Listening. Learning. Leading.®

Understanding What the Numbers Mean: A Straightforward Approach to GRE Predictive Validity

Brent Bridgeman

Nancy Burton

Frederick Cline

September 2008

ETS GRE Board Research Report No. 04-03

ETS RR-08-46

Understanding What the Numbers Mean:
A Straightforward Approach to GRE® Predictive Validity

Brent Bridgeman, Nancy Burton, and Frederick Cline

ETS, Princeton, NJ

GRE Board Research Report No. GREB 04-03

ETS RR-08-46

September 2008

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

Descriptions of validity results for the GRE® General Test based solely on correlation coefficients or percentage of the variance accounted for are not merely difficult to interpret, they are likely to be misinterpreted. Predictors that apparently account for a small percentage of the variance may actually be highly important from a practical perspective. This study used 2 existing data sets to demonstrate alternative methods of showing the value of the GRE as an indicator of 1st-year graduate grades. The combined data sets contained 4,451 students in 6 graduate fields: biology, chemistry, education, English, experimental psychology, and clinical psychology. In one set of analyses, students within a department were divided into quartiles based on GRE scores and the percentage of students in the top and bottom quartiles earning a 4.0 average was noted. Students in the top quartile were 3 to 5 times as likely to earn 4.0 averages compared to students in the bottom quartile. Even after controlling for undergraduate grade point average quartiles, substantial differences related to GRE quartile remained.

Key words: Preadmissions predictors, grade point average (GPA), first-year graduate grades, explained variance, levels of performance

Table of Contents

List of Tables

List of Figures

Numerous studies have been used to demonstrate the validity of the GRE® for a variety of purposes. Kuncel, Hezlett, and Ones (2001) provided a meta-analysis of such studies. They included data from 1,753 independent samples representing more than 85,000 graduate students and used eight different criteria to define success in graduate school. A report by Burton and Wang (2005) related GRE scores from 21 graduate departments to a number of outcome variables including graduate grades over multiple years and teacher ratings of skills valued in graduate school (mastery of the discipline, professional productivity, and communication skill). Results of such studies typically are summarized in terms of simple correlations, multiple correlations, and increments in multiple correlations with additional variables. Although these coefficients provide convenient summaries, they are difficult for lay audiences (and even trained researchers) to interpret. To anyone who is unfamiliar with correlations in the social sciences, a correlation (r) of 0.4 has little intrinsic meaning. As an alternative to the raw correlation, a squared correlation is frequently presented to indicate the amount of variance in the criterion that can be explained by the predictor. Unfortunately, this substitution is of little help because readers cannot picture a variance, much less what 16% of a variance really means. The picture gets even fuzzier when multiple regression methods are used to show the improvement in prediction when GRE scores are added to college grades. Again, the improvement is frequently described in terms of the additional variance in college grades that can be explained by the test scores. The additional explained variance attributable to the test is typically less than 10%. Readers do not understand what 10% of the variance means, but 10% of anything sounds quite unimportant.

Without a correction for restriction in range used in the above studies, the correlation of GRE scores and first-year graduate grades is about 0.3, explaining about 9% of the variance. To test critics, this appears to be a trivially small number. “The ability of the GRE to predict first-year graduate grades is incredibly weak, according to data from the test’s manufacturer. In one ETS study of 12,000 test takers, the exam accounted for a mere 9% of the differences (or variation) among students’ first-year grades” (The National Center for Fair and Open Testing [FairTest], 2001). Methods of describing the value of the GRE that do not rely on “explained variance” would be more comprehensible to the various audiences that evaluate the utility of a prediction measure.

Since at least 1982, there have been clear warnings that even trained social scientists may be severely underestimating the practical importance of apparently small amounts of explained variance (Rosenthal & Rubin, 1982). A recent example clearly showed the potential value of experimental treatments that explain only a miniscule percentage of the variance in the outcome variable. Wainer and Robinson (2003) cited data from a large-scale study in which 22,071 physicians were randomly assigned to take either aspirin or a placebo every other day over a five-year period and the outcome variable was a heart attack. Using a traditional explained variance approach indicated that much less than 1% of the variance in getting a heart attack could be explained by taking (or not taking) aspirin. (The r^2 is .001.) Focusing instead on the number of people in each group who actually had heart attacks told a far different story. In the group taking the aspirin, 104 participants had heart attacks; in the placebo group, there were 189, or almost twice as many.

Various alternatives to r^2 have been proposed. The binomial effect size display (BESD) converts r to a 2 x 2 table with equal marginals and cells defined by (.5 + r/2)*100 and (.5 – r/2)*100 (Rosenthal & Rubin, 1982). Although very useful as a demonstration, many real-life situations are not easily converted to balanced 2 x 2 tables. Even in the case of a simple 2 x 2 table, if the marginals are not equal there is no straightforward conversion of r to a difference in success probability (Falk & Well, 1997). An equally serious problem is that many predictors and criteria cannot be reasonably dichotomized.

A potentially more useful approach, based on a direct interpretation of the unstandardized weights in the regression equation, allows both the predictor and criterion to be continuous. Instead of focusing on differences in r^2 (or for more than one predictor, the multiple correlation, R^2), the focus is on how much performance improves on the criterion for a given improvement on the predictor, holding other variables constant. This is exactly the information provided by the unstandardized regression weight. Note that it is only the unstandardized weights that are directly interpretable on the original score scale; standardized weights have no straightforward interpretation on the original scale. Bowen and Bok (1998) used unstandardized weights to show how much rank in class in college improves as SAT® scores increase. Verbal and math SAT scores were combined and entered in 100-point intervals along with a number of background variables. In their regression equation, the unstandardized weight for the SAT score was 5.93, and Bowen and Bok discussed the results as follows:

> Moreover, the positive relationship between students' SAT scores and their rank in class…remains after we control for gender, high school grades, socioeconomic status,

> school selectivity, and major, as well as for race…. This relationship easily passes tests of statistical significance, but the magnitude of the effects (the “slope”) is modest: for these students, *an additional 100 points of combined SAT score is associated, on average, with an improvement of only 5.9 percentile points in class rank,*[1] No teacher will be surprised to hear that other factors, many of them unmeasureable, affect academic performance—especially in these highly competitive schools where nearly all students have strong academic skills. (p. 74)

When the criterion is dichotomous (e.g., graduate or not graduate), logistic regression is preferable to ordinary least squares regression. Because the dependent variable is in terms of log odds, direct interpretation of regression coefficients is not practical, but with a few simple transformations, results can be expressed as a probability of being in one of the dichotomous outcome categories. Bowen and Bok used this approach for a number of their analyses, and this method has been used to show the probability of getting a 2.5 or higher grade point average (GPA) for given levels of ACT scores (Noble, 2004).

For the above approaches to work satisfactorily, the models must fit the data reasonably well. For the ordinary least squares regression model, there must be a linear relationship between predictors and criteria, and for the logistic regression models, the logistic function must fit the data. A further consideration with these approaches is that although the outcome can be clearly explained to a nontechnical audience, the process of getting to this outcome is somewhat more obscure.

Bridgeman, Pollack, and Burton (2003) addressed these problems by presenting SAT validity results solely in terms of the proportion of students at different levels of performance on predictor measures who succeeded in college. They divided a sample of 41 colleges into four levels based on average SAT scores within each institution. For colleges at a given level, they identified two levels of successful students: those who achieved a 2.5 average and those who achieved a 3.5 average. These averages were computed at two time points: the end of the freshman year and after four years in the college. The percentage of successful students at different levels of three preadmissions predictors (high school curriculum intensity, high school grades, and SAT scores) was determined. All three indicators were strongly related to success in college. For example, in Level 1 colleges (i.e., colleges with mean combined SAT scores below 1100), 26% of the students in the lowest high school grade point average (HSGPA) category were successful by the criterion of a 2.5 GPA by the end of freshman year; 86% of the students in the highest category were successful. Similarly, 25% of the students in the lowest of five SAT levels were successful compared to 90% in the highest level. Even within a single level of curriculum intensity and HSGPA, success rates varied dramatically by SAT score level whether the criterion was freshman grades or four-year GPA. For example, with a success criterion of a four-year GPA of 3.5 or above in Level 4 colleges (i.e., colleges with mean SAT scores over 1250) among students who were very successful in high school (HSGPA over 3.7) and who had taken a very rigorous high school curriculum (at least three advanced placement courses), scores on the SAT still mattered. At the middle of the five SAT levels, fewer than 20% of the students were successful, but at the highest SAT level, 60% were successful. Bridgeman et al. concluded that high school performance and SAT scores may not appear to be strongly related to success in college if the focus is only on “variance accounted for,” but if percentage succeeding is the criterion, then the substantial relationship between SAT scores and college performance is apparent. The current study adapts these methods for the data available on the GRE population. In particular, the need to do analyses within individual academic departments and the small size of these departments, compared to the number of students in an entire college freshman class, provide additional challenges.

Method

Data Source

Two data sets were used in the analysis. The larger data set was selected from departments that participated in the GRE Validity Study Service (VSS) between 1987 and 1991. The initial data set included more than 8,000 students attending graduate school in a variety of departments. A minimum of 10 departments and 100 students was required for a group of departments to be included. The department groups fitting these criteria were natural sciences, engineering, social sciences, humanities/arts, education, and business.

From this universe, an analysis sample of 128 departments with 3,303 students was selected for use in this study—all graduate departments in biology, chemistry, education, English, and psychology. This subset was chosen to be comparable to the second data set, a group of 17 departments from seven different institutions that collaboratively studied the progress through graduate school of students entering graduate programs in 1995–96, 1996–97, or 1997–98 (Burton & Wang, 2005). From the latter study we included 1,148 masters and doctoral students in five disciplines: biology, chemistry, education, English, and psychology. We split psychology departments into two subsets: one subset included traditional experimental psychology programs; the other subset, which we labeled “clinical psychology,” included clinical, counseling, and community psychology programs. In addition to first-year graduate GPA, outcome measures included a transcript of all degree-related courses, credits, and grades; cumulative GPA throughout graduate school; milestones such as passing common examinations, attaining candidacy, and graduation; and faculty ratings of students’ mastery of the discipline, professional productivity, and professional communication skills. Graduate first-year GPA was estimated by averaging grades for the first eight courses taken by each student, weighted by the number of credits per course. This was done because many students in masters programs were not full-time students. Most took one or two courses a term, so first-year GPA could be based on as few as two courses.

