

ETS GRE[®] Board Research Report

ETS GRE[®] – 14-01

ETS Research Report No. RR–14-17

The Validity of Scores from the GRE[®] revised General Test for Forecasting Performance in Business Schools: Phase One

John W. Young

David Klieger

Jennifer Bochenek

Chen Li

Fred Cline

December 2014

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

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in Graduate Record Examinations Board reports do not 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.

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



RESEARCH REPORT

The Validity of Scores from the GRE[®] revised General Test for Forecasting Performance in Business Schools: Phase One

John W. Young, David Klieger, Jennifer Bochenek, Chen Li, & Fred Cline

Educational Testing Service, Princeton, NJ

Scores from the GRE[®] revised General Test provide important information regarding the verbal and quantitative reasoning abilities and analytical writing skills of applicants to graduate programs. The validity and utility of these scores depend upon the degree to which the scores predict success in graduate and business school in specific contexts. To assess the predictive validity of the GRE test for graduate business programs, we collaborated with a number of universities and obtained data from each university's admissions office and registrar. We focused specifically on part-time and full-time students in master's of business administration (MBA) degree programs. Given the nested structure of the data, we used a 2-level (representing students and institutions) hierarchical linear model (HLM) to estimate regression models with first-semester MBA grade point average (GPA) or cumulative MBA GPA as the dependent variable and GRE scores and undergraduate GPA (UGPA) as independent variables. For predicting cumulative MBA GPA, the pseudo R-squared (R^2) value was .04 using UGPA, cohort year, and program type as the only predictors; this value increased substantially to .19 with the addition of GRE scores. The HLM results show that both GRE-Quantitative and GRE-Verbal scores were statistically significant predictors of both first-semester MBA GPA and cumulative MBA GPA.

Keywords GRE; predictive validity; business schools; hierarchical linear modeling; graduate school

doi:10.1002/ets2.12019

In Kane's (2006) argument-based approach to test validity theory, evidence of validity for specific interpretations of test scores can take many forms. For a large-scale assessment program, an agenda for research should incorporate the interpretative and validity arguments in Kane's framework. However, any single validity research study can focus on only a few limited aspects of these arguments. One category of empirical evidence consists of criterion-related studies that investigate the statistical relationship between scores on a test and later performance on an outcome measure of interest. For example, for assessments used for admissions and selection decisions, an important form of validity evidence derives from predictive validity studies that show the degree of association between test scores prior to selection and later relevant outcome measures. This study is designed specifically to obtain evidence on the predictive validity of scores from the GRE[®] revised General Test for forecasting performance in business schools, specifically for students enrolled in master's of business administration (MBA) programs.

Earlier versions of the GRE test, as well as the current GRE revised General Test, have been shown to be valid and reliable measures of the skills needed to succeed in graduate school (see the following section for a brief summary of research on the GRE test). Scores from the GRE revised General Test provide important information regarding the verbal and quantitative reasoning abilities and analytical writing skills of applicants to graduate programs. Given that these same skills are critical for success in business schools, we theorize that the predictive validity of the GRE revised General Test for this student population will be as strong as, if not stronger than, that found for graduate school populations. This study is the first large-scale predictive validity study of scores from the GRE revised General Test for this population of GRE test takers. For this report, we investigated the relationships between scores from the GRE revised General Test and certain measures of students' achievement in MBA programs (first-semester MBA grade point average [GPA] and cumulative MBA GPA). Given the increasing acceptance of GRE scores in graduate business school programs, this study provides evidence on the utility and validity of scores from the GRE revised General Test in the admissions process for MBA programs. This evidence is of significant value to institutional users and prospective users of GRE scores in the global business school community.

