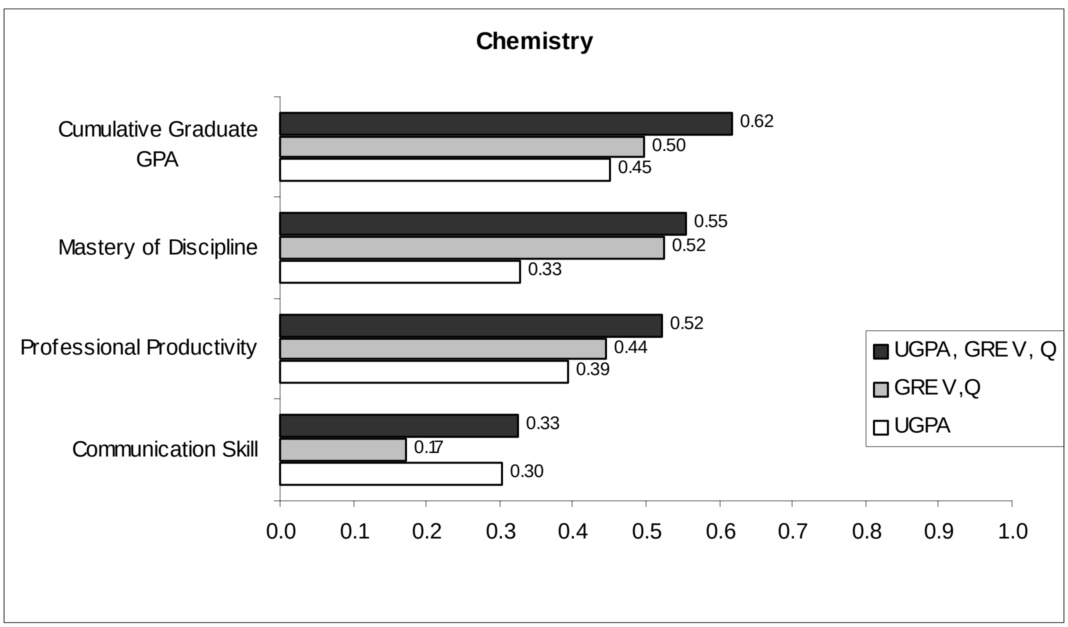

***Figure 2.* Average correlations for predictions of four graduate school outcomes in chemistry.**

Education. Figure 3 shows the results for 699 students in three education departments. Only about 1 in 10 education students were rated (83 out of 699), in part because faculty usually did not remember master's degree students who had been enrolled four or five years previously well enough to rate them. In addition, one school of education did not submit ratings. Education departments have a slightly different pattern from the two science disciplines we have been discussing.

- Despite low faculty participation, GRE scores strongly predict faculty ratings for students the faculty knows well.
- The three faculty ratings are unusually strongly predicted. GRE scores provide all of the prediction for mastery of the discipline and communications skills, and most of the prediction for professional productivity.
- Communication skills, predicted moderately for chemistry students, are strongly predicted for education students. (The same is true for biology students.)
- Cumulative graduate grade point average is predicted moderately well in education departments (it is predicted strongly in both science disciplines); GRE scores and undergraduate grade point average contribute equally to the prediction.

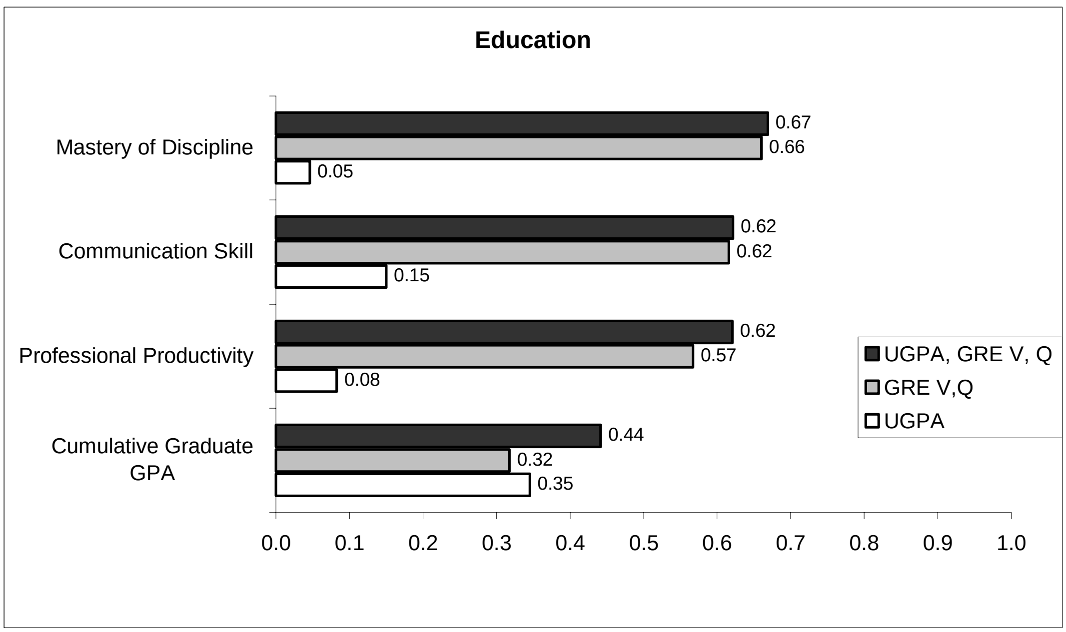

***Figure 3.* Average correlations for predictions of four graduate school outcomes in education.**

English. Figure 4 shows the results for 170 students in five English departments. The pattern of correlations for English departments is similar to the pattern observed in education departments.

- As we observed in education departments, all three faculty ratings are predicted well in English departments.
- Cumulative graduate grade point average is predicted better in English departments than in education departments, but not quite as well as in science departments.
- GRE scores are particularly important predictors of success in English departments. The faculty ratings are predicted entirely by GRE scores; undergraduate grades make a moderate contribution to predicting cumulative graduate grade point average.

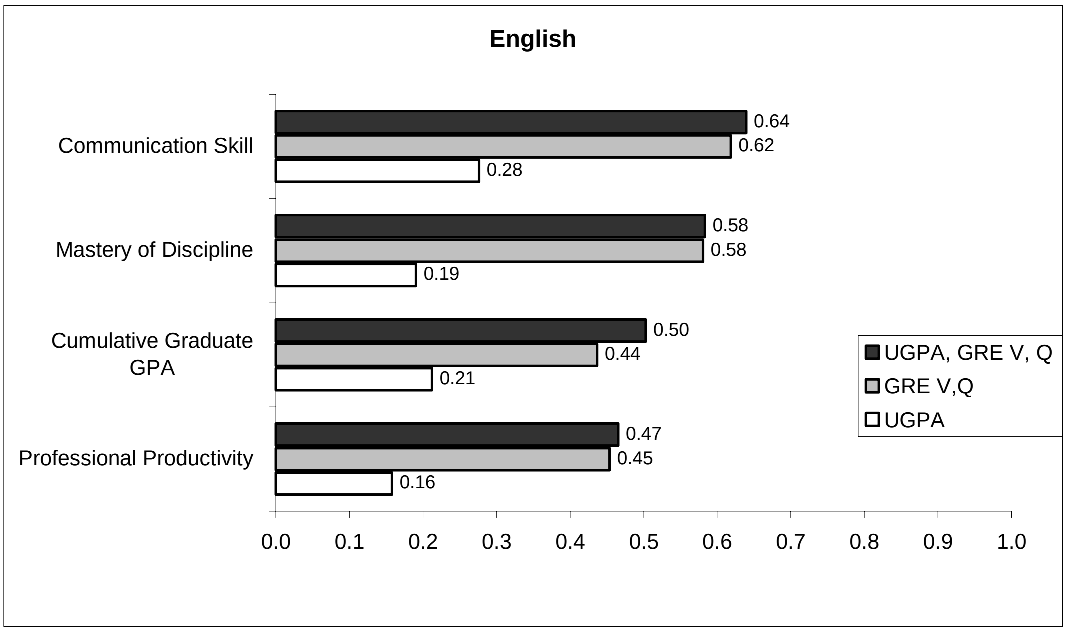

Figure 4. **Average correlations for predictions of four graduate school outcomes in English.**

Psychology. Figure 5 shows the results for 155 students in four psychology departments. The pattern for psychology departments most resembles that for the biology and chemistry departments.

- As we observed in natural science departments, cumulative graduate grade point average is the outcome that is predicted best in psychology departments. It is the only outcome that is predicted strongly.
- The three faculty ratings are predicted moderately well, with professional productivity predicted best of the three.

- In psychology departments, both GRE scores and undergraduate grade point average contribute to the prediction of all outcomes, although the contribution of undergraduate grade point average to predicting communication skill is very small.

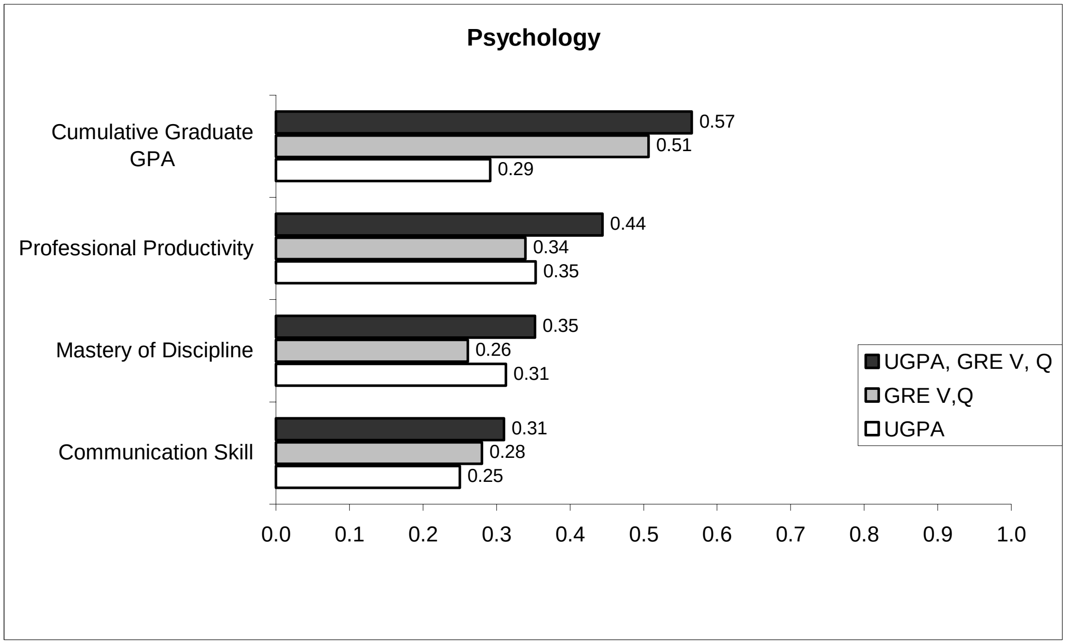

Figure 5. **Average correlations for predictions of four graduate school outcomes in psychology.**

Results Section II: Detailed Findings

In the following two sections we present the more detailed results of the study. The first section presents equations that can be used to predict cumulative graduate grade point average. We focus on cumulative graduate grade point average because it is available for more students (1,310) and more departments (19) than any other outcome measure. In the second section of detailed analysis, we discuss validity results for various important subgroups of the graduate school population, including men and women; ethnic minority students (African American, Asian American, Hispanic, and White); students who are U.S. citizens and those who are citizens of other countries; master's and doctoral degree candidates; and, finally, those who took the computer-adaptive GRE versus those who took the paper-and pencil-version.