From both data sets, students were selected who had complete data on GRE verbal (GRE-V), GRE quantitative (GRE-Q), undergraduate grade point average (UGPA), and graduate first-year GPA. Students who reported that English was not their best language and international students were excluded from the sample because many of these students attended undergraduate schools outside the United States, where grading standards are not known and not comparable. Departments with fewer than 10 students were also dropped from both samples. Table 1 describes the final analysis group from the two data sets.

Analyses

Two basic approaches were explored. The first approach used the ordinary least squares regression equation as the starting point, but rather than focusing on the overall R^2 or increments in R^2, we focused on the direct interpretations that can be derived from the unstandardized regression coefficients. Specifically, we computed a separate regression equation predicting the graduate GPA for each department in a field. We weighted the unstandardized coefficients by the sample size in the department and computed the weighted average for each of the coefficients. One set of regression equations included UGPA and total GRE score (sum of verbal and quantitative) as predictors. A second set of equations included UGPA, GRE-V, and GRE-Q as independent predictors.

Table 1

Description of the Analysis Sample

Discipline	1987–1991 VSS		1995–1998 Long-term validity study		Total *N*	
	Depts.	Students	Depts.	Students	Depts.	Students
Biology	21	453	3	61	24	514
Chemistry	14	334	2	109	16	443
Education	29	765	3	673	32	1,438
English	12	330	5	160	17	490
Experimental psychology	20	573	4	145	24	718
Clinical psychology	32	848	0	0	32	848
Total	128	3,303	17	1,148	145	4,451

The other basic approach did away with the regression equation entirely and instead simply ranked students on the predictors and criteria. Within a field, such as biology, different departments had quite different admissions standards so that top-scoring students at one department might have scores that would put them near the bottom in a more competitive department. Therefore, our definitions of outstanding admissions scores were always department dependent. Within a particular department, we identified three levels based on GRE scores. The top level was the top quarter of the students in that department based on combined verbal and quantitative scores. Combining scores in this manner is essentially an equal weighting of verbal and quantitative scores. When evaluating individuals, departments should consider these scores separately, but for our purposes, the combined score, reflecting the importance of both verbal and quantitative skill, is satisfactory. The middle level was the middle 50%, and the bottom level was the bottom quarter. With small department sizes, these cuts could not always be exact. (For example, with 10 students in a department, the top quarter would be 2.5 students, but we decided that cutting students in half was not advisable.) Furthermore, exact cuts were not necessary; it was sufficient that the top group represented the highest scoring students in the department, and the bottom group represented the lowest scoring students. We then made similar cuts based on UGPA. These within-department cuts were then aggregated across all of the departments in a field in the sample. We then looked at the success of students in these categories, and combinations of the GRE and UGPA categories, in terms of percentage of students in each category who reached a high level of success in their first-year courses.[2] We initially defined this high level as a 4.0 average. Although this was adequate for identifying truly exceptional students in most departments, this standard yielded very few students in chemistry departments. So, we also included a less stringent, but still very high-level standard of a 3.8 or better average. On the other side of the spectrum, we defined students who were in academic difficulty as students with less than a 3.0 average.

Results and Discussion

Analysis of Unstandardized Regression Weights

For biology departments, the unstandardized regression weight for the combined GRE score (i.e., GRE-V + GRE-Q) was .00060. Holding UGPA constant (or for students with identical UGPAs), that means a one-point increase in the GRE combined score would lead to an increase in the predicted graduate GPA of .00060 points on the 0–4 grade scale. This suggests that a one-point increase on the GRE scale is not meaningful, and indeed it is not. Combined scores can range from 400 to 1600, so it makes more sense to think in terms of 100-point differences than in single-point differences. But even a 100-point difference makes only a difference of .06 in predicted graduate average. A 200-point difference in GRE scores yields a noticeable, but hardly impressive, difference in predicted graduate grades. Using the regression equation based on the weighted averages, a student with a 3.5 UGPA and a 1200 GRE would be predicted to have a 3.60 graduate GPA; a student with the same UGPA, but a 1400 GRE, would be predicted to have a 3.72 graduate GPA.

Just as a single-point difference is not realistic in considering differences in GRE scores, a single-point difference in UGPA is not very meaningful, but for the opposite reason. The full range of applicants to a department may differ by only a single point in UGPA units (from a 3.0 to a 4.0). To put GRE scores and UGPA on a more nearly equal footing, while keeping to the original score units rather than possibly confusing standard score units, we provided the weights for a 100-point difference in combined GRE scores and a 0.25 difference in UGPA.[3] Table 2 shows the difference in graduate GPA units that are associated with a 100-point difference in combined GRE score (holding UGPA constant) and the difference in graduate grades associated with a 0.25 difference on the 0–4 UGPA scale (holding GRE constant).

Table 2

Expected Differences in Graduate GPA Associated With 100-Point Differences in Combined GRE Scores and 0.25 Differences in Undergraduate Grade Point Average(UGPA) by Graduate Department

Department	Change in GPA per 100 combined GRE points	Change in GPA per 0.25 UGPA points
Biology	0.060	0.054
Chemistry	0.054	0.083
Education	0.033	0.051
English	0.044	0.028
Experimental psychology	0.066	0.054
Clinical psychology	0.056	0.041

Note. The change in first-year GPA associated with GRE scores assumes UGPA is held constant and change associated with UGPA assumes GRE held constant.

This table is intended to show, in only a general way, how score differences and UGPA differences relate to graduate GPA differences. Differences between departments in these averaged coefficients should be treated very tentatively, or ignored, as the differences among departments in a field are far larger than the differences among the fields.

Table 3 separates the two components of the combined GRE score so that the separate contributions of the verbal and quantitative scores can be considered. As before, the table shows the change in graduate GPA associated with the indicated change on one of the predictors while holding the other predictors constant. For GRE-V, for example, both GRE-Q and UGPA are held constant to show the effects of a change in GRE-V score. Because the combined score has been cut in half, we show differences per 50 points on GRE-V and per 50 points on GRE-Q rather than the 100-point increments used for the combined score.

Analyses of Highly Successful and Less Successful Students by Score Categories

The quite modest changes in expected graduate GPA associated with fairly substantial differences in GRE scores might lead to the conclusion that the GRE is of practically no use in differentiating students who will be very successful from other students. But mean differences on the very compressed graduate GPA scale, with relatively few grades below a B, actually

Table 3

Expected Differences in Graduate GPA Associated With 50-Point Differences in GRE-V and GRE-Q Scores and 0.25 Differences in Undergraduate Grade Point Average (UGPA) by Graduate Department

Department	Change per 50 points on GRE-V	Change per 50 points on GRE-Q	Change per 0.25 UGPA points
Biology	0.034	0.017	0.049
Chemistry	0.024	0.029	0.084
Education	0.029	0.007	0.051
English	0.025	0.021	0.029
Experimental psychology	0.035	0.030	0.052
Clinical psychology	0.020	0.035	0.041

Note. Change in each column assumes scores in other two columns held constant. Q = quantitative, V = verbal.

reveal surprisingly little about how well the GRE identifies the students likely to be very successful in their departments (finishing the first year in the top quartile of GPAs) or those likely to be in academic difficulty (bottom quartile).

For biology departments, Figure 1 shows the percentage of students who were in the top or bottom quartiles of GRE scores in their class with first-year grade point averages (FYA) that were in the top or bottom quartiles of their class.

Figure 1 demonstrates that the small mean differences shown in the previous section can translate into substantial differences in success percentages. Among students in the bottom quartile of GRE scores in a biology department, only 15% earned GPAs in the top quartile; almost 3 times as many students (43%) in the top quartile of GRE scores ended the year with GPAs in the top quartile. Similarly, students in the bottom GRE quartile were more than twice as likely to finish in the bottom GPA quartile as in the top quartile.

Figures 2–6, for the other academic fields, tell essentially the same story as Figure 1. Indeed, the percentages are remarkably similar across fields. Differences were greatest in the clinical psychology departments in which only 10% of the bottom GRE score quartile finished in the top GPA quartile, contrasted with 41% from the top GRE quartile.

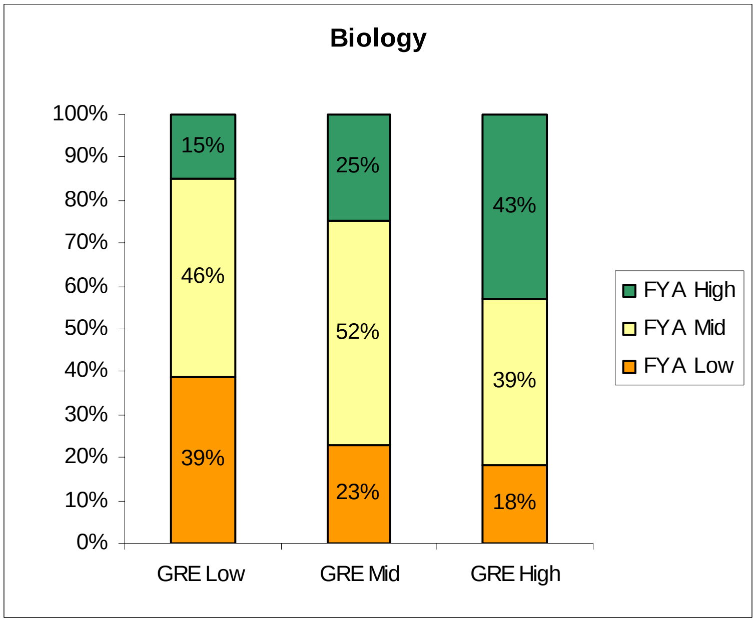

Figure 1. **Percentage of students in three GRE score categories whose first-year grade point averages (FYAs) in biology departments were in the bottom quartile, top quartile, or mid-50%.**[4]