Corresponding author: J.W. Young, E-mail: jwyong@ets.org

The GRE revised General Test

The GRE revised General Test was launched in August 2011. The GRE General Test was first introduced in 1949 and has undergone a number of modifications over the years. Based on input from the GRE Board and the graduate school community, the revisions that led to the creation of the GRE revised General Test represent one of the largest changes in the history of the test. The GRE revised General Test provides three section scores: Verbal Reasoning (GRE-V), Quantitative Reasoning (GRE-Q), and Analytical Writing (GRE-AW). Scaled scores on the GRE-V and the GRE-Q sections range from 130 to 170, whereas GRE-AW scores are reported on a scale from 0 to 6 in half-point increments. At present, the GRE revised General Test is the world's most widely administered graduate-level admissions test. The number of business schools worldwide accepting GRE scores for their MBA programs has grown rapidly in recent years, and more applicants to MBA programs are now beginning to submit GRE scores for admissions. Currently, more than 1,100 business schools around the world accept GRE scores for one or more of their MBA degree programs (Educational Testing Service, 2014a). At present, the majority of applicants submit GMAT scores when applying for admissions into graduate business schools (Educational Testing Service, 2014b).

Research on the Validity of GRE Scores for Predicting Graduate School Performance

The predictive validity of GRE scores has been widely studied over the past several decades with research dating back to the 1940s (Broadus & Elmore, 1983; Cureton, Cureton, & Bishop, 1949; Sleeper, 1961; Stricker & Huber, 1967). Previous empirical and meta-analytical studies have provided a wealth of supporting evidence on the validity of GRE scores from earlier versions of the test (e.g., Burton & Wang, 2005; Klieger, Cline, Holtzman, Minsky, & Lorenz, 2014; Kuncel & Hezlett, 2007; Kuncel, Hezlett, & Ones, 2001; Kuncel, Wee, Serafin, & Hezlett, 2010; Powers, 2004). However, only a few of these studies examined the relationship of GRE scores with performance in graduate business programs.

The large-scale meta-analysis by Kuncel et al. (2001) included more than 82,000 students from over 1,700 samples. They found that GRE-V and GRE-Q scores were good predictors of cumulative graduate GPA (GGPA) and correlated as well as with this criterion measure, if not better than, undergraduate GPA (UGPA). In this meta-analysis, the authors examined the relationship between GRE scores and graduate school performance by subgroups of students in different disciplines. These disciplines were categorized into four broad areas: humanities, social sciences (which included business), life sciences, and mathematics/physical sciences. Not surprisingly, GRE-V scores were found to be better predictors of GGPA for humanities and social science students than GRE-Q scores. GRE-V and GRE-Q scores predicted GGPA about equally well for life science students, whereas GRE-Q scores were found to be better predictors of GGPA for mathematics/physical science students than were GRE-V scores.

In the context of graduate programs in professional schools, Powers (2004) investigated the predictive validity of GRE scores in colleges of veterinary medicine. In his study of 16 institutions, he found that the average correlation between first-year GGPA and GRE scores was .30 for GRE-V and .44 for GRE-Q (these values were corrected for restriction of range and criterion unreliability). Given the similarity in knowledge and skills required for students in graduate mathematics/physical science programs compared to that for students in veterinary medicine, this finding is consistent with results from Kuncel et al. (2001). Holt, Bleckmann, and Zitzmann (2006) evaluated the validity of the GRE for students in an engineering management program at the Air Force Institute of Technology. GRE-V correlated with first-year GGPA at .37, GRE-Q at .20, and GRE-AW at .18, but UGPA did not correlate significantly with first-year GGPA ($r = .12$). These values were not corrected for restriction of range and criterion unreliability, and we did not have the information to compute the corrected values. Milner, McNeil, and King (1984) found that total GRE scores were correlated significant with class GPA ($r = .24$) for students seeking professional degrees in social work. Observed correlations of GRE scores with other outcomes (GPA in courses specific to the field, degree attainment, and retention rate) ranged from .03 to .11 and were not statistically significant. These values were not corrected for restriction of range and criterion unreliability, and we did not have the information to compute the corrected values. In a study of engineering graduate students (Wang, 2013), first-year GGPA correlated with GRE-V at .17 and GRE-Q at .22, but the test scores actually had higher correlations with GGPA (GRE-V, $r = .21$; GRE-Q, $r = .26$).

A recent study investigated the predictive validity of GRE scores for graduate programs in 10 Florida public universities. One significant finding from this study (Klieger et al., 2014) is that the validity of GRE-AW scores for predicting GGPA for master's students was .19 overall and .15 for business, management, and marketing programs. It was not possible

to determine whether any of these students were enrolled in MBA programs or business schools. These validity coefficients were corrected for range restriction but not for measurement error in either the predictor or criterion. Owing to measurement error in the criterion (Cronbach's alpha coefficients were .71–.72), the sizes of the true validity coefficients were attenuated by 15–16%. For this Florida sample, GRE-AW was generally as strong as GRE-V and GRE-Q for predicting GGPA.