Equations Predicting Cumulative Graduate Grade Point Average

Prediction equations are based on that part of a predictor measure that is related to a valued outcome. Verbal reasoning, for example, is related to performance in graduate school especially when students are learning new content, when they are organizing or reorganizing a conceptual system, and when they are communicating what they have learned. (See, for example, Burton, Welsh, Kostin, & Van Essen, in press; Glaser, 1984; Goody, 1977; Nist & Simpson, 2000; Wagner & Stanovich, 1996.) Verbal reasoning is probably not closely related to a student's willingness to do assignments on time, to attend classes, or to the student's interest in and commitment to the discipline. This does not imply there is anything wrong with verbal reasoning as an admission measure, but that other admission measures are necessary if responsibility and dedication are important aspects of success in graduate school. A prediction equation combines the various numerical measures available at admission so as to predict an outcome as accurately as possible. In this case, we will be using GRE verbal and quantitative scores and undergraduate grade point average, each multiplied by its own *regression weight*, to predict cumulative graduate grade point average.

Because many individual department regression weights are unstable, we have computed regression equations for combined departments. Data from the departments in the same discipline were pooled to compute common regression weights, while the regression constants (intercepts) were allowed to vary across departments. These analyses are less sensitive than individual department analyses to random variations. Because the pooled analysis is more stable, we would expect it to apply to subsequent entering classes better. However, a cross-validation study would be needed to determine whether it does.

Table 3 shows the pooled regression weights in all five disciplines. These weights can be used to compute a predicted cumulative graduate grade point average for any applicant with GRE verbal and quantitative scores and undergraduate grade point average. They can be used by any department to check the results of an individual study based on a small number of students, or on an atypical sample of students. They can also be used by departments that have not yet done an individual validity study and would like to profit from the knowledge about selecting graduate students gained by other departments in their discipline.

Table 3

Predicting Cumulative Graduate Grade Point Average: Pooled Department Analysis

	Biology	*Chemistry*	*Education*	*English*	*Psychology*
Number	*145*	*134*	*701*	*175*	*155*
Regression weights					
U	0.164	0.245	0.170	0.021	0.048
V	0.116	0.056	0.116	0.198	0.065
Q	0.107	0.193	0.008	0.054	0.042
Standard error of estimate	0.301	0.308	0.287	0.211	0.176
R pooled over departments					
Multiple R	0.30	0.38	0.35	0.44	0.26
Corrected multiple R (R^C)	0.48	0.59	0.40	0.55	0.37
Weighted average department R^C					
Recommended equation	0.59	0.62	0.44	0.46	0.54
Full equation (V, Q, and U)	0.57	0.62	0.44	0.50	0.57
Mean CGPA	3.62	3.49	3.69	3.75	3.83
SD CGPA	0.313	0.328	0.311	0.234	0.180

Note. CGPA = cumulative graduate grade point average. R=multiple correlation; R^C=multiple correlation corrected for multivariate restriction in range. GRE verbal (V) and quantitative (Q) scores were divided by 200 to reduce the number of decimal places required for regression weights. Pooled estimates include departments below minimum sample size for separate analysis. Recommended equation: highest correlation with no negative regression weights. Note that the recommended equation would usually have a correlation either lower than or equal to the full equation. The slightly higher correlation of the recommended equation for biology (.59, compared to .57 for the full equation), is possible because both are corrected for multivariate restriction of range.

Table 3 also displays multiple correlations for the pooled analysis and, in comparison, average multiple correlations from individual department analyses. In general, the correlations from the pooled analysis are somewhat lower than the average individual correlations, except in English departments, where the pooled corrected multiple R (.55) is slightly higher than the weighted average corrected multiple R (.50). The corrected pooled correlation for psychology (.37) is much lower than the average for the corrected individual correlations (.57). This suggests that the pooled analysis for psychology does not fit the data as well as the within-department analyses. This is possible, given that psychology departments may have quantitative, clinical, experimental, social, or cognitive orientations, which might call for different mixes of skills. It is also true that average grades are higher (3.83 is the mean grade) and have less variation (.18 is the standard deviation) in psychology departments than in the others studied, making graduate grades a narrow, elusive target to predict.

In addition to computing pooled results, we also developed simple rules for specifying a recommended equation computed for individual departments. While there are reasonable explanations for negative weights, it does not make sense to use them in actual admission decisions. Our rule discards any predictor with a negative weight because each measure was considered to be positively related to success and had a positive single correlation with the criterion. We recommend the equation with the highest correlation and no negative weights. Table 3 shows two weighted averages of individual department multiple correlations: one for the recommended equation and one for the full set of predictors. It can be seen that, in general, the recommended equation has essentially the same average correlation as the full three predictor equation. In two of the five disciplines, the average correlations are the same; in one, the recommended equation has a slightly higher average correlation, and in two, the recommended equation has a slightly lower average correlation. The similarity suggests that negative weights do not make an important contribution to prediction.

Predicting Graduate School Outcomes for Subgroups

In this section, we provide an evaluation of the fairness of undergraduate grade point average and GRE scores for several subgroups of the graduate school population. The questions we will discuss are:

- Can graduate school outcomes be predicted equally strongly for various groups? That is, when separate equations are computed for two groups of interest, are the correlations comparable?
- When a single equation is used for combined subgroups, are the predictions fair for all subgroups? That is, do predicted outcomes tend to be systematically lower or higher than the actual outcomes for certain groups?

Because the number of students in a given subgroup is often small, we analyze data only for the most frequently available outcome, cumulative graduate grade point average. Because departments differ in a number of important ways, we required that any comparison be made within a single department. Thus, a department's data was analyzed only if the sample was sufficient for at least one focal group *and* a comparison group. In the ethnic group analysis, the comparison group was always White domestic students. For gender groups, most departments were included in the analysis (13 of 19), so we feel nearly as comfortable discussing the gender results as we do about the overall study results. For the other subgroups, however, we really can only say that the correlations are or are not comparable in the departments analyzed. We cannot infer what the results might have been in other departments. This study allows us to begin to accumulate information about how well conventional predictors work for subgroups, as it was proposed to do. However, more data will have to be accumulated before we can begin to draw conclusions. More data are needed to represent the graduate community adequately, and more data are needed to achieve stable, dependable results.

Specific subgroup analyses follow. We will answer analysis questions about both the strength of correlations and the fairness of predictions for a particular group before moving on to the next group. The demographic groups—gender, ethnic group, and citizenship—will be analyzed first.

Gender comparisons. Figure 6 shows average correlations by discipline for men and women. These correlations are all high. Only one, for men in education departments, does not round to at least .5, which is considered to be a large correlation. The results for men and women are comparable. In two disciplines, the men's coefficients are higher, while in the other three, women's coefficients are higher.

The second question that we wish to pursue about prediction of graduate school success for men and women has to do with the fairness of using the same selection rules for men and women.

If you predict success using the same measures, with the same weightings, what is the typical result? For years, researchers have found that when a formal regression equation is applied to both men and women, women tend to get slightly higher grades than predicted (this is called *underprediction)*, while men tend to get slightly lower grades than predicted (this is called *overprediction)*. This has been found in undergraduate, graduate, and professional schools. See, for example, Linn (1982) and Willingham and Cole (1997). Quite a bit of research has been done on this topic. One explanation is that men and women take a different allocation of courses—men more frequently take math and science courses that tend to be graded stringently, while women more frequently take humanities and social science courses that tend to be graded more liberally. If coursework is held constant by analyzing within discipline or, even better, within individual courses, much of the gender difference in prediction disappears. Further gender differences are accounted for by the fact that women tend to have better *studenting* skills than do men; for example, they attend classes and read assignments more frequently than men (Stricker, Rock, & Burton, 1993; Willingham & Cole, 1997).

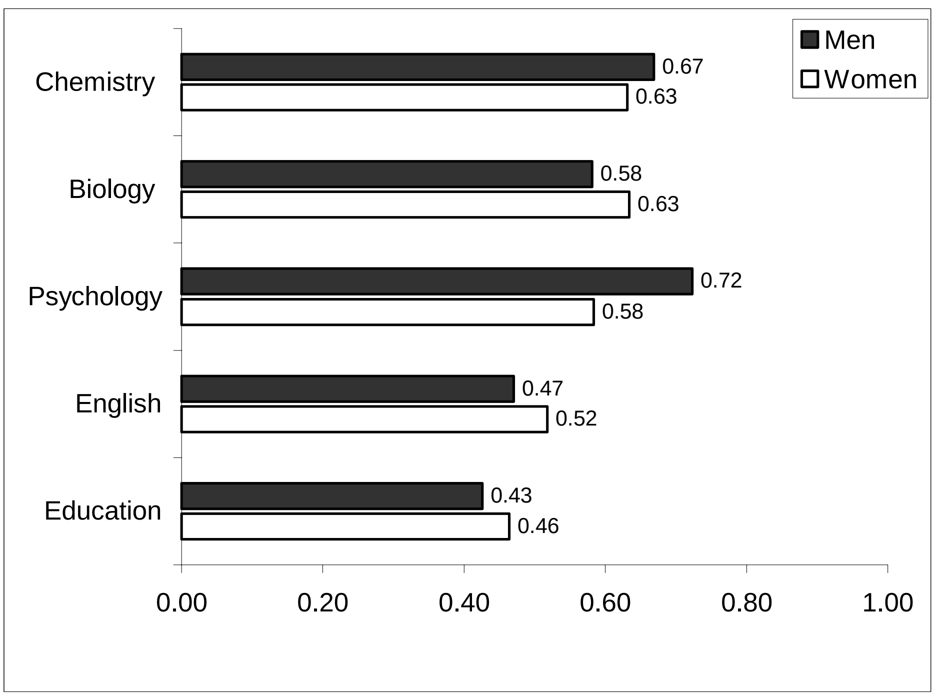

Figure 6. **Average correlations for predictions of cumulative graduate grade point average for women and men.**

Because of the differences in course taking patterns generally observed for men and women, we might expect to find in this study the typical pattern of overprediction and underprediction in graduate school when grades are combined over different graduate disciplines. However, we were not completely sure what we would find in this study, either for individual department analyses or for analyses pooled over all departments in a discipline, and hence, we will look at data on the difference between a person's actual earned grade point average and the same person's predicted grade point average. A negative difference means that the predicted grade was higher than the actual grade: the person's grade was overpredicted. Other things being equal, this is an advantage in admission, since the admission committee believes that the person will do somewhat better than he or she actually will. A positive difference means that the predicted grade was lower than the actual grade: The grade was underpredicted. If our data follows the traditional pattern, the average difference between observed and predicted grades will be negative for men and positive for women.

