



ORIGINAL

Testing the Validity of GRE Scores on Predicting Graduate Performance for National and International Engineering Students

Capacidad Predictiva de los Puntajes REG sobre el Rendimiento de Graduados Nacionales e Internacionales de Ingeniería

Wei Wang^a, and Kurt F. Geisinger^{1 a b}

^aDepartment of Educational Psychology, University of Nebraska-Lincoln

^bBuros Center for Testing

Received March 15, 2019; Accepted July 29, 2019

Abstract

Numerous research studies have examined the validity of GRE scores in predicting graduate success, however, some limitations still exist. This study targeted graduate engineering programs and investigated the validity of GRE scores in predicting graduate engineering GPA (GGPA). In addition, the differences in the validity of GRE scores between American and international students and between masters' and doctoral students were compared. The GRE's incremental predictive ability over undergraduate GPA (UGPA) and TOEFL scores was examined. Data were obtained from 1083 students from the engineering programs in a large, comprehensive midwestern university. Results indicated that GRE was useful in forecasting GGPA of graduate engineering students. The GRE scores explained more criterion variance for American students than for international students, but statistically significant differences were only found when GRE-Quantitative predicted GGPA. The GRE-Verbal and GRE-Quantitative scores had different patterns in predicting graduate grades for master's and doctoral students. UGPA was found to be a very strong predictor, and TOEFL scores were significantly correlated with the criterion variables. GRE scores, however, were found to have significant incremental validity over UGPA and TOEFL scores. TOEFL scores were less able to make predictions.

Keywords: Graduate Record Examinations, Predictive Evidence of Validity, Prediction of International Students, Language Proficiency

Resumen

Gran cantidad de estudios han examinado la validez de puntajes REG (Registro de Examinación de Graduados) como predictores del éxito de graduación, aunque aún existen algunas limitaciones. Este estudio se enfocó en programas de ingeniería y evaluó la validez de puntajes REG como predictor del GPA (*Grade Point Average*, Puntaje Promedio de Calificaciones) (GGPA). Adicionalmente, las diferencias en la validez de puntajes REG entre los estudiantes americanos e internacionales, así como los puntajes entre estudiantes de maestría y doctorado, fueron comparadas. La capacidad predictiva incremental del REG sobre el GPA de los estudiantes de pregrado (UGPA) y

¹ Contact: Kurt Geisinger, mail: Kgeisinger2@unl.edu

los puntajes TOEFL también fue evaluada. Se obtuvieron datos de 1083 estudiantes de programas de ingeniería de una universidad grande de la zona centro-oeste. Los resultados indicaron que el REG fue útil como predictor del GGPA de los estudiantes de ingeniería. Los puntajes REG explicaron mayor cantidad de varianza para estudiantes americanos que para estudiantes internacionales, aunque sólo se encontraron diferencias estadísticamente significativas cuando el REG cuantitativo predijo GGPA. Los puntajes REG verbal y REG cuantitativo presentaron puntajes diferentes al predecir calificaciones de los graduados de maestría y doctorado. El puntaje UPGA resultó ser un buen predictor, y los puntajes del TOEFL estuvieron correlacionados significativamente con las variables de criterio. Los puntajes REG, sin embargo, resultaron tener validez incremental significativa por encima de los puntajes UPGA y TOEFL. Los puntajes TOEFL fueron menos capaces de hacer predicciones.