In summary, a wealth of evidence already exists regarding the predictive validity of the GRE test (both previous and current versions) for forecasting performance across a broad range of graduate programs over a number of years. This study will add to the existing literature by providing evidence of the GRE revised General Test for forecasting the performance of students enrolled in business schools, specifically for a sample of students in MBA programs.

Hierarchical Linear Modeling (HLM)

By simultaneously modeling students at a first level and institutions at a second level, we account for the fact that student characteristics and the determinants of business school performance could differ across institutions. Given that each institution may have a unique effect upon its students, in using hierarchical linear modeling (HLM), we avoid violating the assumption in multiple regression analysis that errors across individuals are uncorrelated (see Raudenbush & Bryk, 2002). Because students were nested within particular institutions, we employed two-level HLM models to estimate regression effects, with MBA GPA (first semester or cumulative) as the dependent variable and GRE scores and UGPA as independent variables. Cohort year (2012–2013 or 2013–2014) and program type (part time or full time) served as covariates. We did not treat cohort year or program type as additional model levels because of the limited number of institutions and students and a desire to achieve the most parsimonious and interpretable results. The intraclass correlation coefficients (ICC) for the null models were large enough to support a multilevel approach. Although we theorized that models with random intercepts and slopes for Level 2 predictors could be more appropriate, we ultimately selected models limited to random intercepts because they fit the data better.

Research Methods

Research Question

In this study, the main research question is as follows: What is the predictive validity of scores from the GRE revised General Test (alone and in conjunction with UGPA) for forecasting significant educational outcomes for MBA students?

Samples and Institutions

Because the GRE revised General Test was first administered in August 2011, the first uses of these scores for admissions are for students applying for the spring semester of 2012 and later. Thus, we studied full-time and part-time MBA students who matriculated in the 2012–2013 and 2013–2014 academic years. Institutions were recruited for this study through communications from the GRE program to institutional representatives of MBA programs. Institutional data were sent in two rounds: The first round of data collection was completed in Winter 2014 and included admissions and performance data through the fall semester of 2013.¹

Overall, 12 institutions participated in this first phase of the study. Institutional admissions selectivity ranged from 21% to 66% in terms of the percentage of applicants admitted. Together, participating institutions submitted data for 16 programs (full time or part time) for a total of 30 unique samples (or groups) of students when organized by enrollment status and cohort year (see Table 1). All of the MBA programs in this study are traditional face-to-face programs with none of the programs being online or executive MBA programs. The study includes a total of 480 students (across all programs and both cohort years) with GRE revised General Test scores, with 61.7% from the 2013–2014 cohort year, 70.6% from part-time programs, and 42.7% being female. Students who only had scores from earlier versions of the GRE General Test or who did not have any admissions test scores were not included in our analyses.

The first cohort year consists of students with initial matriculation in the 2012–2013 academic year. For these students, their first-semester MBA GPA is based on courses taken in Fall 2012, whereas their cumulative MBA GPA includes all courses taken from the fall of 2012 through Fall 2013. This includes up to three semesters of taking courses as well as any

Table 1 Number of Samples by Enrollment Status and Cohort Year

Enrollment status	Cohort year	
	2012–2013	2013–2014
Part time	10	11
Full time	3	6

courses completed in the intervening summer term. The second cohort year consists of students with initial matriculation in the 2013–2014 academic year. For these students, their first-semester MBA GPA is based on courses taken in Fall 2013, whereas their cumulative MBA GPA includes these courses plus any transfer courses.

Statistical Analyses

We used standard statistical methods for this study, including computing summary statistics and correlation coefficients. In addition, we employed two-level HLM models to estimate regression models with first-semester MBA GPA and cumulative MBA GPA as separate dependent variables and the three GRE revised General Test scores (GRE-V, GRE-Q, and GRE-AW) and UGPA as independent variables. We also included program type and cohort year as student-level variables as we hypothesized that these variables may moderate the relationships between GRE revised General Test scores and the MBA GPAs. The second level of the HLM model represents the different institutions in the study.