Table 4 presents the average observed minus predicted difference for men and for women in each of the five disciplines, and Figure 7 presents the information visually. Because predicted grades are based on total group equations, we do not have the sample size problem we encountered when computing separate equations for men and women, so we are able to report differences for the full dataset of 1,300 students in 19 departments. There are small average differences between men and women, mostly in the expected direction. Men's grades are overpredicted in all disciplines but English; note, however, that men on average receive higher grades in English. In all other disciplines, women receive higher grades. The amount of underprediction for women (or men) is very small. Overall, women's grades are underpredicted by one one-hundredth of a grade point. In other words, the average woman who is predicted to get a 3.00 cumulative graduate grade point average actually gets a 3.01 cumulative graduate grade point average. The largest average underprediction occurs in chemistry departments, where women's cumulative graduate grade point average is underpredicted by six one-hundredths of a grade point. None of these differences is practically significant, and the differences would not be worth mentioning if they were not consistent with a great deal of previous data. Table C3 gives overprediction and underprediction information by department. Note that results are somewhat inconsistent at the departmental level.

Table 4

Average Overprediction (-) or Underprediction (+) of Cumulative Graduate Grade Point Average for Men and Women Students

			Means		
		N	*Over/under-prediction*	*CGPA*	*SD CGPA*
Biology	Men	67	–0.037	3.58	.33
	Women	78	0.032	3.66	.30
Chemistry	Men	92	–0.029	3.47	.33
	Women	42	0.064	3.57	.32
Education	Men	193	–0.026	3.67	.34
	Women	506	0.010	3.70	.30
English	Men	64	0.024	3.78	.23
	Women	106	–0.014	3.74	.23
Psychology	Men	56	–0.026	3.81	.19
	Women	99	0.015	3.84	.17
Total	Men	472	–0.022	3.65	.33
	Women	831	0.012	3.71	.28

Note. Overprediction and underprediction computed by subtracting cumulative graduate grade point average predicted using the recommended equation (the highest correlation with no negative regression weights) from observed cumulative graduate grade point average. Average over-/underprediction weighted by the number of students in each department.

Ethnic group comparisons. The next results we will discuss are for ethnic minority group performance as compared to White performance. We have followed GRE program policy and classified only domestic U.S. students by ethnic group. Only the large education departments in three participating universities had the minimum required samples of both minority and White students. Table 5 summarizes the results of computing separate prediction equations for the ethnic groups with nine or more students in these education departments in three institutions. In total, about 350 White students, 130 African American students, 70 Asian American students, and 70 Hispanic American students were available for analysis. They represent over 600 of the 700 education students included in this study. The correlations for students in comparable situations (i.e., in the same graduate department) were quite comparable across ethnic groups. The one very high correlation, for African American students in Institution B, was for a group of 9 students, the very lowest number we would analyze. Note that this correlation of .84 was only slightly adjusted by the correction for shrinkage to .72. We believe that this more likely means that the shrinkage was inadequate than that the correlation is correct.

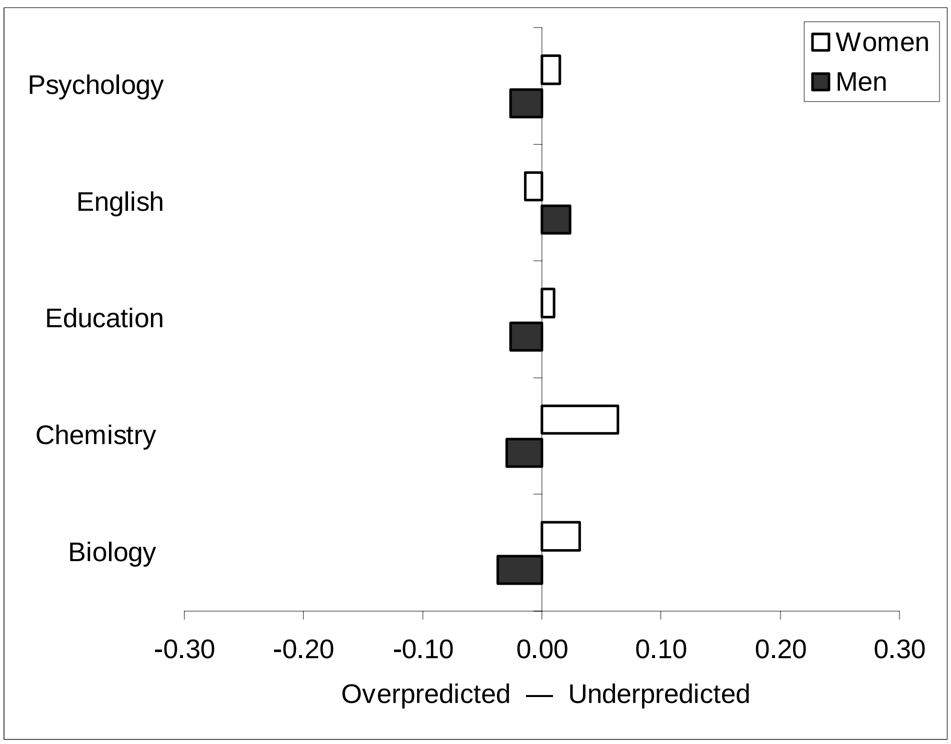

Figure 7. **Over- and underprediction of women and men.**

Table 5

Ethnic Groups in Three Education Departments: Multiple Correlations of GRE Verbal and Quantitative Scores and Undergraduate Grade Point Average With Cumulative Graduate Grade Point Average

		White	*African American*	*Asian American*	*Hispanic American*
Education Dept. A	Number	212	35	66	70
	Multiple R	0.29 (0.27)	0.33 (0.16)	0.30 (0.22)	0.38 (0.32)
	Corrected Multiple R	0.38	0.40	0.39	0.46
	Mean CGPA	3.75	3.55	3.67	3.70
	SD CGPA	0.29	0.27	0.36	0.28
Education Dept. B	Number	116	9		
	Multiple R	0.42 (0.39)	0.84 (0.72)		
	Corrected Multiple R	0.44	0.86		
	Mean CGPA	3.77	3.48		
	SD CGPA	0.27	0.53		
Education Dept. C	Number	19	85		
	Multiple R	0.50 (0.32)	0.38 (0.34)		
	Corrected Multiple R	0.44	0.57		
	Mean CGPA	3.85	3.49		
	SD CGPA	0.15	0.27		

Note. CGPA = Cumulative graduate grade point average. Multiple correlations reported uncorrected and corrected for multivariate restriction of range. The multiple correlation tends to be overestimated when samples are small. Correlations in parentheses corrected for shrinkage (Pedhazur, 1997, p. 208), which adjusts for capitalization on chance, but it can reduce correlations to less than zero.

The second question we will discuss is how students in the various ethnic groups fare when the same prediction equation is applied to all students in a department. Since this analysis is based on the total group prediction equation for each department, we are able to report results for the scattered small numbers of ethnic minority students in all departments. As for the gender group analysis, we look at the difference between the graduate school grades that are predicted by each department's equation and the actual grades attained by students in that department. We then average these differences by ethnic group to look for any systematic over- or underprediction. (Note that the amount of over- or underprediction observed depends, in part, on group size. Because in the total group the differences sum to zero, the average of over- and underpredictions for all groups, weighted by the size of each group, will also sum to zero. In general, large groups, necessarily close to the mean, have small average differences, while small groups can have quite large differences.)

Figure 8 displays the average over- or underprediction for African American, Asian American, Hispanic American, and White students in each of the five disciplines. Table C3 documents the numbers used in creating the figure. The non-White groups are small, except in education. There are about 20 African American students each in biology and English, and about 20 Hispanic American students in psychology—all other groups are smaller. The education results are the best guide; the other department results tend to confirm the direction of the differences between observed and predicted, but may exaggerate their size. In education departments, graduate grades tend to be slightly overpredicted for African American and Asian American students, and slightly underpredicted for Hispanic American and White students. African American students' grades are consistently overpredicted (except in biology). The tendency to overpredict African American students' grades is also observed for undergraduates (Bowen & Bok, 1998; Jencks & Phillips, 1998; Ramist et al., 1994). Hispanic American students' grades, underpredicted in education, are overpredicted in English and chemistry, and right at zero in psychology departments, leaving any general trend in doubt.