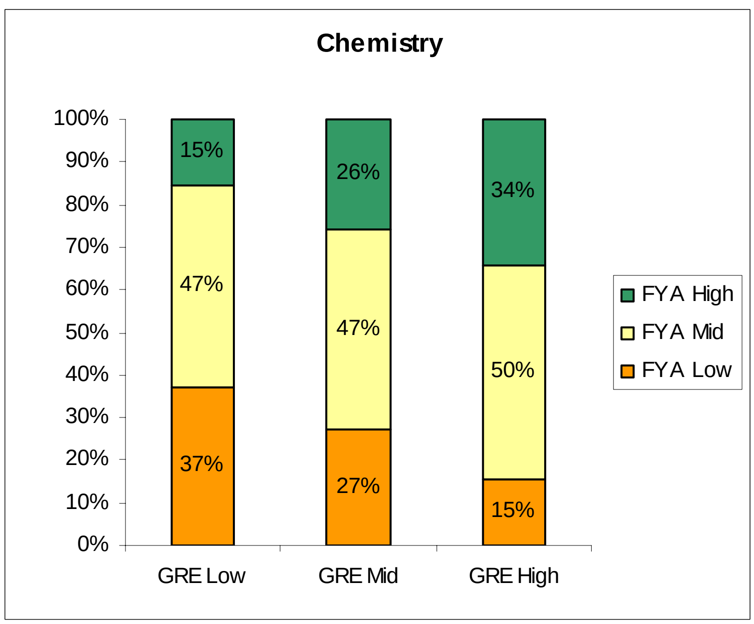

Figure 2. **Percentage of students in three GRE score categories whose first-year grade point averages (FYAs) in chemistry departments were in the bottom quartile, top quartile, or mid-50%.**[4]

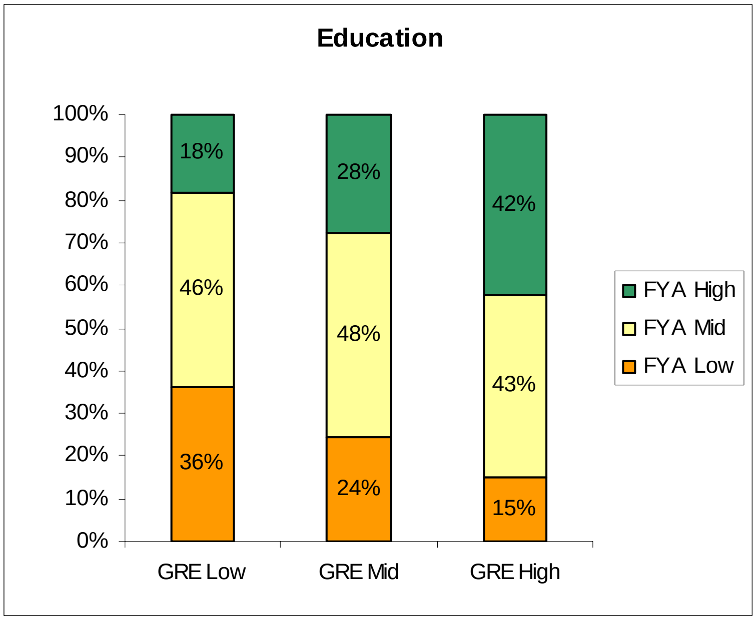

Figure 3. **Percentage of students in three GRE score categories whose first-year grade point averages (FYAs) in education departments were in the bottom quartile, top quartile, or mid-50%.** [4]

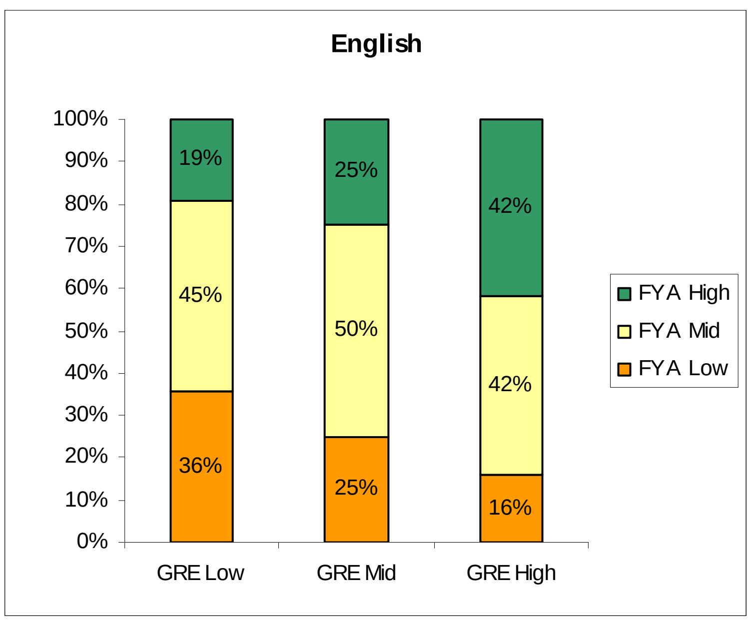

Figure 4. **Percentage of students in three GRE score categories whose first-year grade point averages (FYAs) in English departments were in the bottom quartile, top quartile, or mid-50%.**[4]

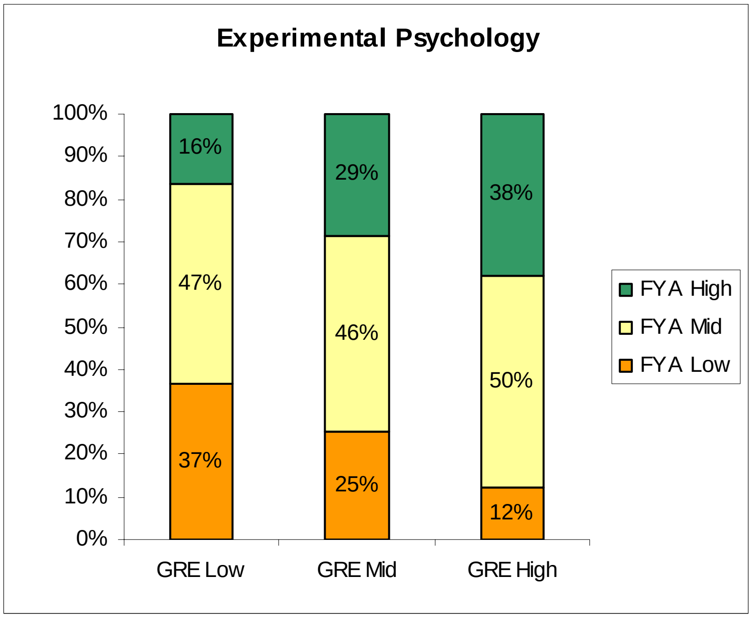

Figure 5. **Percentage of students in three GRE score categories whose first-year averages (FYAs) in experimental psychology departments were in the bottom quartile, top quartile, or mid-50%.** [4]

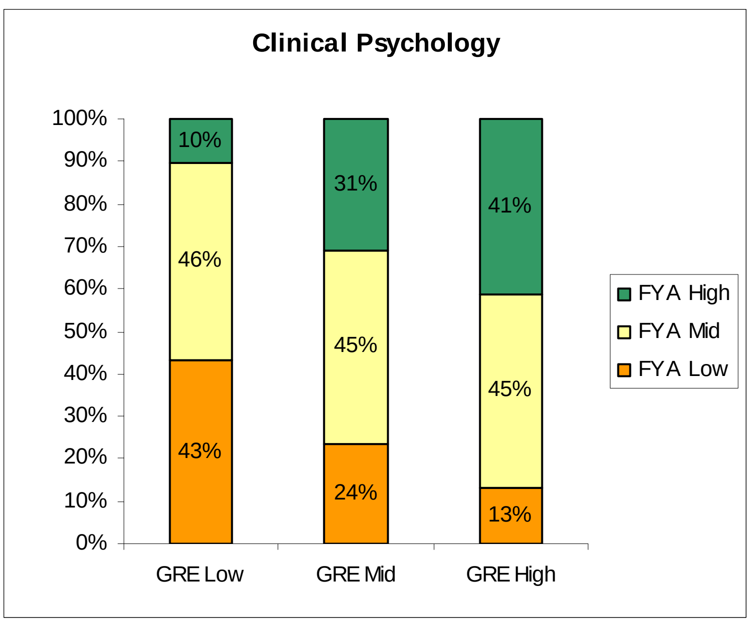

Figure 6. **Percentage of students in three GRE score categories whose first-year averages (FYAs) in experimental psychology departments were in the bottom quartile, top quartile, or mid-50%.** [4]

Although performing in the top quartile is a notable accomplishment, it does not reflect the true academic superstars. To identify the best of the best, at least in terms of first-year grades, we selected a sample of students with 4.0 averages. Figure 7 shows the percentage of students reaching this high level in the bottom and top quartiles of GRE scores within biology departments.

Differences were striking. Students in the top quartile of GRE scores were more than 5 times as likely to earn 4.0 averages compared to students in the bottom quartile. Figure 8 presents a comparable analysis for the chemistry departments. Because of the low percentage of students with 4.0 averages in chemistry departments, both bars are quite short and the difference does not appear to be as compelling. Nevertheless, twice as many students earned 4.0 averages in the high GRE category as in the low (5.4% versus 2.7%). When we lowered the standard to a still demanding 3.8 or higher level, a difference more similar to the one noted for biology departments in Figure 7 emerged. Students in the top quartile of GRE scores were about 2.5 times as likely to earn a 3.8 GPA as students in the bottom quartile (see Figure 9).

For the other fields, there were sufficient numbers of students with 4.0 averages that we returned to that standard for Figures 10–13.