Results

Summary Statistics

Descriptive information on the students included in the first data collection is shown in Table 2. Additionally, in comparison to the norms for all US GRE revised General Test test takers, the GRE revised General Test test takers in this study scored higher on all three sections as the 2012–2013 national averages were 150.6 for GRE-V, 152.2 for GRE-Q,

Table 2 Characteristics of Participants

	GRE test takers (<i>N</i> = 480)	
	<i>N</i>	%
Cohort		
2012–2013	184	38.3
2013–2014	296	61.7
Program		
Full time	141	29.4
Part time	339	70.6
Gender		
Male	275	57.3
Female	205	42.7
	<i>M</i>	<i>SD</i>
GRE scores		
Verbal reasoning	156.68	6.66
Quantitative reasoning	155.00	6.26
Analytical writing	4.10	0.83
UGPA	3.34	0.39
MBA GPA		
First-semester GPA	3.27	0.50
Cumulative GPA	3.30	0.43

Note. GPA = grade point average; MBA = master's of business administration; UGPA = undergraduate GPA.

Table 3 Weighted Correlations and 95% Confidence Intervals of GRE Scores and Undergraduate Grade Point Average (UGPA) With Master's of Business Administration (MBA) GPAs

	GRE revised General Test Weighted correlations (95% confidence intervals)			UGPA
	Verbal reasoning	Quantitative reasoning	Analytical writing	
First-semester MBA GPA	.22 (.13–.31)	.39 (.32–.50)	.07 (–.02–.16)	.13 (.04–.22)
Cumulative MBA GPA	.24 (.15–.34)	.37 (.30–.48)	.11 (.02–.20)	.19 (.10–.28)

and 3.5 for GRE-AW (Educational Testing Service, 2013). For those institutions for which we had course-level information, we computed summary statistics on the number of courses that were included in the first-semester MBA GPA and cumulative MBA GPA. For first-semester MBA GPA, the mean number of courses per student was 2.74 (with a standard deviation of 1.70); for cumulative MBA GPA, the mean number of courses was 5.41 (with a standard deviation of 3.45).

Correlations

We computed average correlation coefficients between each of the four admissions variables (GRE-V, GRE-Q, GRE-AW, and UGPA) and the two outcome variables (first-semester MBA GPA and cumulative MBA GPA). The correlations were computed by first calculating the correlations within each of the 30 student groups and then averaging across groups weighted by sample size. Table 3 shows that, of the four admissions variables, GRE-Q has the highest averaged correlation with each of the two outcome variables: 0.39 with first-semester MBA GPA and 0.37 with cumulative MBA GPA (these values have not been corrected for restriction of range or criterion unreliability).

Hierarchical Linear Modeling (HLM) Results

For each of the dependent variables (first-semester MBA GPA and cumulative MBA GPA), we first conducted analyses of model fit. Although we believed that a multilevel approach was reasonable because of students being nested within institutions, we sought statistical evidence to support a multilevel approach. In our analyses, we included different combinations of admissions variables in order to model their possible uses in admissions: (a) GRE-V, GRE-Q; (b) GRE-V, GRE-Q, GRE-AW; (c) GRE-V, GRE-Q, GRE-AW, UGPA; and (d) UGPA only. Cohort year and program type served as covariates at the student level. For both dependent variables, we first examined ICCs of the null versions of the models (multilevel models with no predictors) to ascertain whether multilevel models of any kind were appropriate. As Table 4 shows, ICCs range from 38% to 42% across null models, indicating that variance between institutions is large enough to justify multilevel models for all predictor combinations for both dependent variables (see Luke, 2004; Raudenbush & Bryk, 2002; Snijders & Bosker, 1994, 1999).