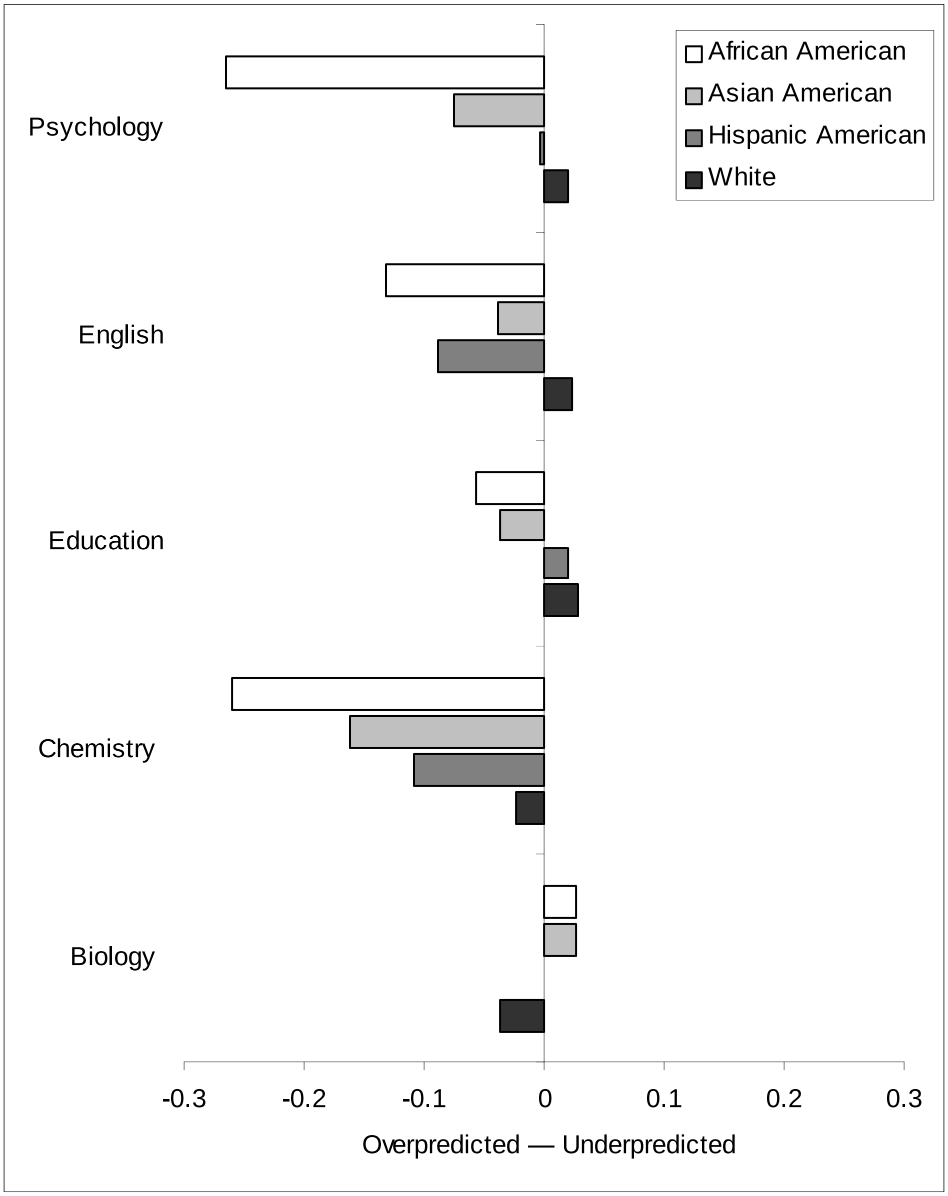

Figure 8. **Over- and underprediction for African American, Asian American, Hispanic American, and White students.**

In summary, the small amount of data available in this study on ethnic minority students suggests that GRE scores and undergraduate grade point average provide similar predictive information for all groups. Correlations are about the same size for White and minority students in the same department. While the overprediction results look large for African American students, they are based on very small groups of students. In education departments where there was a large sample of African American students, their grades are also overpredicted on average, but by a relatively small amount—by six one-hundredths of a grade point.

Students who are not U.S. citizens. The final demographic groups we analyze are U.S. citizens compared to noncitizens.[4] Noncitizenship is a proxy for a possible lack of familiarity with U.S. culture and education, and it is associated with (but not identical to) being a nonnative speaker of English. Table 6 summarizes results for three departments in a university with many international students. An adequate sample size was available in the biology, chemistry, and education departments to give results for both citizens and noncitizens. The corrected correlations are large for both citizens and noncitizens in the biology and chemistry departments, and moderate for both groups in the education department.

Table 6

Citizens and Noncitizens in Three Departments: Multiple Correlations of GRE Verbal and Quantitative Scores and Undergraduate Grade Point Average With Cumulative Graduate Grade Point Average

	Biology		*Chemistry*		*Education*	
	Citizen	*Noncitizen*	*Citizen*	*Noncitizen*	*Citizen*	*Noncitizen*
Number	23	35	23	26	400	50
Multiple R	0.36 ([a])	0.41 (0.30)	0.53 (0.40)	0.35 (0.06)	0.31 (0.30)	0.42 (0.36)
Corrected multiple R	0.51	0.48	0.73	0.56	0.39	0.31
Mean CGPA	3.67	3.66	3.39	3.59	3.70	3.65
SD CGPA	.37	.21	.29	.31	.31	.35

Note. CGPA = Cumulative graduate grade point average. Multiple correlations reported uncorrected and corrected for multivariate restriction of range. The multiple correlation tends to be overestimated when samples are small. Correlations in parentheses corrected for shrinkage (Pedhazur, 1997, p. 208), which adjusts for capitalization on chance, but it can reduce correlations to less than zero (see note below).

[a] The correction for shrinkage reduced the estimated R^2 to -.01.

Because the data in this study come from three departments in a single institution, we will not attempt to generalize about the specific contribution of verbal versus quantitative measures to prediction for international students. However, there have been several large and representative studies reported recently. SAT® data on undergraduates (Burton & Cline, in press) and GRE data on graduate students collected prior to this study (Wang, 2002) show that verbal scores contribute to prediction for nonnative speaking students, although not as strongly as for native speakers. Wang's results are based on data from GRE validity studies conducted between 1987 and 1991, by 468 departments enrolling 8,281 students. These validity data, originally analyzed using empirical Bayes[5] methods, were recomputed using the same methods as this study. Because Wang's results are not available elsewhere, they are reprinted in Tables C5 by discipline, C6 by gender, and C7 for students whose best language is, or is not, English. Kuncel et al. (2001) report single correlations with first year graduate grade point average that are higher for GRE quantitative scores than for GRE verbal scores for nonnative speakers of English. We analyzed over- and underprediction for these students as well, but there were no substantial results, so we did not produce an over- and underprediction figure like Figures 7 and 8. Noncitizens were slightly overpredicted (by three one-hundredths of a grade point); citizens were underpredicted by less than one one-hundredth of a grade point. Table C4 documents the negligible under- and overprediction results for citizenship and the two next analysis categories, described below.

Next, we will present analyses for two different kinds of subgroups—master's degree students compared to doctoral degree students, and students who took a computer-adaptive GRE compared to those who took the paper-and-pencil version of the test. This comparison was possible because the students included in the study generally took the GRE in 1994, 1995, or 1996, while the GRE was in transition to computer delivery.

Degree level. The final analysis sample for cumulative graduate grade point average contained 639 master's degree and 664 doctoral degree students. Most departments had almost exclusively one level of student. Only four departments had the minimum sample size to compute results for both master's and doctoral degree students. Two are English departments and two are education departments. Table 7 displays the corrected and uncorrected multiple correlations of GRE verbal and quantitative scores and undergraduate grade point average with cumulative graduate grade point average for master's and doctoral degree students in these four departments. The corrected correlations are what we have come to expect in general, with one exception—high correlations for English departments, moderate correlations for education departments. The correlations are quite similar for master's and doctoral degree students, except in English department 2. For the small group of doctoral degree students, the correlation is extremely high (.9), but for master's degree students it is low (.1). For both groups of students, the verbal score has a small negative correlation, suggesting that this is an unusual group of English graduate students.

Table 7

Master's and Doctoral Degree Students in Two Education and Two English Departments: Multiple Correlations of GRE Verbal and Quantitative Scores and Undergraduate Grade Point Average With Cumulative Graduate Grade Point Average

		Education		*English*	
		Master's	*Doctoral*	*Master's*	*Doctoral*
Department 1	Number	127	11	22	23
	Multiple R	.47 (.45)	.34 ([a])	.62 (.53)	.68 (.61)
	Corrected multiple R	.48	.47	.73	.79
	Mean CGPA	3.74	3.94	3.60	3.75
	SD CGPA	.31	.17	.20	.18
Department 2	Number	248	205	50	12
	Multiple R	.34 (.32)	.32 (.30)	.10 ([a])	.76 (.65)
	Corrected multiple R	.44	.36	.12	.92
	Mean CGPA	3.67	3.73	3.81	3.90
	SD CGPA	.35	.25	.16	.09

Note. CGPA = Cumulative graduate grade point average. Multiple correlations reported uncorrected and corrected for multivariate restriction of range. The multiple correlation tends to be overestimated when samples are small. Correlations in parentheses corrected for shrinkage (Pedhazur, 1997, p. 208), which adjusts for capitalization on chance, but it can reduce correlations to less than zero (see note below).

[a] The correction for shrinkage reduced the estimated R^2 to -.27 (Ed. Dept. 1) and -.05 (Eng. Dept 2).

Similar to our earlier discussion of citizenship, there was no substantial over- or underprediction by degree level, and so no overprediction/underprediction figure was produced. The average is plus or minus one one-hundredth of a grade point (see Table C4).

Delivery mode. In our final analysis sample, approximately 1,400 students have GRE scores. In most analyses, we used the GRE scores supplied by the institutions, since those were the scores used in making admission decisions. For this analysis, however, it was necessary to use scores from the GRE file in order to be certain whether the score was earned on a computer-delivered test or a paper-and-pencil test. We found 256 students with scores from computer-delivered tests and 867 students with scores from paper-and-pencil tests in ETS files. Please note that this sample of GRE scores is different from all other analysis samples in this report. Table 8 summarizes correlations for three education departments.

Table 8

Computer and Paper-and-Pencil Test Delivery in Three Education Departments: Multiple Correlations of GRE Verbal and Quantitative Scores and Undergraduate Grade Point Average With Cumulative Graduate Grade Point Average

	Education dept. A		*Education dept. B*		*Education dept. C*	
	Comp.	*Paper*	*Comp.*	*Paper*	*Comp.*	*Paper*
Number	40	91	75	224	42	66
Multiple R	0.48 (0.40)	0.48 (0.45)	0.36 (0.30)	0.39 (0.37)	0.71 (0.68)	0.43 (0.38)
Corrected multiple R	0.53	0.48	0.44	0.46	0.83	0.52
Mean CGPA	3.79	3.73	3.69	3.71	3.47	3.63
SD CGPA	0.27	0.32	0.32	0.31	0.25	0.31

Note. CGPA = Cumulative graduate grade point average. Scores taken from GRE files as follows: Highest computer test score; if no computer test score, highest paper test score. Multiple correlations reported uncorrected and corrected for multivariate restriction of range. The multiple correlation tends to be overestimated when samples are small. Correlations in parentheses corrected for shrinkage (Pedhazur, 1997, p. 208), which adjusts for capitalization on chance, but it can reduce correlations to less than zero.

The corrected correlations in Table 8 are all large; all but one round to .5 or higher. With one exception, the correlations are also quite comparable for the computer-delivered and paper-and-pencil test takers within each department. The very high correlation for computer test takers in Department C has no immediate explanation. It is based on 42 students; not an unusually small sample, but smaller than desirable in a study using three predictors. The correction for shrinkage (from .71 to .68) does not suggest an explanation for this result.

Finally, there is little under- or overprediction: Computer test takers are slightly overpredicted (by three one-hundredths of a grade point). See Table C4 for these results.

In summary, the subgroup analyses provide baseline evidence that needs to be supplemented by further study. Subgroup analyses seek to determine whether, for example, the correlation for a subgroup of interest is as high as the correlation for some reference group. Because graduate departments, even within the same discipline, differ from each other, these comparative analyses are best interpreted within a single department. This is true even for gender groups, although we did risk presenting a cross-department summary of correlations for men and women in Figure 6. (The gender results reflected the total group results by discipline pretty well, which provide support for that decision.) Few graduate departments are large enough to allow subgroup comparisons for variables other than gender, so the subgroup data we report are limited. The evidence does support the appropriateness of using GRE scores and undergraduate grade point average to predict academic success for the subgroups studied. Within a department, correlation coefficients are of comparable size. Over the departments studied, over- or underpredictions tend to be small. A plausible start has been made in collecting evidence about the appropriateness of using GRE scores together with undergraduate grade point average to select women, ethnic minority students, international students, and master's as well as doctoral degree students. GRE scores for tests administered by computer appear to be as useful as those administered on paper.