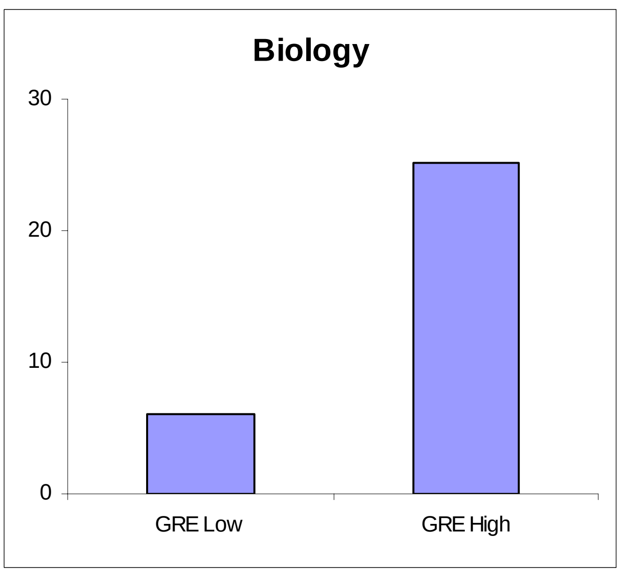

***Figure 7.* Percentage of students in bottom and top quartiles of GRE score within biology departments with first-year GPAs of 4.0.**

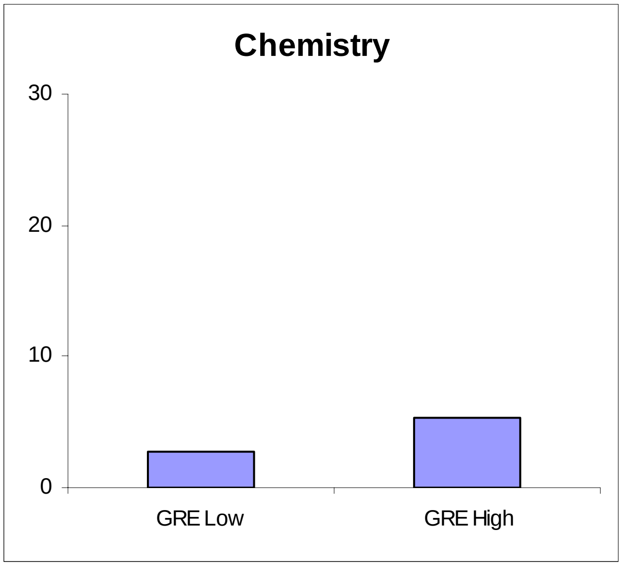

Figure 8. **Percentage of students in bottom and top quartiles of GRE score within chemistry departments with first-year GPAs of 4.0.**

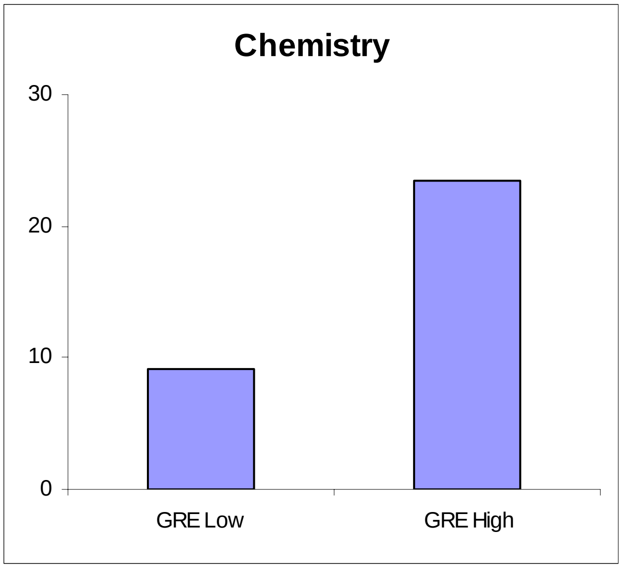

Figure 9. **Percentage of students in bottom and top quartiles of GRE score within chemistry departments with first-year GPAs of 3.8 or higher.**

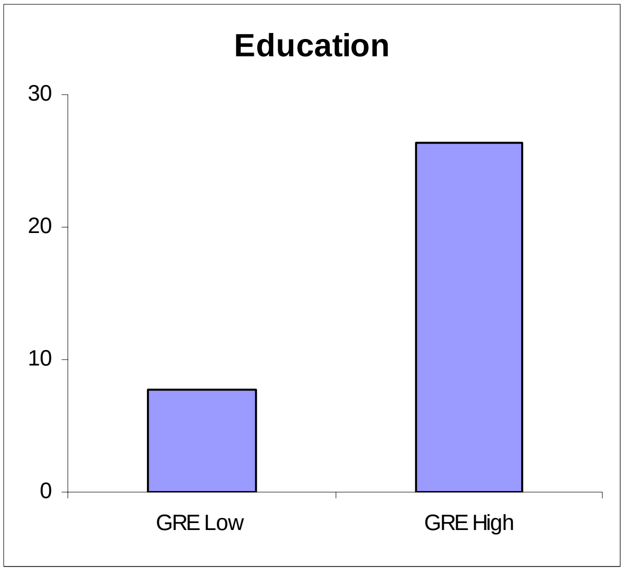

Figure 10. **Percentage of students in bottom and top quartiles of GRE score within education departments with first-year GPAs of 4.0**

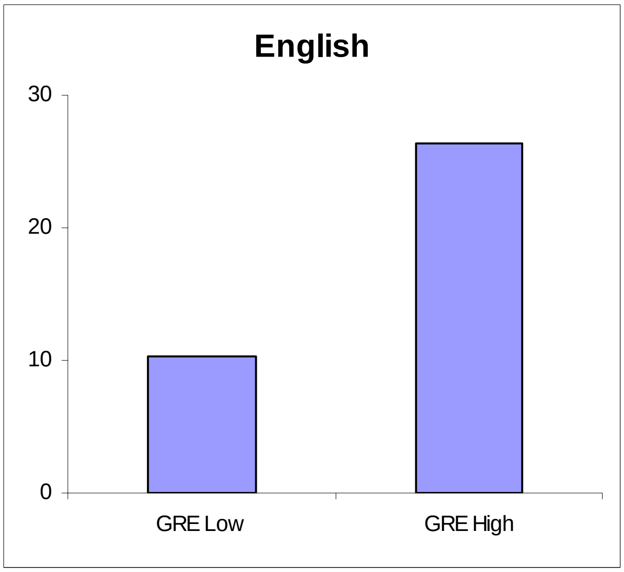

Figure 11. **Percentage of students in bottom and top quartiles of GRE score within English departments with first-year GPAs of 4.0.**

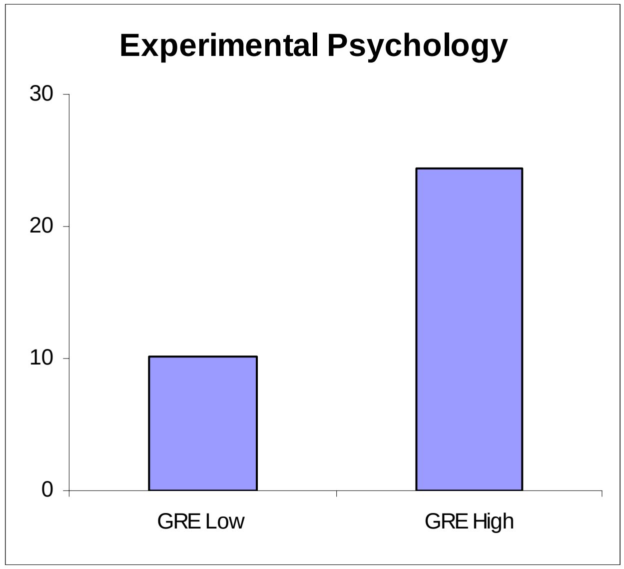

Figure 12. **Percentage of students in bottom and top quartiles of GRE score within experimental psychology departments with first-year GPAs of 4.0.**

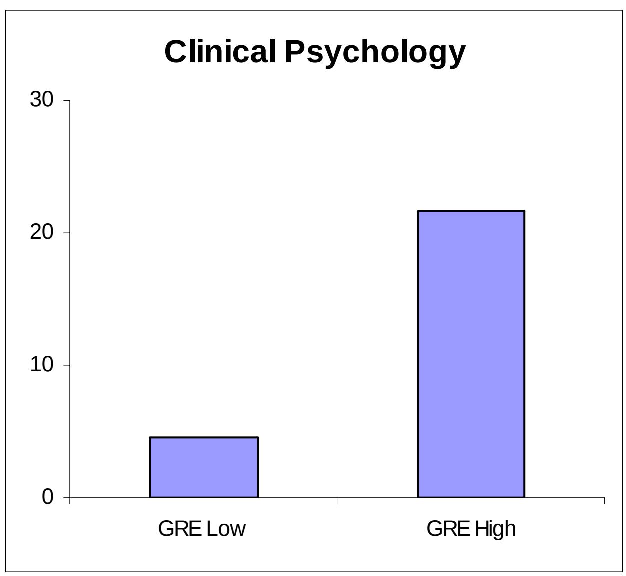

Figure 13. **Percentage of students in bottom and top quartiles of GRE score within clinical psychology departments with first-year GPAs of 4.0.**

At the other end of the academic spectrum are the graduate students who failed to attain at least a B average in their first year. As shown in Figure 14, students in the bottom quartile of the GRE score distribution within biology departments were slightly more than twice as likely to earn less than a B average compared to students in the top quarter of the within-department GRE score distribution. Although this difference between top and bottom quartiles of the GRE scores is substantial, it is considerably smaller than the differences noted for the 4.0 GPA students. This may reflect the multiple reasons for very low grades that are unrelated to verbal and quantitative reasoning skills. Poor motivation or personal adjustment problems can cause academic problems even for students with strong reasoning skills.

As shown in Figures 15–19, this pattern is repeated in the other departments, although the number of students with less than a 3.0 GPA is considerably smaller in the education, English, and psychology departments than in the biology and chemistry departments.

The figures presented thus far show the value of the GRE in identifying successful and unsuccessful graduate students, but they do not address the incremental validity question of how the GRE improves on what is already known from the undergraduate average. Figure 20 combines the GRE and UGPA predictors for biology departments. As previously, we used the bottom quartile, middle 50%, and top quartile of GRE scores and also performed the same type of quartile division for the UGPA. The figure shows that both GRE scores and UGPA make a difference. Among students in the bottom quartile in terms of their UGPAs, those with high GRE scores earned substantially[5] higher grades, on average, than those with low GRE scores. And among students in the bottom GRE quartile, those with UGPAs in the top quartile got noticeably higher grades than those in the bottom UGPA quartile.

Figures 21–25 provide similar information for the other departments. Within a UGPA level, students in the top GRE quartile consistently earned higher grades than those in the bottom quartile, but the middle 50% group was not always as clearly in the middle as it was in the biology departments. In the low UGPA group in chemistry departments (see Figure 21), for example, there was essentially no difference in mean graduate grades from the low to the mid GRE groups, though grades in the high GRE group were still somewhat higher. In English departments (see Figure 23), the middle and low GRE groups within each GPA level performed comparably.