We believed that the multilevel models should include the three GRE section scores and UGPA as Level 2 predictors because we expected the school-level means of these variables to moderate the Level 1 relationships of these predictors with the MBA GPA criterion. In other words, because of effects such as differential grading and possible range restriction, institutions having students with higher mean GRE revised General Test scores and undergraduate grades may have different predictor–criterion relationships at the student level than do institutions having students with lower mean GRE revised General Test scores and undergraduate grades. In Tables 4 and 5, we exclude GRE-AW from Level 2 of Models 1-3 and 2-3 to avoid the multicollinearity that prevented the solutions from converging. The reason for the differences in the numbers of schools and students across models is the absence of GRE-AW scores for two institutions that do not use these scores in their admissions decisions. We recognized that the models are not based on the same samples of students in all cases. However, we wished to maximize the stability of the models to the greatest extent possible, so we opted to use data for all 12 schools whenever possible. We had no a priori theoretical reason to believe that the validity findings for the two omitted schools would differ from that of the other 10 schools.

Moreover, we theorized that the multilevel models should include random slopes and not just random intercepts because the nature of the predictor–criterion relationships could vary across institutions. However, empirical findings did not support the use of random slopes models. As Table 4 indicates, Level 1 and Level 2 errors change relatively little as a function of whether slopes are modeled as random in the models. Deviance values (–2LL) are larger when slopes are

Table 4 Fit of Multilevel Models

Model	DV	Level 1 predictors ^a	Level 2 predictors ^b	N	k	ICC	Intercepts as outcomes models (random intercepts only)						Intercepts and slopes as outcomes models (random intercepts and random slopes)					
							Level 1			Level 2			Level 1			Level 2		
							error (σ^2)	error (τ^2)	-2LL	p	AIC	BIC	error (σ^2)	error (τ^2)	-2LL	p	AIC	BIC
1-1	FSGPA	C P V Q	MeanV meanQ	449	12	41%	0.108	0.067	344	2	348	356	0.105	0.067	397	16	429	494
1-2	FSGPA	C P V Q AW	MeanV meanQ meanAW	369	10	42%	0.110	0.081	299	2	303	310	0.106	0.081	361	22	405	491
1-3	FSGPA	C P V Q AW U	MeanV meanQ meanU ^c	335	10	38%	0.114	0.070	286	2	290	298	0.108	0.061	349	29	407	517
1-4	FSGPA	C P U	MeanU	413	12	38%	0.127	0.077	365	2	369	377	0.124	0.077	349	11	371	416
2-1	GGPA	C P V Q	MeanV meanQ	480	12	40%	0.094	0.057	297	2	301	309	0.089	0.058	346	16	378	444
2-2	GGPA	C P V Q AW	MeanV meanQ meanAW	397	10	41%	0.096	0.064	266	2	270	278	0.090	0.065	321	22	365	452
2-3	GGPA	C P V Q AW U	MeanV meanQ meanU ^c	354	10	39%	0.099	0.065	251	2	255	262	0.090	0.058	310	29	368	480
2-4	GGPA	C P U	MeanU	435	12	38%	0.111	0.071	324	2	328	336	0.106	0.071	308	11	330	375

Note. DV = dependent variable; FSGPA = first-semester MBA GPA; GGPA = cumulative MBA GPA; C = cohort year (2012–2013 [coded 1] or 2013–2014 [coded 2]); P = program type (full time [coded 1] or part time [coded 2]); V = GRE-Verbal; Q = GRE-Quantitative; AW = GRE-Analytical Writing; U = undergraduate GPA; N = number of students; k = number of schools; ICC = intraclass correlation coefficient for null version of the models; -2LL = deviance = $-2 \times \log$ likelihood; p = # of parameters; AIC = Akaike information criterion; BIC = Schwarz's Bayesian information criterion. ^aLevel 1 predictors were group-centered under the assumption that it is a student's score or GPA relative to that of other students in the school that is most salient here (see discussion in Snijders & Bosker, 1999, p. 81). ^bLevel 2 predictors for the intercepts and slopes as outcomes models were grand mean centered to permit solutions to converge. To the extent possible, the Level 1 predictors for a model were included as group means in Level 2 of that model. ^cAnalytical writing is absent as a predictor in the second level of Models 1-3 and 2-3, because inclusion of that section contributed to multicollinearity that prevented these models from converging on a solution.