Discussion

This collaborative validity study provides up-to-date information about the predictive validity of GRE verbal and quantitative scores and undergraduate grade point average. The design provides an enhanced set of outcome measures designed to assess those skills and qualities that are most valued by the graduate community today. The study sample includes a small but diverse group of institutions and coverage of several disciplines that attract large numbers of graduate applicants and require a wide variety of knowledge and skills. We would like to discuss what we have learned, and some of the issues that still remain, about predicting success in graduate school.

Earlier, we presented the top five qualities and skills of successful graduate students mentioned by GRE users:

- Persistence, drive, motivation, enthusiasm, positive attitude
- Amount and quality of research or work experience

- Interpersonal skills/collegiality
- Writing/communication
- Personal and professional values and character, such as integrity, fairness, openness, honesty, trustworthiness, consistency (Walpole et al., 2002, p. 14)

This list certainly supports the comments in the introduction about graduate grade point average not being the most important outcome of graduate school. Indeed, the list is rather remarkable for its lack of academic accomplishments. Perhaps because academic accomplishments are carefully screened at admission, excellent academic performance is assumed. Furthermore, students who get grades below B are soon persuaded to leave.

We have learned that GRE scores and undergraduate grade point average do predict a variety of outcomes of graduate school. Recent studies by Kuncel et al. (2001) and Wang (2002) show that first-year graduate school grades are predicted strongly when studies done at different universities are adjusted to be comparable. This study found that these earlier trends can be extended to students who are just now receiving graduate degrees. This study showed that cumulative graduate grades can also be strongly predicted. Key professional skills of graduate students, including their mastery of the discipline, their potential for professional productivity, and their ability to communicate what they know are predicted strongly by GRE scores and undergraduate grade point average. More limited data on subgroups indicate that prediction of long-term success in graduate school is good for women and men, ethnic minority students and White students, citizens and noncitizens, master's and doctoral degree students, and students who took the GRE by computer and those who took the pencil-and-paper version.

Suggestion for the future. The largest problem with our attempt to study long-term success in graduate school is that faculty were not willing, or not able, to rate most of the students included in the study. Only about 25 percent as many students with complete predictor data were rated (350) as had graduate grade point average (1,300). Several departments did no ratings at all, and most others did not rate all of their students. Departments did not always submit ratings from two different faculty members for each student on each rating measure. Two ratings were requested both to improve the reliability of the ratings, and to allow us to estimate the reliability of the raters. Asking faculty to rate students who enrolled four or five years previously is probably not the best strategy. This sort of effort might be more successful if it were undertaken as part of a longitudinal study monitoring the ongoing progress of a group of graduate students.

Of the three ratings, mastery of the discipline appeared to be the most satisfactory in that its results are strong and consistent across disciplines. Although knowledge did not appear on the list of the top five outcomes mentioned by deans and faculty, the kind of deep knowledge included in the definition of mastery of the discipline is likely to require the motivation and enthusiasm mentioned in the top-ranked persistence category.

Our definition of professional productivity more directly includes, among other things, the persistence most highly valued by the graduate community. However, this measure appears to have the strongest results in disciplines that involve empirical research. The definition of professional productivity was originally conceived of as a measure of research productivity and then generalized to fit the scholarly and applied work done in disciplines that do not typically engage in empirical research. The origin of this variable may be one reason for its greater importance in the three science disciplines included in this study. There may also have been differences in how different departments within a discipline define productivity; for example, one would expect different views of this variable in psychology departments that train professional counselors than in departments that train school psychologists who would be mainly concerned with testing duties, which might differ from departments training academics or researchers.

Our definition of communication skill combines both interpersonal skills and communication, and, in addition, basic standard English for nonnative speakers: the ability to judge the needs of one's audience; a mastery of the language of the discipline; a mastery of standard English; and the ability to communicate and work cooperatively with others. The somewhat overloaded definition of communication skills may account for its inconsistent performance as an outcome in prediction equations. Also, departments may differ in the extent to which they recruit students with existing communications skills and/or train their students in communication as part of their graduate program. The measure of communication skills used in this study deserves further refinement and simplification. Measures of communication skills should probably separate the concepts of interpersonal skills and collegiality from the academic skills of reading, writing, listening, and speaking. Furthermore, the communication skills of international students may better be treated separately from the skills of native English speakers. The original measure appeared to work best in English and education departments, the two most verbally oriented disciplines. These are both areas where communication skills are important and are very likely to be a prominent part of the graduate curriculum.

Besides the specific problems of the three different ratings, there is a general problem with the reliability (or consistency) of rating measures based on a single, global question, and based only on one or two separate ratings. Further research is needed to study faculty ratings, and perhaps to develop sets of several questions or observations that would more reliably measure these valued outcomes. Unreliable criterion measures cannot be predicted as well as reliable measures. Therefore, they underestimate the validity that predictors would have if the outcome could be measured cleanly.

A measure of progress to degree also needs further work. Our efforts to define progress to degree did not provide a reasonable outcome measure for a prediction study. This is a regrettable result, since degree attainment is one of the first and most obvious measures of success in graduate school. The research literature, however, consistently shows that it is difficult to predict which students will attain a degree, undergraduate or graduate. See, for example, Burton and Ramist (2001), Kuncel et al. (2001), and Willingham (1985). Our descriptive analysis of degree progress showed how complex the data are. Part of the problem is the long lag time between enrollment and degree attainment, especially for doctoral degree students. Only about one quarter of the doctoral degree candidates in our study had graduated four or five years after entry. Furthermore, the students in this study were not well matched on their length of enrollment.[6]

Degree attainment can be difficult to predict if it is essentially an oversimplified true/false question (did graduate/did not graduate), since such a stark distinction poorly captures a complicated process. Kuncel et al. (2001), for example, report an average corrected single correlation of .18 for GRE verbal and .20 for GRE quantitative in their meta-analysis of graduate admissions. Wilson (1978, 1980) demonstrated somewhat stronger correlations of predictors with a seven-point scale of levels of education reached by undergraduates, from *returned for sophomore year* to *enrolled in graduate or professional school*. We attempted something of the same nature by combining stages of degree progress, but had only moderate success. The main difficulty was practical. Few programs today require an orderly progression from bachelor's to master's to doctoral degree; instead, many master's degree programs are ends in themselves, and many doctoral degree programs do not require an intermediate master's degree.

The measurement of progress to degree could be improved in a number of ways. Graduate departments or graduate schools may actually possess much better information about degree progress, for example in the data they use for accreditation evidence. Or graduate departments could be asked for the total number of credits required in their degree program, and the number of credits attained by each student. Our recommendation is that progress to degree needs to be measured very carefully, and with full consultation with the institution, to make sure that the data are the best available and that the researchers understand what they have. We also suggest that the data would best be collected longitudinally, since tracking students who may have slowly faded from the program is difficult retrospectively. Many departments have graduate student coordinators or oversight committees responsible for monitoring the progress of current students; such information, unlikely to be permanently maintained, would be available to a longitudinal study. Also, we suggest that collecting students' perceptions of their own progress could be revealing.

Among the outcomes mentioned in user interviews that have not been used in validity studies as far as we know, one that seems worth developing, is a measure of pertinacity. The interviews suggested that such a measure would be welcome to the graduate community. While the literature on persistence in graduate school suggests that it is mainly determined by external factors such as funding and family support (Kyllonen, Walters, & Kaufman, in preparation), personal pertinacity is a psychological trait that may also affect graduate school completion. This is a promising variable. Willingham (1985) showed that a measure of *follow through*, defined as a student's continuing successful effort in two or more extracurricular activities in high school, is a good predictor of leadership and accomplishments in undergraduate school. It is possible that a similar measure, based on a student's successful persistence in undergraduate activities, might help predict persistence in graduate school.

Finally, we need to discuss an outcome considered by many to be of little value, graduate school grades. Over all participating departments in this study, the three predictors generally correlated .5 or higher with cumulative graduate grade point average, generally .4 without correction for restriction of range (Table 1). The correlations observed in this study are somewhat smaller than those computed by Wang (2002) for first-year grades in graduate school. She found an average corrected correlation of .65 for GRE verbal and quantitative scores and undergraduate grade point average with first-year graduate grades and an uncorrected correlation of .52. The difference between Wang's (2002) results and the results of this study may be because first-year graduate grades are more predictable than cumulative grades or faculty ratings of long-term success. Students in a given department may take a more comparable set of courses in the first year. The most fundamental knowledge in the field is likely to be covered in that year. In addition, weak students may leave during or after the first year, which means that there would be a greater range of performance among first-year students.

One important drawback of using first-year grades as a criterion of success in graduate school is that it tacitly assumes that all graduate students are full-time students. This has never been true, and is growing less common every year. However, it may be possible to generalize the criterion to include grades in the first 12 to 18 credits, or in common core courses, or some other definition that approximates a functionally equivalent criterion across a wide array of programs.

It is true that graduate school grades have always ranged between A and B, and in several of our participating departments ranged between A+ and A-. We suggest, however, that despite the very narrow scope of grades, there appears to be systematic information distinguishing different levels of accomplishment captured within that narrow scope. Common core grades represent a substantial number of hours of graduate school work, supervised by as many as seven or eight different faculty members. That accumulated evidence is almost bound to be important. Furthermore, grades are available for nearly all graduate students, and the data are almost universally available on central data bases. This provides a very important advantage, since it means that the simplest validity studies need not involve faculty at all. Thus we come to the suggestion that, for normal purposes, a study of common core grades seems like a reasonable way to check on the continuing appropriateness of an institution's admission requirements or a particular department's admission procedures. These are the kinds of studies that then could be accumulated in a national database if agreement could be reached about common format across institutions.

Periodically, it may be advisable to involve faculty in a discussion of long-term goals, and in gathering more specific information about student outcomes. Such a study might best be done longitudinally, following a group of students through graduate school and even into their professional life. Such a study would probably not focus only on admission, but on the entire process of finding, teaching, professionalizing, and placing graduate students. To succeed, the study would need commitment by faculty, since it would require agreement on goals, take a number of years, and require careful observation of student accomplishments.

References

Bowen, W. G., & Bok, D. (1998). *The shape of the river.* Princeton, NJ: Princeton University Press.

Braun, H. I., & Jones, D. H. (1985). *Use of empirical Bayes method in the study of the validity of academic predictors of graduate school performance* (GRE Board Professional Rep. No. 79-13P; ETS RR-84-34). Princeton, NJ: ETS.

Burton, N. W., & Cline, F. (in press). *The predictive validity and fairness of SAT item types for ethnic, gender, and language groups: Two studies.* New York: College Board.