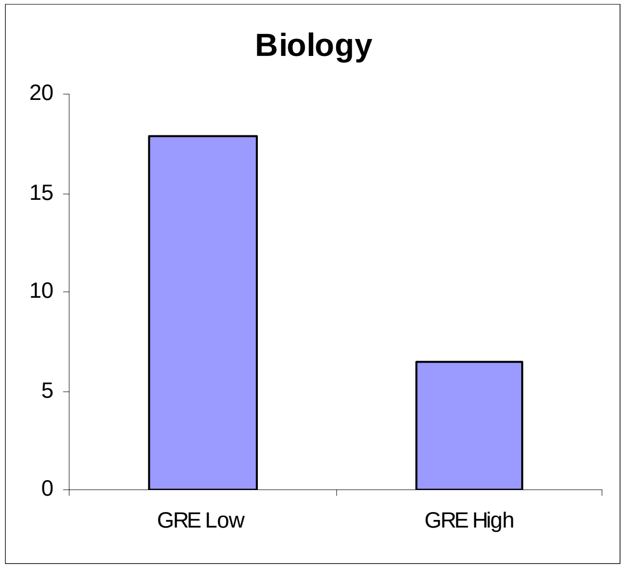

***Figure 14.* Percentage of students in bottom and top quartiles of GRE score within biology departments with first-year GPAs below 3.0.**

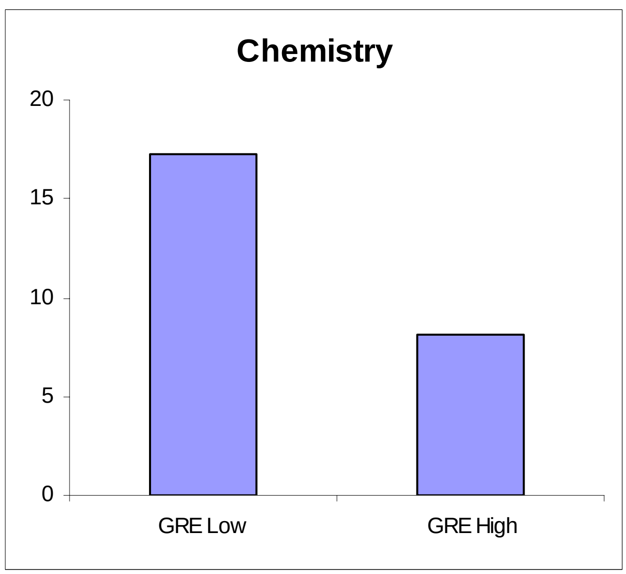

***Figure 15.* Percentage of students in bottom and top quartiles of GRE score within chemistry departments with first-year GPAs below 3.0.**

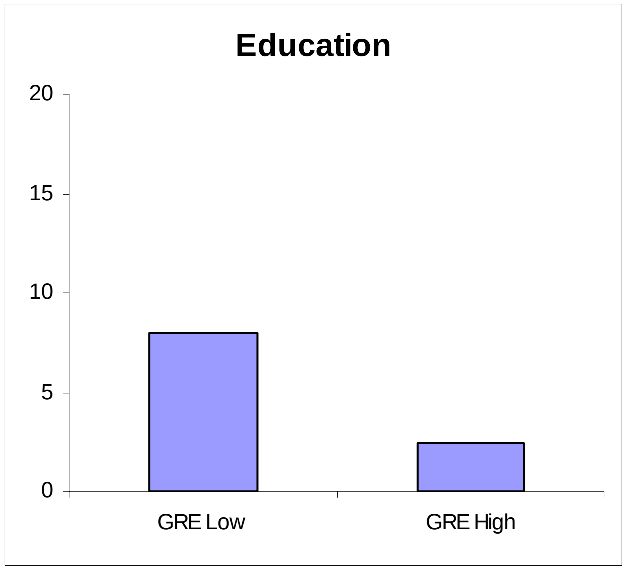

Figure 16. **Percentage of students in bottom and top quartiles of GRE score within education departments with first-year GPAs below 3.0.**

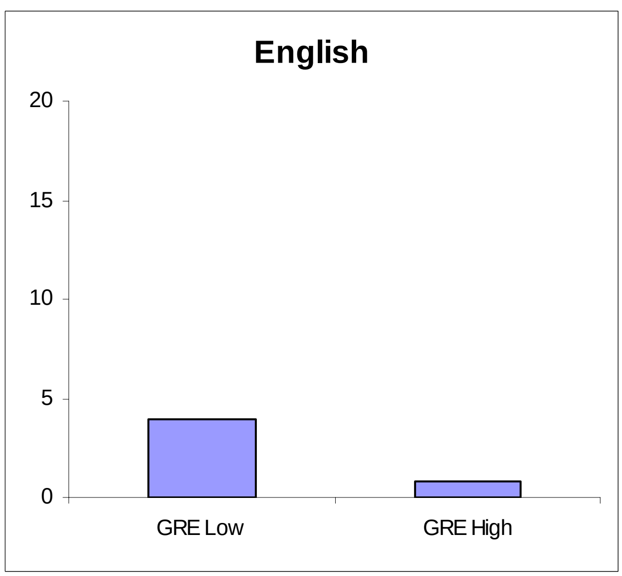

Figure 17. **Percentage of students in bottom and top quartiles of GRE score within English departments with first-year GPAs below 3.0.**

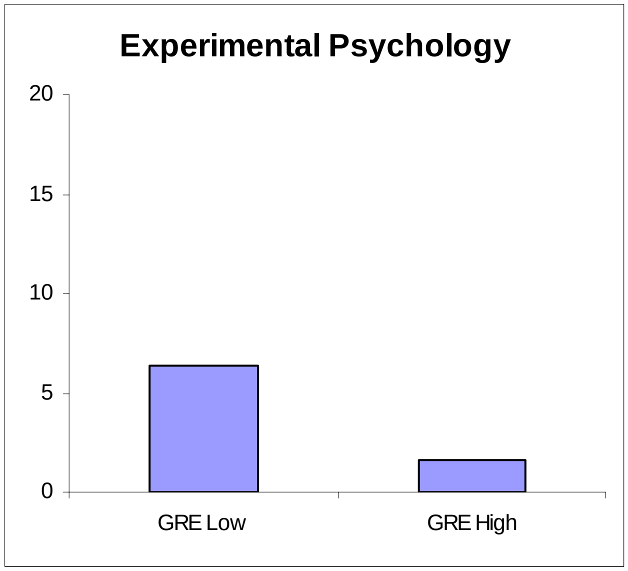

Figure 18. **Percentage of students in bottom and top quartiles of GRE score within experimental psychology departments with first-year GPAs below 3.0.**

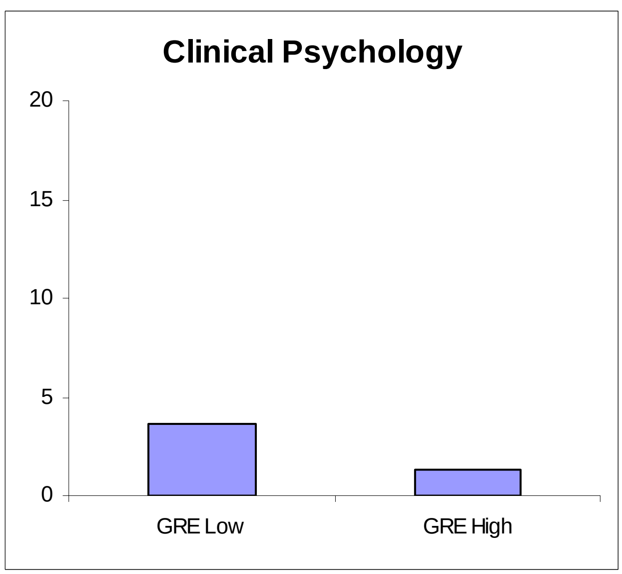

Figure 19. **Percentage of students in bottom and top quartiles of GRE score within clinical psychology departments with first-year GPAs below 3.0.**

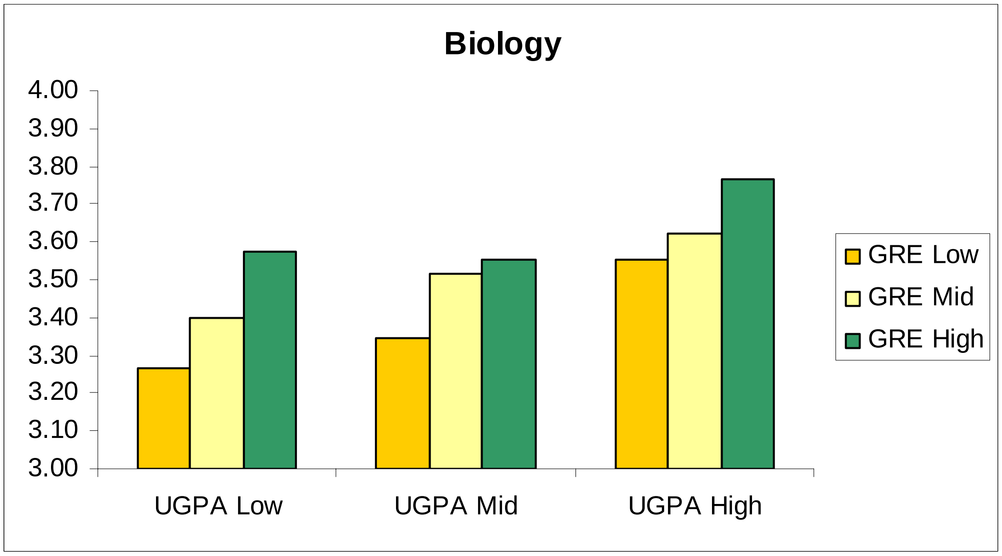

Figure 20. **Mean graduate GPA in biology departments by undergraduate GPA (UGPA) and GRE quartiles.** [6]

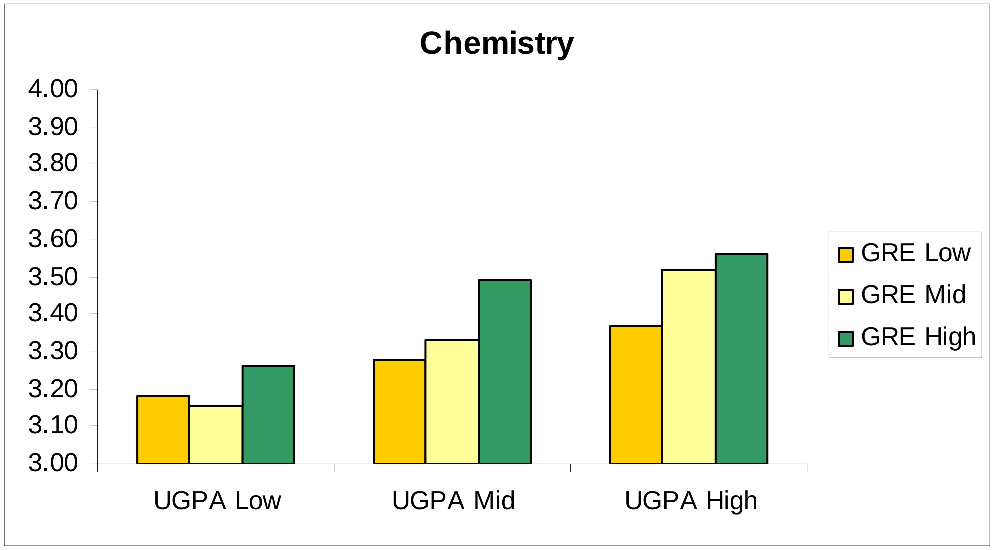

Figure 21. **Mean graduate GPA in chemistry departments by undergraduate GPA (UGPA) and GRE quartiles.** [6]