Table 5 Statistical Significance of Multilevel Predictors, Multilevel Variance Components, and Multilevel Reduction in Prediction Error (Pseudo R-Squared Values)

Model	DV	Level 1 predictors ^a	Level 2 predictors ^b	N	k	Null model ^d				Full model				Pseudo R-squared values			
						Level 1 error (σ^2)		Level 2 error (τ^2)		Level 1 error (σ^2)		Level 2 error (τ^2)		Raudenbush and Bryk (2002)		Snijders and Bosker (1999)	
						error (σ^2)	error (τ^2)	error (σ^2)	error (τ^2)	error (σ^2)	error (τ^2)	R_1^2	R_2^2	R_1^2	R_2^2	R_1^2	R_2^2
1-1	FSGPA	C P V*** Q***	Mean V mean Q	449	12	0.126	0.087	0.108	0.067	0.145	0.224	0.177	0.219	21.614			
1-2	FSGPA	C P V*** Q*** AW	Mean V mean Q mean AW	369	10	0.132	0.095	0.110	0.081	0.168	0.139	0.156	0.141	18.941			
1-3	FSGPA	C P V*** Q*** AW U*	Mean V mean Q mean U ^c	335	10	0.137	0.084	0.114	0.070	0.171	0.170	0.170	0.170	17.431			
1-4	FSGPA	C P U***	Mean U	413	12	0.130	0.079	0.127	0.077	0.020	0.022	0.021	0.022	19.175			
2-1	GGPA	C* P V*** Q***	Mean V mean Q	480	12	0.111	0.074	0.094	0.057	0.153	0.226	0.182	0.222	26.146			
2-2	GGPA	C* P V*** Q*** AW	Mean V mean Q mean AW	397	10	0.116	0.080	0.096	0.064	0.171	0.200	0.183	0.198	23.482			
2-3	GGPA	C P V*** Q*** AW U**	Mean V mean Q mean U ^c	354	10	0.122	0.077	0.099	0.065	0.187	0.154	0.174	0.156	21.771			
2-4	GGPA	C P U***	Mean U	435	12	0.115	0.072	0.111	0.071	0.039	0.010	0.028	0.012	24.100			

Note. DV = dependent variable; FSGPA = first-semester MBA GPA; GGPA = cumulative MBA GPA; C = cohort year (2012–2013 [coded 1] or 2013–2014 [coded 2]); P = program type (full time [coded 1] or part time [coded 2]); V = GRE-Verbal; Q = GRE-Quantitative; AW = GRE-Analytical Writing; U = undergraduate GPA; N = number of students; k = number of schools; R_1^2 = the proportional reduction in error for predicting student-level FSGPA or GGPA; R_2^2 = the proportional reduction in error for predicting school-level FSGPA or GGPA; H = harmonic mean used in pseudo R-squared calculations per Snijders and Bosker (1999).

^aLevel 1 predictors were group-centered under the assumption that it is a student's score or GPA relative to that of other students in the school that is most salient here (see discussion in Snijders & Bosker, 1999, p. 81). ^bLevel 2 predictors for the intercepts and slopes as outcomes models were grand mean centered to permit solutions to converge. To the extent possible, the Level 1 predictors for a model were included as group means in Level 2 of that model. ^cAnalytical writing is absent as a predictor in the second level of Models 1-3 and 2-3 because inclusion of that section contributed to multicollinearity that prevented these models from converging on a solution. ^dThe null model has two levels but does not include any predictor variables; the full model is intercepts as outcomes (random intercepts only). * $p \leq 0.05$. ** $p \leq 0.01$. *** $p \leq 0.001$.

random, and the values for the Akaike information criterion (AIC) and Schwartz's Bayesian information criterion (BIC) fit indices are always larger for the random slopes models (see discussion in Luke, 2004, about use of these fit indices). Therefore, considerations of model parsimony led us to choose the random intercepts models as our best fitting models.

Once we selected these models, we sought to determine if the predictors had statistically significant relationships to either of the MBA GPAs. As expected, GRE revised General Test scores were always statistically significant predictors across the models. Furthermore, we used the methods of Raudenbush and Bryk (2002) and Snijders and Bosker (1994, 1999) to estimate values of pseudo R-squared (R^2) to communicate the multilevel equivalent of effect sizes. Across the models, GRE revised General Test section scores (particularly GRE-V and GRE-Q) always reduced error for predicting both first-semester MBA GPA and cumulative MBA GPA by a substantial degree—and far more than UGPA did.