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (College Board Research Rep. No. 2001-2). New York: College Board.

Burton, N. W., Welsh, C., Kostin, I., & Van Essen, T. (in press). *Toward a definition of verbal reasoning in higher education*. Princeton NJ: ETS.

Cohen, J. (1977). *Statistical power analyses for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 65-108.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.

Goldberg, E. L., & Alliger, G. M. (1992). Assessing the validity of the GRE for students in psychology: A validity generalization approach. *Educational and Psychological Measurement 52*, 1019-1027.

Goody, J. (1977). *The domestication of the savage mind.* New York: Cambridge University Press.

Gulliksen, H. (1987). *Theory of mental tests.* Hillsdale, NJ: Erlbaum.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

How not to pick a physicist. (1996). *Science 274*(1), 710-712.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis.* Newbury Park, CA: Sage Publications.

Jencks, C., & Phillips, M. (Eds.). (1998). *The black-white score gap.* Washington, DC: Brookings Institution Press.

Kyllonen, P., Walters, A. M., & Kaufman, J. C (2004). *Noncognitive constructs in graduate education: A review and some applications.* Manuscript in preparation.

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.

Lannholm, G. V. (1960). *Abstracts of selected studies of the relationship between scores on the Graduate Record Examinations and graduate school performance* (GRE Special Rep. No. 60-3). Princeton, NJ: ETS.

Lannholm, G. V. (1968). *Review of studies employing GRE scores in predicting success in graduate study, 1952-1967* (GRE Special Rep. No. 68-1). Princeton, NJ: ETS.

Lannholm, G. V. (1972). *Summaries of GRE validity studies 1966-70* (GRE Special Rep. No. 72-1). Princeton, NJ: ETS.

Lannholm, G. V., & Schrader, W. B. (1951). *Predicting graduate school success: An evaluation of the effectiveness of the Graduate Record Examinations*. Princeton, NJ: ETS.

Linn, R. L. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A. Wigdor & W. Garner (Eds.), *Ability testing: Uses, consequences, and controversies, Part II. Report of the National Academy of Sciences Committee on Ability Testing* (pp. 335-388). Washington, DC: National Academy Press.

Longford, N. T. (1991). *Negative coefficients in the GRE validity study service* (GRE Board Professional Rep. No. 89-05P; ETS RR-91-26.). Princeton, NJ: ETS.

Morrison, T., & Morrison, M. (1995). A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the Graduate Record Examination with graduate grade point average representing the criterion of graduate success. *Educational and Psychological Measurement 55*(2), 309-316.

Nist, S. T., & Simpson, M. L. (2000). College studying. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III). Mahwah NJ: Lawrence Erlbaum Associates.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York: Harcourt College Publishers.

Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA as the criterion for the predictive validity of the SAT. In W. W. Willingham, C. Lewis,

R. Morgan, & L. Ramist (Eds.), *Predicting college grades: An analysis of trends over two decades* (pp. 253-258). Princeton, NJ: ETS.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic group* (College Board Rep. No. 93-1; ETS RR94-27). New York: College Board.

Schneider, L. M., & Briel, J. B. (1990). *Validity of the GRE: 1988-89 summary report*. Princeton, NJ: ETS.

Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? *American Psychologist 52*(6), 630-641.

Stricker, L., Rock, D., & Burton, N. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology, 85*(4), 710-718.

Wagner, R. K., & Stanovich, K. E. (1996). Expertise in reading. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games.* Mahwah, NJ: Lawrence Erlbaum Associates.

Walpole, M. B., Burton, N. W., Kanyi, K., & Jackenthal, A. (2002). *Selecting successful graduate students: In-depth interviews with GRE users* (GRE Board Research Rep. No. 99-11R; ETS RR-02-06). Princeton, NJ: ETS.

Wang, M. (2002). *GRE validity tables* (Internal memorandum). Princeton, NJ: ETS.

Willingham, W. W. (1974). Predicting success in graduate education. *Science 183*, 273-278.

Willingham, W. W. (1985). *Success in college: The role of personal qualities and academic ability.* New York: College Board.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment.* Mahwah, NJ: Erlbaum.

Wilson, K. M. (1978). *Predicting the long-term performance in college of minority and nonminority students: A comparative analysis in two collegiate settings* (ETS RB-78-6). Princeton, NJ: ETS.

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 Rep. No. 75-8R). Princeton, NJ: ETS.

Wilson, K. M. (1980). The performance of minority students beyond the freshman year: Testing a "late bloomer" hypothesis in one state university setting. *Research in Higher Education 13,* 23-47.

Wilson, K. M. (1986). *The relationship of GRE General Test scores to first-year grades for foreign graduate students: Report of a cooperative study* (GRE Board Professional Rep No. 82-11P; ETS RR-86-44). Princeton, NJ: ETS.

Yee, C. (2003, September 25). Committee calls for new GRE changes. *The Chronicle Online.* Retrieved October 2, 2003, from http://www.chronicle.duke.edu/vnews/display.v/ART/2003/09/25/3f72d242e272d?in_archive=1

Notes

[1] Originally, engineering departments were also included in the design, but we were unable to find departments willing to participate. The recruiting was done in 2000, at a time when the economy was booming, especially in technical areas, and the engineering programs we contacted were focused on recruiting issues.

[2] The multivariate correction assumes invariant regression weights, and the associated error variances are based on the recommended prediction equation. When the recommended prediction equation does not include all three predictors, the weight for any predictor absent in the recommended equation is set to zero. In this way, a common corrected variance-covariance matrix of the predictors and one outcome is used to estimate the corrected multiple correlations for all predictor combinations. Thus the results for the recommended equations correspond to that of the usual corrections for explicit selection on the predictors present in the equations. (Gulliksen, 1987, pp. 164–165). Let Σ_{xx} be the covariance matrix of the predictors for the target (reference) population; S_{xx} be the sample covariance matrix of the predictors x; $\hat{\beta}_{y.x}$ and $\hat{s}_e^2$ be the vector of estimated regression coefficients for the predictors (with value 0 if a predictor is absent in the equation) and the estimated error (residual) variance, respectively, for the recommended prediction equation. The population variance of the outcome variable is estimated as $\hat{\sigma}_y^2 = \hat{s}_e^2 + \hat{\beta}'_{y.x} S_{xx} \hat{\beta}_{y.x}$, and the population covariance of the predictors with the outcome is estimated as $\hat{\Sigma}_{xy} = \Sigma_{xx} \hat{\beta}_{y.x}$. Then the corrected multiple correlation of a combination v of the predictors x (i.e., v is a subset of x) is obtained as $R_{y.v}^C = (\hat{\Sigma}_{yv} \Sigma_{vv}^{-1} \hat{\Sigma}_{vy} / \hat{\sigma}_y^2)^{\frac{1}{2}}$. Note that $\hat{\Sigma}_{vy}$ is a subvector of $\hat{\Sigma}_{xy}$ containing the elements corresponding to the predictors v; $\hat{\Sigma}_{yv} = \hat{\Sigma}'_{vy}$; and Σ_{vv} is a submatrix of Σ_{xx} corresponding to the covariance matrix of the predictors v.

[3] The correction formula used is the one suggested by Pedhazur (1997, p. 208). In general, larger correlations shrink less with this correction. The correction sometimes produces negative squared multiple correlations, which cannot be interpreted.

[4] The GRE background questionnaire has two pertinent questions. Applicants are asked directly if they are U.S. citizens (these were counted as citizens), or if they are resident aliens or citizens of another country (these were counted as noncitizens.) In addition, GRE registrants are asked to

specify their ethnic group only if they are citizens of the United States. Thus, for students who did not respond directly to the citizenship question, an ethnic group designation was taken as evidence that the applicant is a U.S. citizen. If neither the citizenship nor the ethnic question were answered, the graduate school's classification was used. Students' responses were given priority, since it was assumed that they are likely to be more aware of their citizenship status than their department or graduate institution would be.

[5] Empirical Bayes procedures are used to counteract the imprecision of regression equations computed for small samples of students. Ordinary least squares regression equations are computed separately for a group of departments; then the results are adjusted toward an equation based on pooled departmental results. See Braun & Jones (1985). This method differs from the method of pooled department analysis used in this study. Empirical Bayes is more radical in that all parameters, including the intercept, are adjusted. Empirical Bayes is less radical in that each department's parameters are adjusted *toward* the pooled result, but, especially for larger departments, maintain some independence.

[6] There are other issues with quality of data to keep in mind as well. Institutional records of progress toward degree are not ideal. They do not have good data on reasons for withdrawal, but these may vary greatly among students and programs. Policies and degree requirements vary a great deal—in some departments, good students may be held for years after passing comprehensive exams to do research. The records seldom track how many times a student may have failed common examinations. They may overwrite degree status in the records, so that a student who started as a master's student may be shown as a doctoral student as soon as that student enters a doctoral program. So the student, who attained the intended degree and more, may appear to be making very poor progress, given the length of enrollment, toward the next degree. Documentation should be requested from the institution about how such records are updated, as well as information on when such records were last updated. Special data on what happens to withdrawing students are probably necessary. The researchers also need to keep accurate records of when the data were received and/or updated once collected.

Appendix A

Participating Institutions and Departments

RESEARCH

PENN STATE UNIVERSITY

GRADUATE SCHOOL

Vice President for Research and Graduate Dean: Dr. Eva Pell

Former Assistant Dean: Dr. Richard Yahner

Assistant Dean: Dr. Barbara W. Pennypacker

CHEMISTRY, Eberly College of Science

Chair: Dr. Andrew Ewing

Graduate Director: Dr. Karl Mueller

ENGLISH, College of Liberal Arts

Chair: Dr. Dan Bialostosky

Graduate Director: Dr. Jack Selzer

PSYCHOLOGY, College of Liberal Arts

Chair: Dr. Keith Crnic

UNIVERSITY OF COLORADO HEALTH SCIENCES CENTER

Director of Admissions and Student Support: Fran Osterberg

MOLECULAR BIOLOGY

Graduate Advisor: Dr. Judith Jaehning

BIOCHEMISTRY AND MOLECULAR GENETICS

Graduate Advisor: Dr. Robert Sclafani

CELL AND DEVELOPMENTAL BIOLOGY

Graduate Advisor: Dr. Kathryn Howell

MICROBIOLOGY

Former Graduate Advisor: Dr. Kathryn Holmes

Graduate Advisor: Dr. Ron Gill

UNIVERSITY OF SOUTHERN CALIFORNIA

Director of Graduate and Professional Programs, Office of the Provost: Dr. Jonathan Kotler

GRADUATE SCHOOL

Vice Provost for Academic Programs and Dean of the Graduate School:

Dr. Joseph B. Hellige

BIOLOGY

Former Chair: Dr. Donal T. Manahan

Chair: Dr. Sarah W. Bottjer

Dr. William Trusten

PSYCHOLOGY

Chair: Dr. Stephen Read

CHEMISTRY

Chair: Dr. Curtis Wittig

ENGLISH

Chair: Dr. Percival Everett

ROSSIER SCHOOL OF EDUCATION

Dean: Dr. Karen S. Gallagher

Associate Dean for Academic Programs: Dr. David Marsh

DOCTORAL

FORDHAM UNIVERSITY

GRADUATE SCHOOL

Graduate Dean: Dr. Nancy Busch

Former Assistant Dean: Dr. Craig Pilant

Assistant Dean: Dr. Tony DeCarlo

ENGLISH

Chair: Dr. Frank Boyle

Graduate Director: Dr. Philip Sicker

Graduate Director: Dr. Nicola Pitchford

PSYCHOLOGY

Former Chair: Dr. Mary E. Procidano

Chair: Dr. Fred Wertz

Dr. Charles Lewis, Professor

BIOLOGY

Chair: Dr. Berish Y. Rubin

MASTER'S

TEXAS A&M INTERNATIONAL UNIVERSITY (Hispanic Serving Institution)

Associate Vice President of Student Services and Research: Dr. Mary Treviño

SCHOOL OF ARTS AND HUMANITIES

Former Dean: Dr. Jerry D. Thompson

Current Dean: Dr. Nasser Momayezi

PSYCHOLOGY

Chair: Dr. Cecilia Garza

ENGLISH, LANGUAGE, LITERATURE AND ARTS

Chair: Dr. Thomas Mitchell

SCHOOL OF EDUCATION

Dean: Dr. Rosa Maria Vida

DEPARTMENT OF PROFESSIONAL PROGRAMS (Education)

Chair: Dr. Humberto Gonzalez

UNIVERSITY OF SOUTH ALABAMA

Associate Vice President for Research and Graduate Dean: Dr. James L. Wolfe

BIOLOGY

Chair: Dr. John Freeman

ENGLISH

Former Chair: Dr. Linda Payne

Chair: Dr. Sue Walker

SCHOOL OF EDUCATION

Associate Dean of Graduate Studies and Research: Dr. William Gilley

VIRGINIA STATE UNIVERSITY (HBCU)

GRADUATE SCHOOL

Director of Graduate Studies, Research and Outreach: Dr. Wayne Virag

Dr. James F. McClelland, Professor Emeritus

LIFE SCIENCES: BIOLOGY

Chair: Dr. Larry C. Brown

ENGLISH LANGUAGES & LITERATURE

Chair: Dr. Freddy L. Thomas

EDUCATIONAL ADMINISTRATION, GUIDANCE & COUNSELING

Chair: Dr. Raymond Griffin

EDUCATIONAL CURRICULUM & INSTRUCTION

Chair: Dr. Vykuntapathi Thota

MATHEMATICS EDUCATION

Chair: Dr. George Wimbush

Appendix B

Definitions of Success in Graduate School

Cumulative Graduate GPA: Average of all credit courses taken in graduate school that were academically graded and relevant to the degree being sought, weighted by the number of credit hours for each. Reported on a 0 (failing) to 4 (A) scale, with + and – counted as +/- 1/3 of a grade point when available.

Progress to Degree: A variable constructed at ETS on a 0 to 8 scale:

blank = Department did not report this information for any student.

0 = Failed common exams

1 = Withdrew before common exams or did not register after first 2 terms

2 = Still enrolled in master's program, but has reached no further milestone

3 = Passed master's common exams

4 = Entered doctoral degree program but got master's degree and left

5 = Entered master's degree program and attained master's degree

6 = Still enrolled in doctoral degree program, but has reached no further milestone

7 = Passed doctoral common exams

8 = Attained doctoral degree or defended thesis

(This variable was not used in the final analysis.)

Faculty ratings: All students were rated by faculty on the following characteristics:

Mastery of the discipline includes knowledge of the discipline, ability to apply that knowledge to new situations; ability to structure, analyze, and evaluate problems; and an independent ability to continue learning.

Professional productivity includes the extent to which the student shows good judgment in selecting professional problems to attack, and the practical abilities of planning, flexibility in overcoming obstacles, and determination in carrying problems to successful completion.

Communication skills include the ability to judge the needs of one's audience; a mastery of the language of the discipline; a mastery of standard English; and the ability to communicate and work cooperatively with others.

Faculty used the following 0 to 6 scale to rate students:

blank = Department did not report this information for any student.

0 = I do not know student well enough to rate; *treated as missing.*

1 = Unsatisfactory relative to this department's recent standards.

2 = Adequate; marginal performance relative to this department's recent standards.

3 = Good; a solid representative of recent students, with few weaknesses.

4 = Excellent; a fine representative of recent students with clear strengths and few weaknesses.

5 = Distinguished; among the best students this department recently has had.

6 = Outstanding; no more than one or two of this department's recent students compare.

Appendix C

Discipline-Specific Results

Table C1

Average Correlations for Four Outcomes by Discipline

		Numbers		*U, V, Q*		*V, Q*		*U*	
		Dept.	*Stud.*	R*(corr)*	*R*	*R(corr)*	*R*	*r(corr)*	*r*
Biology	CGPA	5	145	0.57	0.40	0.51	0.33	0.34	0.22
	Mastery of discipline	3	70	0.57	0.39	0.56	0.37	0.24	0.10
	Professional productivity	2	47	0.55	0.36	0.51	0.29	0.30	0.21
	Communication skill	3	67	0.55	0.39	0.54	0.37	0.24	0.15
Chemistry	CGPA	2	134	0.62	0.46	0.50	0.36	0.45	0.28
	Mastery of discipline	1	48	0.55	0.31	0.52	0.27	0.33	0.16
	Professional productivity	1	48	0.52	0.31	0.44	0.21	0.39	0.23
	Communication skill	1	48	0.33	0.23	0.17	0.10	0.30	0.21
Education	CGPA	3	699	0.44	0.38	0.32	0.29	0.35	0.29
	Mastery of discipline	2	83	0.67	0.49	0.66	0.48	0.05	0.04
	Professional productivity	2	83	0.62	0.45	0.57	0.41	0.08	0.04
	Communication skill	2	83	0.62	0.47	0.62	0.47	0.15	0.12
English	CGPA	5	170	0.47	0.40	0.45	0.39	0.16	0.11
	Mastery of discipline	3	73	0.58	0.50	0.58	0.49	0.19	0.11
	Professional productivity	3	64	0.50	0.42	0.44	0.38	0.21	0.14
	Communication skill	3	63	0.64	0.56	0.62	0.54	0.28	0.17
Psychology	CGPA	4	155	0.57	0.41	0.51	0.37	0.29	0.16
	Mastery of discipline	2	78	0.35	0.28	0.26	0.18	0.31	0.24
	Professional productivity	2	77	0.44	0.32	0.34	0.20	0.35	0.24
	Communication skill	2	78	0.32	0.26	0.28	0.21	0.25	0.18
All depts.	CGPA	19	1,303	0.53	0.41	0.46	0.35	0.29	0.20
	Mastery of discipline	11	352	0.55	0.41	0.53	0.38	0.21	0.12
	Professional productivity	10	319	0.53	0.38	0.46	0.31	0.25	0.16
	Communication skill	11	339	0.52	0.41	0.49	0.38	0.24	0.16

Note. V = GRE verbal; Q = GRE quantitative; U= undergraduate grade point average; CGPA = cumulative graduate grade point average; *R* = uncorrected multiple correlation; *R*(corr) = corrected multiple correlation; *r* = correlation of one predictor with the criterion; *r*(corr) = corrected correlation of one predictor with the criterion. Average correlations weighted by number of students in each department. Correlations reported uncorrected and corrected for multivariate restriction of range.

Table C2

Equations Predicting Cumulative Graduate Grade Point Average: Within Department and Pooled Within Department

	Institution	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*Pooled*
Biology	Number	15	0	0	38	10	58	24	145
	U	0.093			0.144	0.300	0.111	0.248	0.164
Regression wts	GRE V /200	-0.087			0.092	–0.044	0.140	0.360	0.116
	GRE Q /200	0.911			0.472	0.123	–0.036	–0.081	0.107
	Standard error of estimate	0.388			0.266	0.178	0.280	0.369	0.301
Multiple	Multiple R	0.363			0.540	0.607	0.274	0.448	0.299
Correlations	Corrected R	0.840			0.813	0.739	0.325	0.537	0.482
Cumulative	Mean	3.51			3.64	3.70	3.66	3.54	3.62
Graduate GPA	Standard dev.	0.369			0.303	0.183	0.283	0.385	0.313
Chemistry	Number	0	85	0	0	0	49	0	134
	U		0.298				0.147		0.245
Regression wts	GRE V /200		0.213				– 0.069		0.056
	GRE Q /200		-0.013				0.389		0.193
	Standard error of estimate		0.310				0.281		0.308
Multiple	Multiple R		0.436				0.498		0.380
Correlations	Corrected R		0.567				0.701		0.590
Cumulative	Mean		3.49				3.50		3.49
Graduate GPA	Standard dev.		0.338				0.314		0.328
Education	Number	0	0	2	0	138	453	108	701
	U					0.098	0.189	0.153	0.170
Regression wts	GRE V /200					0.029	0.117	0.301	0.116
	GRE Q /200					0.197	–0.042	–0.024	0.008
	Standard error of estimate					0.267	0.294	0.263	0.287
Multiple	Multiple R					0.486	0.325	0.479	0.350
Correlations	Corrected R					0.497	0.392	0.578	0.402
Cumulative	Mean					3.76	3.70	3.57	3.69
Graduate GPA	Standard dev.					0.302	0.310	0.295	0.311
English	Number	45	62	5	0	19	34	10	175
	U	0.018	0.018			–0.153	0.033	0.152	0.021
Regression wts	GRE V /200	0.297	-0.001			0.390	0.113	1.330	0.198
	GRE Q /200	0.005	0.014			0.256	0.064	-0.318	0.054
	Standard error of estimate	0.151	0.158			0.363	0.171	0.174	0.211
Multiple	Multiple R	0.678	0.068			0.600	0.385	0.909	0.437
Correlations	Corrected R	0.762	0.088			0.663	0.501	0.972	0.551
Cumulative	Mean	3.68	3.82			3.66	3.80	3.69	3.75
Graduate GPA	Standard dev.	0.199	0.154			0.415	0.177	0.340	0.234
Psychology	Number	52	41	13	0	0	49	0	155
	U	0.017	0.188	0.039			0.046		0.048
Regression wts	GRE V /200	0.157	-0.080	0.346			0.042		0.065
	GRE Q /200	–0.080	0.076	0.061			0.131		0.042
	Standard error of estimate	0.120	0.247	0.269			0.114		0.176
Multiple	Multiple R	0.441	0.256	0.469			0.504		0.255
Correlations	Corrected R	0.536	0.424	0.640			0.695		0.368
Cumulative	Mean	3.81	3.82	3.790			3.86		3.83
Graduate GPA	Standard dev.	0.129	0.246	0.264			0.129		0.180

Note. U= undergraduate grade point average. GRE verbal (V) and quantitative (Q) scores divided by 200 to reduce decimal places in the table. Correlations reported uncorrected and corrected for multivariate restriction of range.