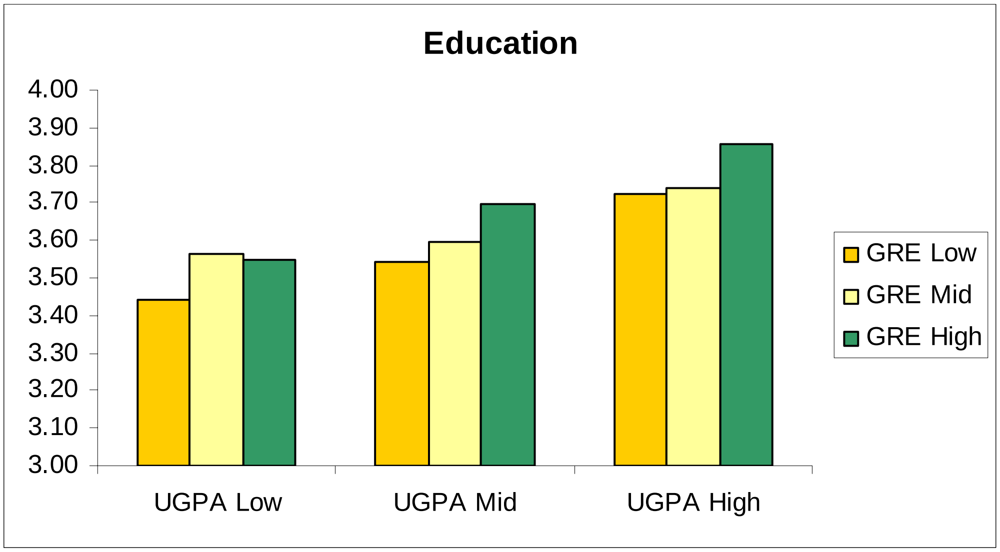

Figure 22. **Mean graduate GPA in education departments by undergraduate GPA (UGPA) and GRE quartiles.** [6]

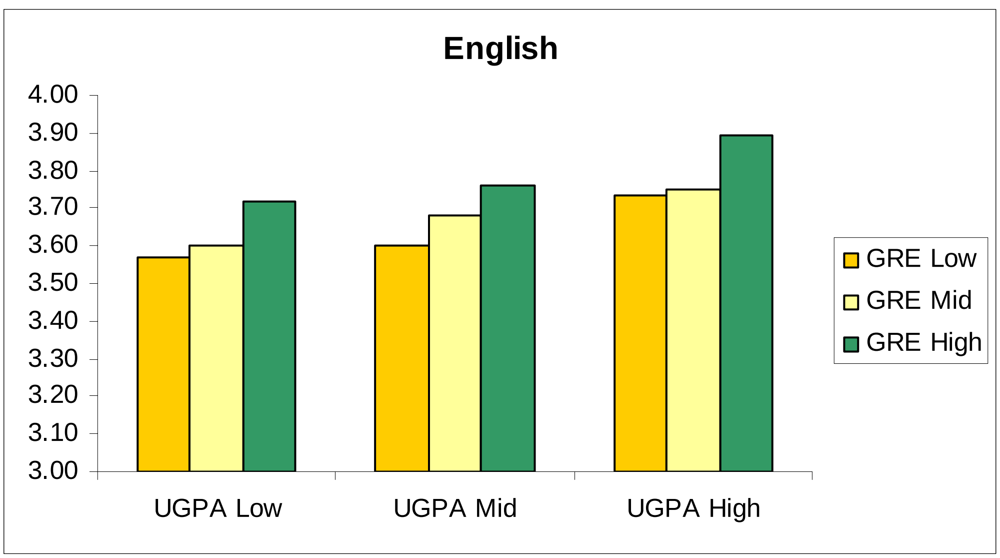

Figure 23. **Mean graduate GPA in English departments by undergraduate GPA (UGPA) and GRE quartiles.** [6]

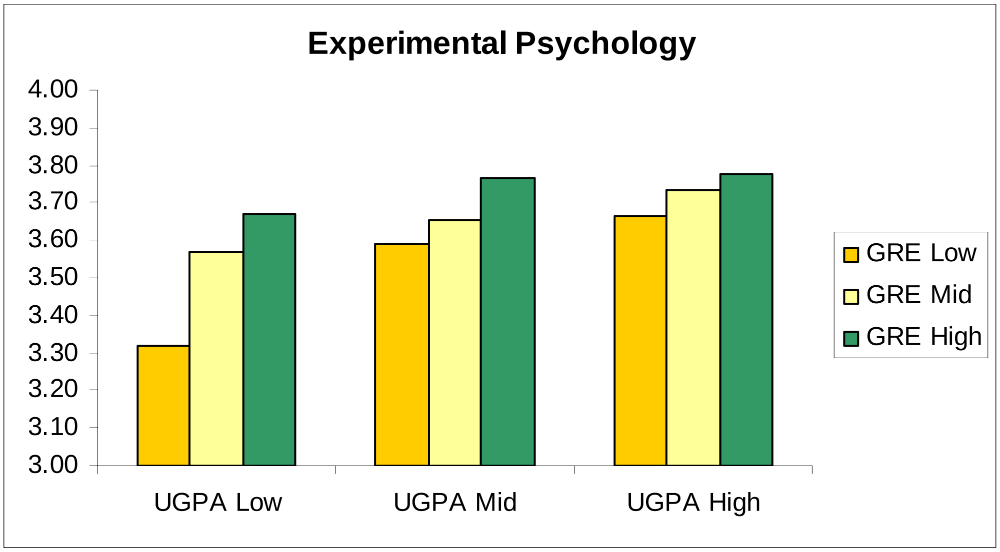

Figure 24. **Mean graduate GPA in experimental psychology departments by undergraduate GPA (UGPA) and GRE quartiles.** [6]

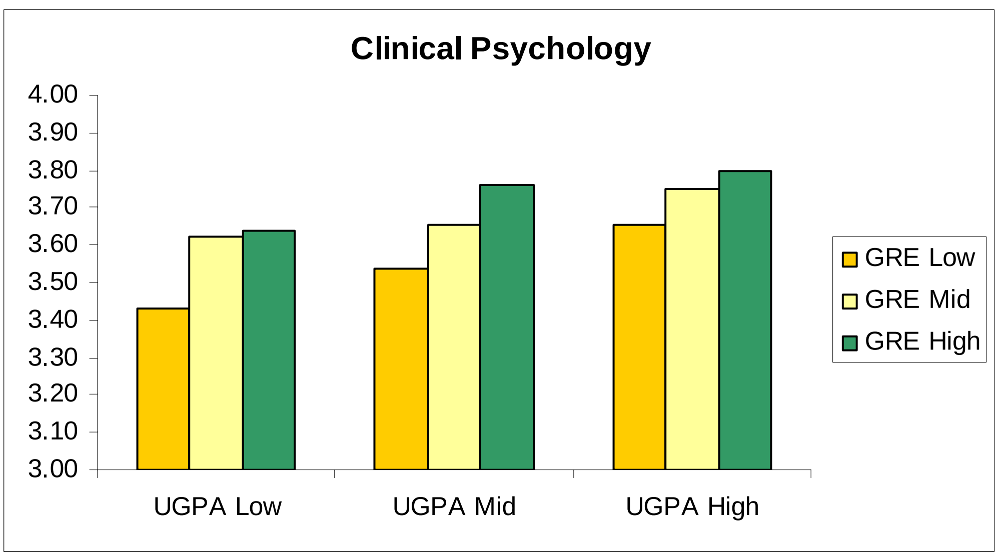

Figure 25. **Mean graduate GPA in clinical psychology departments by undergraduate GPA (UGPA) and GRE quartiles.** [6]

Figures 26–31 show the percentage of students earning a 4.0 first-year graduate GPA for high and low GRE quartiles within high and low UGPA quartiles. These figures address the question of whether knowing the UGPA quartile is sufficient for predicting who might get a 4.0 or whether the GRE assists in this prediction. If the GRE adds nothing, then the two bars on the left side of each graph should be the same height, indicating that in the bottom UGPA quartile students with high or low GRE scores are equally likely to excel. Similarly, if the two bars on the right side of each graph are the same height, it would suggest that, among students in the top quartile of UGPA, GRE scores do not matter. But, in fact, GRE scores do appear to matter. In biology departments (see Figure 26), among the students in the bottom UGPA quartile and bottom GRE quartile, not one student completed the year with a 4.0. Staying in the bottom UGPA quartile, but considering students who were also in the top GRE quartile, the rate of students earning a 4.0 jumped to 18%. Similarly, among students in the top UGPA quartile, 13% of the students who were in the bottom GRE quartile earned 4.0 first-year GPAs, but 28% in the top GRE quartile reached this distinction. In some departments, the GRE seemed to make a difference at one end of the scale but not at the other. In English departments (see Figure 29), among students with low UGPAs, differences in GRE quartile did not seem to matter, but among students with high UGPAs, students who also had high GRE scores were much more likely to be highly successful. In experimental psychology departments (see Figure 30), the opposite pattern was observed—essentially no difference by GRE score at the high UGPA level, but a substantial difference at the low level.

Figures similar to Figures 26–31, but for students earning a 3.8 or better, are in the appendix. Percentages of students meeting this success criterion are much higher, reaching as high as 80% in some departments. Nevertheless, the basic conclusion remains unchanged; even within a UGPA quartile, students with high GRE scores are markedly more successful than students with low scores.