As Table 5 shows, statistically significant relationships exist at the student level only. Institutional mean GRE revised General Test scores and mean UGPA did not predict first-semester MBA GPA or cumulative MBA GPA to a statistically significant extent. Across all models, GRE-V and GRE-Q had statistically significant relationships ($p \leq .01$ or $.001$) to both first-semester and cumulative GPAs. Although by themselves GRE-AW scores were significantly related to cumulative GPA (see Table 3), with GRE-V and GRE-Q already in the equation, GRE-AW scores did not significantly contribute to GPA predictions. UGPA always had statistically significant relationships with both MBA GPAs ($p \leq .05$, $.01$, or $.001$). The effect of cohort year had a statistically significant relationship with cumulative MBA GPA ($p \leq .05$) for Models 2-1 and 2-2, indicating that the relationships of GRE revised General Test scores with cumulative MBA GPA are related to cohort year. Lastly, program type (part time vs. full time) was never a statistically significant predictor ($p > .05$).

Using two different formulations (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999), we calculated pseudo R-squared values to estimate how much a two-level modeling approach reduced the proportion of error for predicting first-semester MBA GPA and cumulative MBA GPA. Given that they are based on a multilevel model, pseudo R-squared values are analogous to, but not precisely the same as, R^2 measures of variance-accounted-for in ordinary least squares (OLS) regression approaches (Luke, 2004). In Table 5, R_1^2 indicates the proportional reduction in error for predicting individual outcomes (student-level MBA GPA); R_2^2 indicates the proportional reduction in error for predicting group means (school-level MBA GPA). As Table 5 shows, GRE-V and GRE-Q (and covariates), with or without GRE-AW or UGPA included in the models, reduce the error for predicting both individual and institution-level MBA GPAs (first semester and cumulative) by 14–23%. This was roughly analogous to OLS multiple correlation values (multiple R) of .37 to .48. UGPA did not significantly reduce the prediction error at either the individual or institution level when all three GRE revised General Test section scores were included in the model; sometimes, in predicting cumulative MBA GPA, the inclusion of UGPA actually increased prediction error. By itself, UGPA reduced prediction error only 1–4% for any model.² To summarize, the HLM results show that both GRE-Q and GRE-V scores are statistically significant predictors of both the MBA GPAs.

Discussion

Our results indicate that GRE revised General Test scores, particularly GRE-Q and GRE-V scores, have a high degree of predictive validity for forecasting the academic performance of students enrolled in MBA programs. This finding is accurate whether or not UGPA, along with GRE revised General Test scores, was included in the multilevel models. Moreover, our estimates of the predictive validity of GRE revised General Test scores would have been even higher if, as in previous GRE validity studies, we had adjusted the validity coefficients for the known statistical artifacts of range restriction and/or unreliability in the criterion variables (see, e.g., discussions in Klieger et al., 2014; Kuncel et al., 2001). Furthermore, the relative predictive validity for each of the GRE revised General Test sections aligns with our understanding of the curricula demands that students encounter in most MBA programs. Thus, the finding that GRE-Q is the single best predictor across programs is consistent with the strong emphasis on quantitative reasoning and skills in many MBA courses, particularly in the first year. In addition, the finding that GRE revised General Test scores have high overall predictive validity (as evidenced by the HLM results) was expected given the direct connections between the verbal and quantitative reasoning abilities and analytical writing skills assessed by the GRE revised General Test and the cognitive demands of MBA programs.

In terms of higher education policy for graduate admissions, the results of this study demonstrate that scores from the GRE revised General Test are valid for predicting the academic performance of students in graduate business schools,

specifically for a sample of students currently enrolled in MBA programs. For graduate business schools that currently accept GRE revised General Test scores, as well as for institutions that are considering accepting these scores, our findings provide evidence that can now be reviewed in order to potentially modify policies and practices with regard to candidate admissions into MBA degree programs. The validity evidence from this study indicates that, compared to other information that applicants may submit (e.g., UGPA), the GRE revised General Test captures information about a candidate that is well-aligned with skills desired by and necessary for success in MBA programs.

Note that these findings are based only on data from the first round of data collection; the second round of data collection will be completed by July 2014 and will include performance data for up to 2 years for the students in this study, including whether they completed their MBA degrees within that time frame.