Palabras Clave: Registros de Examinación de Graduados, Evidencia Predictiva de Validez, Predicción de Estudiantes Internacionales, Competencia Lingüística

The Graduate Record Examination (GRE), published by the Educational Testing Service (ETS), is a battery of standardized tests designed to determine the scholastic potential of graduate students and is commonly used in admission decisions by many universities and institutions. Because of the wide use of GRE scores in admission decisions and the critical importance of the predictive evidence of validity of the GRE, numerous studies about the GRE validation have been conducted since GRE was created and administered by ETS in 1949. Kuncel, Hezlett and Ones (2001) conducted a comprehensive meta-analysis of the predictive evidence of validity of the GRE. Their study integrated a variety of previous studies, including 1,753 independent samples and 82,659 graduate students, and considered the topic from different aspects including multiple disciplines, different criterion measures, and correction for statistical artifacts. It demonstrated that GRE scores were generalizably a valid predictor of graduate performance. However, as Kuncel et al. (2001) stated, there existed inconsistent results across studies and strong opinions of both sides on the usefulness of the GRE in predicting graduate performance. The limitations and gaps of previous studies are impetus for further research about the validity of GRE scores.

Issue of Language Proficiency

The relationship between GRE scores and graduate performance may be moderated by some variables. One potential moderator is language proficiency. The

GRE tests are focused on students' cognitive abilities but also reflect their language proficiency to some extent. As Sandoval and Durán (1998) stated, "When used with nonnative speakers, a test in English must be interpreted as measuring English proficiency in addition to the constructs it was designed to measure (p. 181)". Stricker (2004) found high correlations of TOEFL (the Test of English as a Foreign Language, a test that evaluates the English proficiency of people whose native language is not English) scores with GRE-V and GRE-A and moderate correlation with GRE-Q. A qualitative study by Mupinga and Mupinga (2005), by exploring the perceptions of international students toward the GRE, found that it was very hard for a test to measure all aspects of cognitive performance no matter how well it was developed, and that the content and context of the GRE test, especially the GRE verbal section, were considered to be biased against international students. Pennock-Roman (2002) found that many Puerto Rican students performed better on a Spanish language standardized test than on the GRE.

English language proficiency is a critical factor of the academic performance for non-native speaker students in a setting where English is used for teaching and learning. For this reason, besides GRE scores, colleges or universities in the United States usually request international students to reach a minimum level of language proficiency –usually determined by TOEFL scores –as a threshold of admission (Cho & Bridgeman, 2012; Wait & Gressel, 2009). Some studies have

showed that TOEFL scores played an important role on the academic performance, as indicated by factors such as GPA, pass rates on comprehensive assessment examinations, and graduation rate for international students (Cho & Bridgeman, 2012; Wait & Gressel, 2009). However, Cho and Bridgeman (2012) concluded that research findings on the power of TOEFL scores to predict academic success were mixed and inconsistent, so it was difficult to make a definitive conclusion about the validity of TOEFL. Thus, how language proficiency predicts graduate performance and how much more variance GRE explains beyond language proficiency are worthy to study.

Undergraduate Academic Performance in Graduate Admission

Undergraduate GPA (UGPA) is considered a good indicator of students' academic knowledge, abilities and performance during undergraduate studies. Applicants' GRE scores and UGPAs are the two most heavily weighted numerical and objective pieces of information in graduate admissions process (Kuncel et al., 2001; Norcross, Hanych, & Terranova, 1996). The combination of GRE scores and UGPA was a fairly robust predictor on multiple measures of performance, and the combination explained more variance in the dependent variable than either independent variable did by itself (McKee, Mallory, & Campbell, 2001; Power, 2004; Reisig & DeJong, 2005). Milner, McNeil and King (1984) found a significant increase in the minority enrollment rate –doubled from 9.85% to 17.56% –when GRE scores were eliminated from admission process but using only UGPA as the sole quantitative admission variable. However, in a very few cases, the power of UGPA in predicting graduate success was not obvious. For example, Smaby, Maddux, Richmond, Lepkowski, and Packman (2005) found that GRE scores and UGPA were of limited value when used to predict success for graduate students in counseling programs.