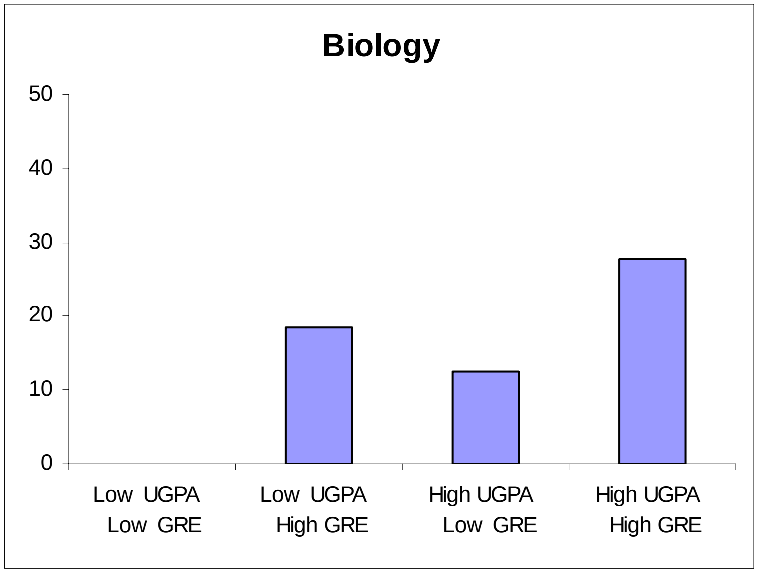

Figure 26. **Percentage of students earning a 4.0 in biology departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

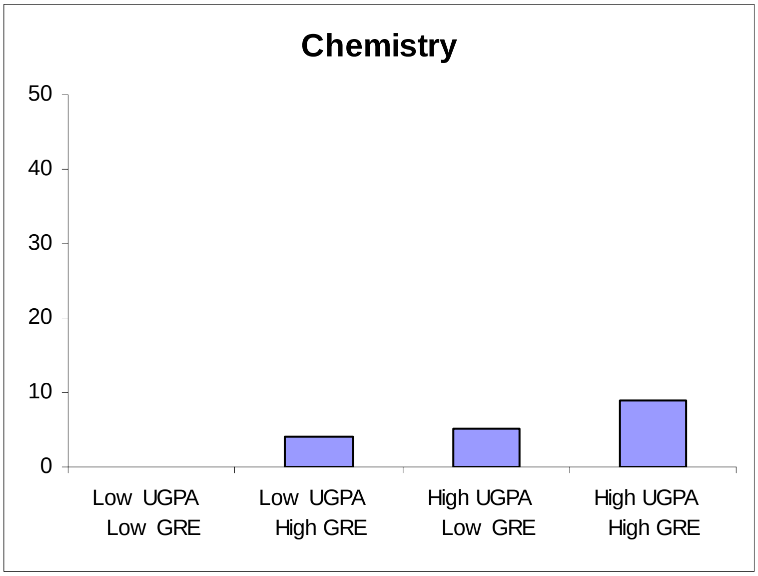

Figure 27. **Percentage of students earning a 4.0 in chemistry departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

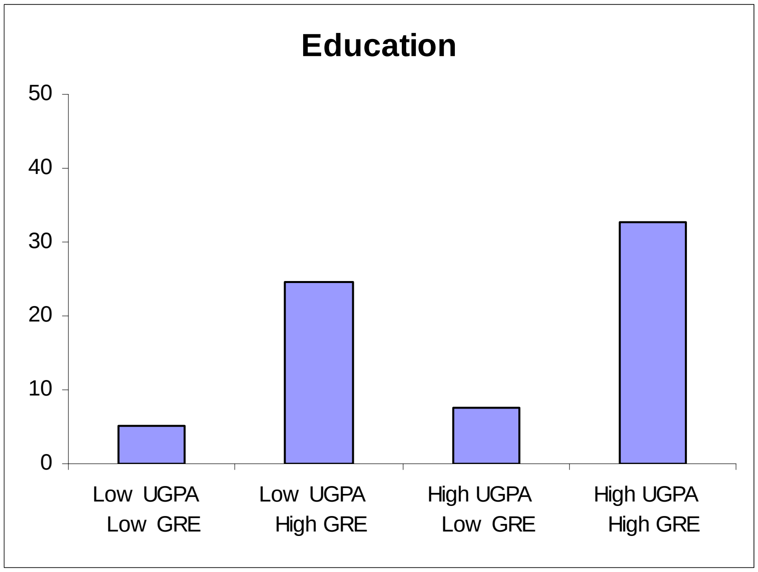

Figure 28. **Percentage of students earning a 4.0 in education departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

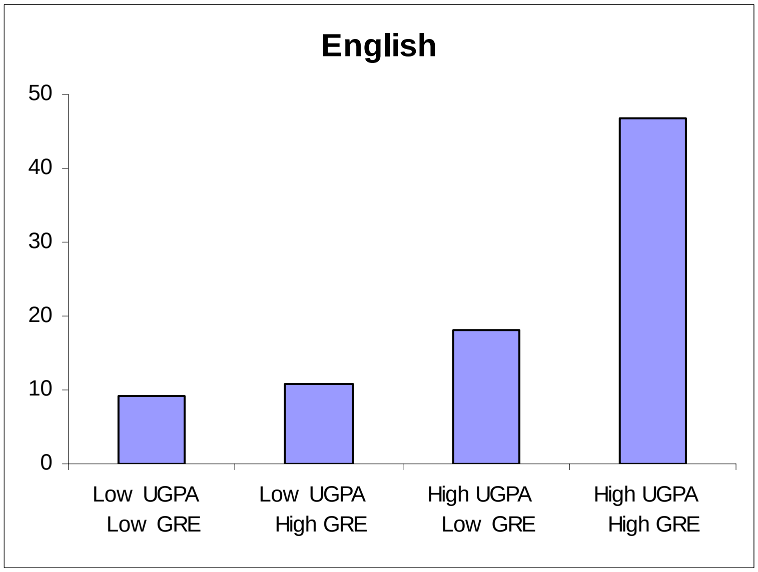

Figure 29. **Percentage of students earning a 4.0 in English departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

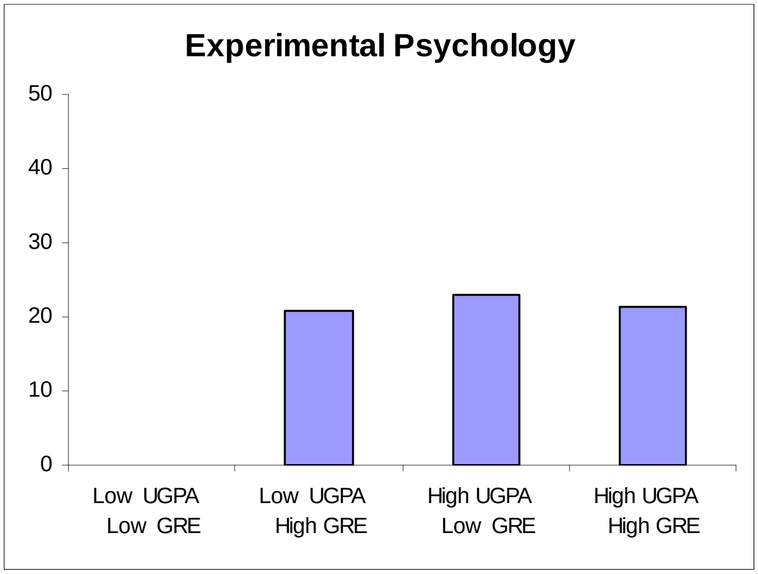

Figure 30. **Percentage of students earning a 4.0 in experimental psychology departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

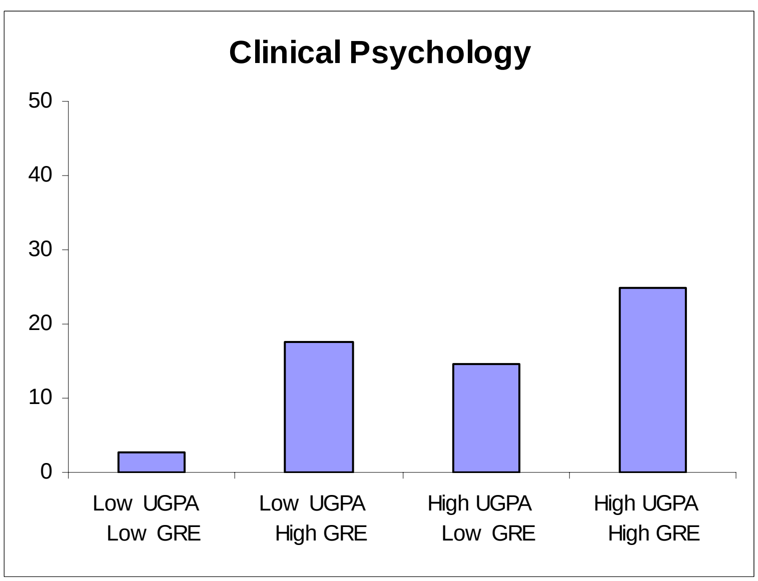

Figure 31. **Percentage of students earning a 4.0 in experimental psychology departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

Conclusion

Although correlations can be useful summary statistics, they are not particularly useful for conveying information on the utility of admissions tests, especially to nontechnical audiences. A test that explains only 9% of the variance in grades may appear to lack validity because a percentage of a variance is a difficult quantity to picture. Using unstandardized regression weights to express differences directly in grade point units may help some, but because of the highly restricted range of graduate grades, apparently large differences in test scores predict grade averages that differ by only hundredths of a point. A clearer indicator of the potential value of test scores comes from a comparison of success rates among students with high and low test scores. Although a 4.0 FYA is not the only indicator of a successful student, it is nevertheless a significant academic accomplishment. It is therefore meaningful to observe that this accomplishment is much more likely among students with relatively high GRE scores. In biology departments, for example, students in the top GRE quartile were 5 times as likely to earn a 4.0 as students in the bottom GRE quartile. “Five times as likely” carries a very different message than “9% of the variance.”

Because all of the analyses presented here are on students who were already admitted and enrolled, they probably understate the value of the test scores. If our bottom quartile could include estimates of the success of applicants who were rejected, the differences between the bottom and top quartiles would doubtless be larger.

For this study our focus was on ways of displaying validity information. For this purpose, it was sufficient to focus on the criterion data that we had most easily available: graduate grades. We recognize that other criteria such as graduation rates and professional productivity are equally or more important, and we believe that the analysis approaches we used could be easily adapted to these other criteria. Similarly, we are aware separate analyses by gender and ethnic groups would be valuable and believe that these areas need further research.