Notes

- 1 The results we report are from this first data collection. The second data collection will be completed in July 2014 and will include data through the spring semester of 2014. For full-time students who began their MBA program in the fall of 2012, this data collection will include performance data for two full academic years, including whether they completed their degrees within that time frame.
- 2 We also conducted OLS regression analyses for the two outcome variables using the GRE section scores as predictors (but not UGPA) and added indicator variables for institutions, program type, and interaction terms for institutions by program type (we excluded cohort year as that was already shown to be not highly significant in the HLM analyses). We obtained multiple *R* values of .68 for first-semester MBA GPA and .66 for cumulative MBA GPA. Because the contributions of the individual-level variables cannot be separated from that of the institutional-level variables in these analyses, the HLM findings are more appropriate for gauging the effectiveness of GRE scores in predicting the MBA GPAs.

References

- Broadus, R. N., & Elmore, K. E. (1983). The comparative validities of undergraduate grade point average and of part scores on the Graduate Record Examinations in the prediction of two criterion measures in a graduate library program. *Educational and Psychological Measurement*, 43, 543–546.
- Burton, N. W., & Wang, M.-M. (2005). *Predicting long-term success in graduate school: A collaborative validity study* (GRE Board Research Report No. 99-14P). Princeton, NJ: Educational Testing Service.
- Cureton, E. E., Cureton, L. W., & Bishop, R. (1949). Prediction of success in graduate study in psychology at the University of Texas. *American Psychologist*, 4, 361–362.
- Educational Testing Service. (2013). *A snapshot of the individuals who took the GRE revised General Test*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (2014a, May 19). *MBA programs that accept the GRE revised General Test*. Retrieved from http://www.ets.org/gre/revised_general/about/mba/programs?WT.ac=rx42
- Educational Testing Service. (2014b). *More MBA applicants are choosing the GRE revised General Test* [Press release]. Retrieved from http://www.ets.org/gre/news/more_mba_applicants
- Holt, D. T., Bleckmann, C. A., & Zitzmann, C. C. (2006). The Graduate Record Examination and success in an engineering management program: A case study. *Engineering Management Journal*, 18(1), 10–16.
- Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Greenwood.
- Klieger, D. M., Cline, F. A., Holtzman, S. L., Minsky, J. L., & Lorenz, F. (2014). *How much does the size of validity coefficients matter? Perspectives from the GRE General Test*. Manuscript submitted for publication.
- Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315, 1080–1081.
- 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.
- Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, 70, 340–352.
- Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.
- Milner, M., McNeil, J. S., & King, S. W. (1984). The GRE: A question of validity in predicting performance in professional schools of social work. *Educational and Psychological Measurement*, 44, 945–950.
- Powers, D. E. (2004). Validity of Graduate Record Examination (GRE) general test scores for admissions to colleges of veterinary medicine. *Journal of Applied Psychology*, 89, 209–219.

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.
- Sleeper, M. L. (1961). Relationship of scores on the Graduate Record Examination to grade point averages of graduate students in occupational therapy. *Educational and Psychological Measurement*, 21, 1039–1040.
- Snijders, T., & Bosker, R. (1994). Modeled variance in two-level models. *Sociological Methods & Research*, 22(3), 342–363.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, England: Sage.
- Stricker, G., & Huber, J. T. (1967). The Graduate Record Examination and undergraduate grades as predictors of success in graduate school. *The Journal of Educational Research*, 60, 466–468.
- Wang, W. (2013). *Testing the validity of GRE scores on predicting graduate performance for engineering students* (Unpublished master's thesis). University of Nebraska, Lincoln.

Suggested citation:

Young, J. W., Klieger, D., Bochenek, J., Li, C., & Cline, F. (2014). *The validity of scores from the GRE[®] revised General Test for forecasting performance in business schools: Phase one* (GRE Board Research Report No. 14-01; ETS Research Report No. RR-14-17). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12019

Action Editor: Brent Bridgeman

Reviewers: was reviewed by the GRE Technical Advisory Committee and the Research Committee and Diversity, Equity and Inclusion Committee of the GRE Board.

ETS, the ETS logo, GRE, and LISTENING. LEARNING. LEADING. are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS RESEARCHER database at <http://search.ets.org/researcher/>