Table C3

Average Over- and Underprediction of Cumulative Graduate Grade Point Average by Ethnic Group

		Number	*Over- / underprediction*
Biology	African American	20	0.027
	Asian American	6	0.026
	Hispanic American		
	White	60	-0.032
Chemistry	African American	3	-0.260
	Asian American	5	-0.162
	Hispanic American	4	-0.108
	White	81	-0.023
Education	African American	129	-0.057
	Asian American	69	-0.037
	Hispanic American	75	0.020
	White	347	0.029
English	African American	17	-0.132
	Asian American	6	-0.039
	Hispanic American	9	-0.088
	White	123	0.023
Psychology	African American	6	-0.265
	Asian American	10	-0.076
	Hispanic American	24	-0.004
	White	90	0.020
Total	African American	175	-0.065
	Asian American	96	-0.044
	Hispanic American	112	0.001
	White	711	0.015

Note. Predicted cumulative graduate grade point average based on recommended combination of undergraduate grade point average, GRE verbal and quantitative scores for all students in each department, excluding predictors with negative regression weights. Over-/underprediction computed by subtracting predicted cumulative graduate grade point average from observed cumulative graduate grade point average. Averages weighted by the number of students in each department.

Table C4

Average Over- and Underprediction of Cumulative Graduate Grade Point Average by Citizenship, Degree, and Mode of Test Delivery

		Citizenship			*Degree*			*Test delivery*	
		N	*Over/under prediction*		*N*	*Over/under prediction*		*N*	*Over/under prediction*
Biology	Citizen	98	-0.016	Master's	35	0.010	Computer	28	-0.120
	Noncitizen	47	0.034	Doctoral	110	-0.003	Paper	99	0.026
Chemistry	Citizen	94	-0.040	Master's	4	0.137	Computer	9	0.025
	Noncitizen	40	0.093	Doctoral	130	-0.004	Paper	113	0.015
Education	Citizen	639	0.002	Master's	483	-0.124	Computer	163	-0.018
	Noncitizen	57	-0.027	Doctoral	216	0.027	Paper	403	0.013
English	Citizen	157	-0.001	Master's	104	-0.012	Computer	27	-0.036
	Noncitizen	12	0.026	Doctoral	66	0.020	Paper	138	0.012
Psychology	Citizen	139	-0.003	Master's	13	-0.000	Computer	29	-0.056
	Noncitizen	12	0.068	Doctoral	142	0.000	Paper	114	0.010
Total	Citizen	1,127	-0.004	Master's	639	-0.010	Computer	256	-0.034
	Noncitizen	168	0.029	Doctoral	664	0.010	Paper	867	0.014

Note. Predicted cumulative graduate grade point average based on recommended combination of undergraduate grade point average, GRE verbal and quantitative scores excluding predictors with negative regression weights. Over- /underprediction computed by subtracting predicted cumulative graduate grade point average from observed cumulative graduate grade point average. Averages weighted by the number of students in each department.

Table C5

1987-1991 GRE Validity Study Service Data: Average Correlations of GRE Scores and Undergraduate Grade Point Average With Graduate First-Year Grade Point Average by Department Type

		Numbers		*Predictors*							
Department	*Corrected*	*Depts.*	*Studs.*	*V*	*Q*	*A*	*U*	*VQ*	*VQ A*	*VQ U*	*VQ AU*
Natural sciences	No	192	3,557	.30	.31	.28	.34	.40	.43	.52	.54
	Yes			.41	.48	.45	.44	.55	.57	.68	.69
Engineering	No	47	824	.30	.33	.31	.39	.41	.45	.54	.57
	Yes			.40	.46	.44	.47	.53	.54	.66	.68
Social sciences	No	143	2,442	.32	.33	.31	.33	.42	.46	.53	.55
	Yes			.46	.47	.46	.43	.54	.56	.64	.66
Humanities & arts	No	33	550	.32	.24	.19	.26	.37	.38	.46	.46
	Yes			.40	.34	.33	.33	.45	.45	.53	.53
Education	No	43	703	.29	.27	.29	.34	.38	.41	.50	.52
	Yes			.38	.37	.39	.40	.45	.47	.58	.60
Business	No	10	205	.26	.36	.30	.38	.41	.44	.55	.57
	Yes			.40	.47	.45	.47	.51	.53	.65	.66
All departs.	No	468	8,281	.26	.31	.29	.34	.40	.43	.52	.54
	Yes			.42	.46	.44	.43	.53	.55	.65	.66

Note. V = GRE verbal, Q = GRE quantitative, A = GRE analytical, U = undergraduate grade point average. The departments included in these analyses participated in the GRE Validity Study Service between 1987 and 1991. A minimum of 10 departments and 100 students in any departmental grouping were required. Only students for whom English is the best language are included, since international students (a large proportion of non-EBL students) do not, in general, have a comparable undergraduate grade point average. Correlations are the weighted averages of the individual departments. For each department, the composite of predictors with the highest correlation and no negative weights was used. Correlations are reported uncorrected and corrected for multivariate restriction of range. From Wang (2002). Copyright 2002 by ETS. Reprinted with permission.

Table C6

1987-1991 GRE Validity Study Service Data: Correlations of GRE Scores and Undergraduate Grade Point Average With Graduate First-Year Grade Point Average by Department Type for Men and Women

		Numbers			*Predictors*							
Department	*Corrected*	*Depts.*	*Studs.*	*Sex*	*V*	*Q*	*A*	*U*	*VQ*	*VQA*	*VQ U*	*VQA U*
Natural sciences	No	81	1,620	M	.24	.29	.25	.32	.35	.38	.47	.49
		78	1,726	F	.29	.32	.29	.37	.39	.42	.53	.55
	Yes			M	.38	.48	.44	.43	.53	.55	.65	.66
				F	.41	.47	.45	.49	.53	.55	.68	.69
Engineering	No	31	636	M	.27	.30	.27	.35	.39	.42	.51	.53
		2	28	F	.35	.23	.25	.08	.43	.44	.50	.50
	Yes			M	.39	.46	.42	.46	.53	.54	.66	.67
				F	.52	.47	.47	.40	.62	.62	.69	.70
Social sciences	No	51	833	M	.27	.34	.31	.32	.40	.44	.51	.53
		71	1,255	F	.35	.37	.32	.30	.46	.50	.54	.56
	Yes			M	.42	.46	.44	.43	.52	.53	.62	.64
				F	.50	.50	.48	.41	.58	.60	.66	.67
Humanities & arts	No	13	231	M	.30	.24	.25	.35	.36	.38	.50	.51
		13	249	F	.28	.27	.24	.32	.37	.38	.48	.48
	Yes			M	.41	.33	.35	.42	.45	.46	.57	.58
				F	.37	.39	.36	.45	.45	.46	.58	.59
Education	No	12	193	M	.39	.38	.31	.35	.53	.56	.59	.62
		20	395	F	.30	.25	.27	.34	.33	.37	.47	.49
	Yes			M	.52	.46	.46	.44	.59	.61	.70	.71
				F	.37	.34	.36	.41	.41	.43	.57	.59
Business	No	5	91	M	.28	.34	.31	.31	.43	.48	.55	.58
		5	97	F	.20	.43	.26	.48	.44	.46	.60	.61
	Yes			M	.50	.52	.52	.41	.60	.61	.68	.69
				F	.37	.51	.42	.59	.53	.54	.71	.72
All departments	No	193	3,604	M	.26	.31	.27	.33	.38	.41	.50	.52
		189	3,750	F	.31	.33	.29	.34	.41	.44	.53	.54
	Yes			M	.40	.46	.43	.43	.53	.54	.64	.65
				F	.43	.46	.44	.46	.53	.55	.66	.67

Note. V = GRE verbal, Q = GRE quantitative, A = GRE analytical, U = undergraduate grade point average. The departments included in these analyses participated in the GRE Validity Study Service between 1987 and 1991. A minimum of 10 departments and 100 students in any departmental grouping were required. Only students for whom English is the best language are included since international students (a large proportion of non-EBL students) do not, in general, have a comparable undergraduate GPA. Correlations are the weighted averages of the individual departments. For each department, the composite of predictors with the highest correlation and no negative weights was used. Correlations are reported uncorrected and corrected for multivariate restriction of range. From Wang (2002). Copyright 2002 by ETS. Reprinted with permission.

Table C7

1987-1991 GRE Validity Study Service Data: Uncorrected Correlations of GRE Scores With Graduate First-Year Grade Point Average for Students Whose Best Language Is (EBL) and Is Not (Non-EBL) English

Department	*Numbers*		Best language	*Predictors*							
	Depts.	*Studs.*		*V*	*Q*	*A*	*U*	*VQ*	*VQA*	*VQU*	*VQAU*
Natural sciences	192	3,557	EBL	.30	.31	.28	.34	.40	.43	.52	.54
	50	717	Non EBL	.28	.29	.25	-	.44	.52	-	-
Engineering	47	824	EBL	.30	.33	.31	.39	.41	.45	.54	.57
	39	734	Non EBL	.25	.31	.27	-	.42	.47	-	-
Social sciences	143	2,442	EBL	.32	.33	.31	.33	.42	.46	.53	.55
	13	189	Non EBL	.30	.39	.24	-	.50	.53	-	-
All departments above	382	6,823	EBL	.31	.32	.29	.34	.41	.44	.53	.55
	102	1,640	Non EBL	.27	.31	.26	-	.44	.50	-	-

Note. V = GRE verbal; Q = GRE quantitative; A = GRE analytical; U = Undergraduate grade point average; EBL= English best language. Undergraduate GPA was not used as a predictor for students whose best language is not English, since many of these students attended undergraduate schools outside the United States, where curriculums and grading standards are not known and not comparable. The departments included in these analyses participated in the GRE Validity Study Service between 1987 and 1991. A minimum of 10 departments and 100 students in any departmental grouping were required. Correlations are the weighted averages of the individual departments. For each department, the composite of predictors with the highest correlation and no negative weights was used. Correlations are not corrected for multivariate restriction of range. From Wang (2002). Copyright 2002 by ETS. Reprinted with permission.

GRE-ETS

PO Box 6000

Princeton, NJ 08541-6000

USA

To obtain more information about GRE programs and services, use one of the following:

Phone: 1-866-473-4373

(U.S., U.S. Territories*, and Canada)

1-609-771-7670

(all other locations)

Web site: www.gre.org

* America Samoa, Guam, Puerto Rico, and US Virgin Islands