Only two scores from the GRE General Test were included in these analyses as the data were collected before the addition of the analytical writing score. Future research should include these scores as well as provide data on the revised GRE that is being introduced in 2006.

References

Bowen, W. G., & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.

Bridgeman, B., Pollack, J., & Burton, N. (2003). *Understanding what SAT I scores add to high school grades: A straightforward approach.* (College Board Rep. No. 2004-4). New York: College Entrance Examination Board.

Burton, N. W., & Wang, M. (2005). *Predicting long term success in graduate school: A collaborative validity study* (ETS GRE Board Research Rep. No. 99-14). Princeton, NJ: ETS.

Falk, R., & Well, A. D. (1997). Many faces of the correlation coefficient. *Journal of Statistics Education, 5*. Retrieved June 28, 2007, from the American Statistical Association Web site: http://www.amstat.org/publications/jse/v5n3/falk.html

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.

The National Center for Fair and Open Testing (FairTest). (2001, October 18). *Examining the GRE: Myths, misuses, and alternatives.* Retrieved May 30, 2008, from http://www.fairtest.org/facts/gre.htm

Noble, J. (2004). The effects of using ACT composite scores and high school averages on college admissions decisions for ethnic groups. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 303-319). New York: RoutledgeFalmer.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166–169.

Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher, 32*, 22–30.

Notes

[1] Note that one of the reasons this difference appears to be so small is that the SAT scale, like the GRE scale, has an extra 0 at the end. Without this extraneous 0, the quote would read, “an additional 10 points…” would be related to a 5.9% jump in percentile rank, which appears to be a much more significant difference.

[2] In the VSS database, we had only the institution’s definition of a first-year GPA. In the Burton and Wang (2004) database, we had information on individual courses taken and so could account for students who took only a few courses per year. For these students, we defined “first year” average as the average of the first eight courses taken even if this stretched over more than one year.

[3] The choice of 100 for GRE scores and 0.25 for UGPA can also be justified on psychometric grounds. The standard deviation of the combined GRE score is about 180 and the standard deviation of UGPA is 0.45. In standard deviation units, the 100-point GRE difference is equivalent to a 0.25 UGPA difference, as both reflect a difference of 0.56 standard deviation units.

[4] GRE low is bottom quartile within a department and GRE high is top quartile; FYA low is bottom quartile for first-year average within a department and FYA high is top quartile.

[5] On one hand, a difference of only 0.31 grade points within a UGPA category may seem trivial, but with the highly restricted range of grades, this still represents a difference of three-quarters of a standard deviation.

[6] UGPA low is bottom quartile within a department, UGPA mid is middle 50%, and UGPA high is top quartile.

Appendix

Percentage of Students Earning a 3.8 First-Year Grade Point Average (GPA) by Undergraduate GPA (UGPA) and GRE Top and Bottom Quartiles for Biology, Chemistry, Education, English, Experimental Psychology, and Clinical Psychology Departments

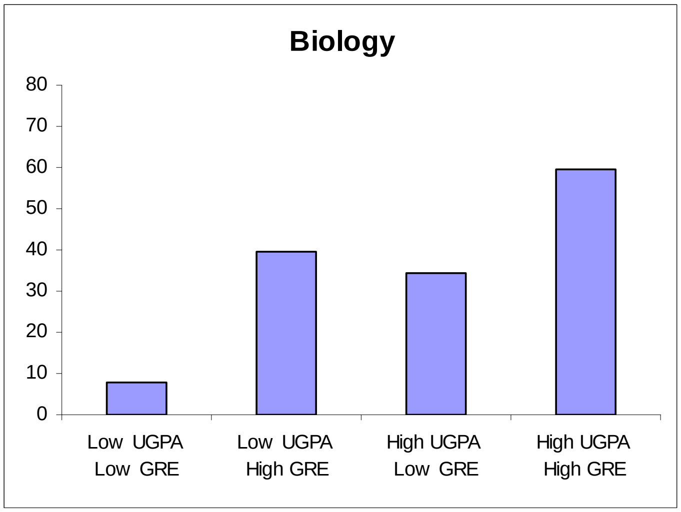

Figure A1. **Percentage of students earning a 3.8 or better in biology departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

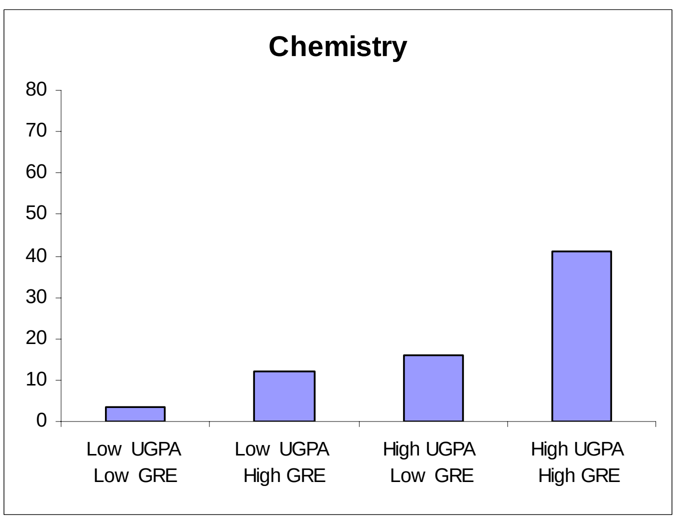

Figure A2. **Percentage of students earning a 3.8 or better in chemistry departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

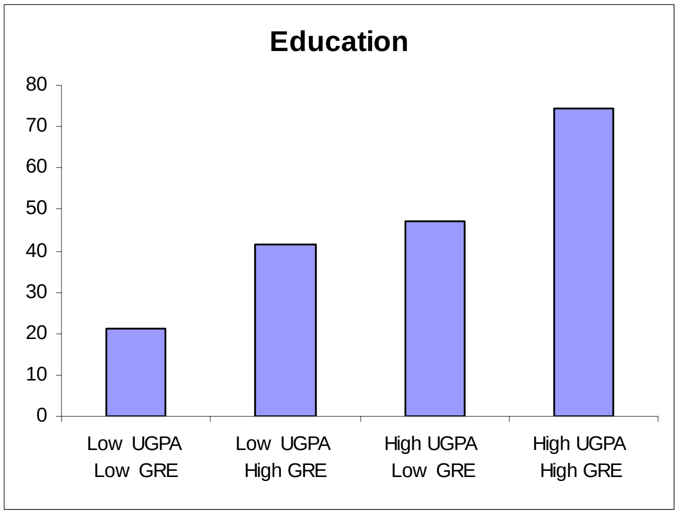

Figure A3. **Percentage of students earning a 3.8 or better in education departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

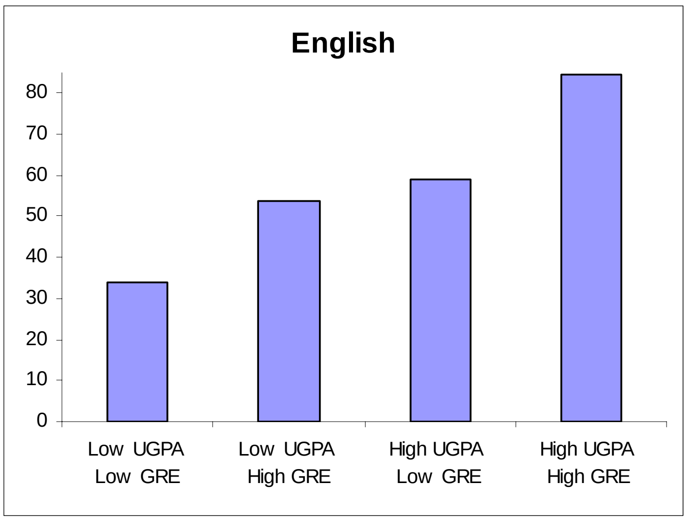

Figure A4. **Percentage of students earning a 3.8 or better in English, departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

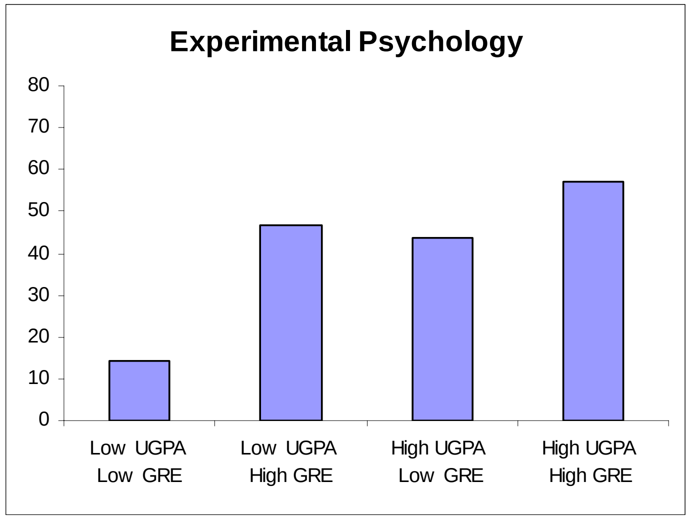

Figure A5. **Percentage of students earning a 3.8 or better in experimental psychology departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

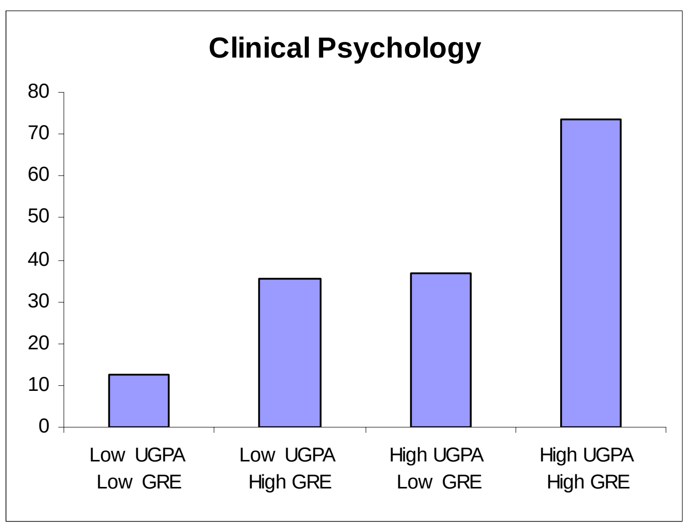

Figure A6. **Percentage of students earning a 3.8 or better in clinical psychology departments by undergraduate GPA (UGPA) and GRE high and low quartiles.**

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