Specificity of Academic Disciplines in GRE Validity

The relationship of GRE scores and graduate performance may be dependent on the particular discipline

or academic background (House & Johnson, 1993; Stack & Kelley, 2002; Thornell & McCoy, 1985). Though there are similarities in some of the fundamental tasks required, the types of training, the demands, the grading standards and evaluation criteria differ from discipline to discipline. Thus, the predictive evidence of validity of the GRE for different disciplines needs consideration. Kuncel et al. (2001) categorized different disciplines into four different fields: social sciences, math-physical sciences, humanities, and life sciences. Their meta-analysis found that the predictive values of the GRE were inconsistent across disciplines and across test segments (i.e., GRE-Q and GRE-V). For example, the GRE-V accounted for more variance in graduate GPA (GGPA) in the social sciences than in the math-physical sciences, and the GRE-Q was less predictive of GGPA in the social sciences, life sciences, math-physical sciences than in the humanities. In addition, Stack and Kelly (2002) stated that GRE scores might be more predictive of GGPA in disciplines with low mean GRE scores than in disciplines with high mean GRE scores. House and Johnson (1993) also found that the relationships between predictor variables and degree completion varied by areas of graduate study or academic background.

Among various disciplines, engineering is one of the very important graduate programs in many universities. The importance of engineering graduate degrees is increasingly recognized by the professional engineering community (National Academy of Engineering, 2005; Rogers & Goktas, 2010). As a consequence of the awareness of this importance, the number of applicants to U.S. engineering graduate programs increased annually by an average of about 4% over the period of ten years from 1997 to 2007 (Bell, 2008). However, GRE validity studies have not widely extended to the discipline of engineering. Thus, the present study intended to help fill this gap and to target the population of engineering students to study the predictive validity of GRE scores in predicting graduate performance.

Purpose and Research Questions

Summarizing the studies reviewed, it was found that using GRE scores as one of the admission tests to

predict graduate performance has both a strong theoretical rationale and empirical support. However, the predictive evidence of validity of the GRE is inconsistent across studies, with the variance of graduate success explained by GRE ranging from less than 10% to as high as about 36% (Goldberg & Alliger, 1992; Fenster, Markus, Wiedemann, Brackett, & Fernandez, 2001; Morrison & Morrison, 1995; Powers, 2004; Sternberg & Williams, 1997). More research about this topic is needed to fill the gaps existing in previous studies. A test is considered to be biased if its predictive power is not equivalent for different subgroups (Johnson, Carter, Davison, & Oliver, 2001). In the present study, the difference in the validity of GRE scores between American students and international students in engineering was tested. Both the language issue and previous academic performance were taken into account to investigate the incremental validity of GRE scores over language proficiency and UGPA. In addition, the influence of degree level on the validity of GRE was also examined, considering that there are differences in program complexity and structure by degree level and that the effectiveness for predicting performance at both the master's and doctoral levels had gone unexamined until recent years (Kuncel, Wee, Serafin, & Hezlett, 2010).

Although GGPA has been shown to have limited value in reflecting how much students master the material and acquire the knowledge of the field of study, GGPA (especially the 1st-year GGPA and cumulative GGPA) is by far the most widely used criterion of graduate school performance (Fenster, Markus, Wiedemann, Brackett, & Fernandez, 2001; Kuncel, Crede, & Thomas, 2007; Kuncel et al., 2001). For most of the specific research studies, either 1st-year grades or cumulative grades were used as the measure of graduate success, but the performance of the second year was less frequently studied (Educational Testing Services, 2008a; Goldberg & Alliger, 1992; Perez, 2011; Sternberg & Williams, 1997). As is the 1st-year GPA and overall GPA, the 2nd-year GPA is also a reasonable criterion of graduate performance, one that was also included in the present study.

Specifically, the research questions of the present study were:

- 1) How do GRE scores predict engineering students' 1st-year, 2nd-year, and total GGPA in graduate program?
- 2) What are the differences in the validities of GRE scores between American students and international students?
- 3) What are the differences in the validities of GRE scores between masters' and doctoral students?
- 4) How does UGPA predict engineering students' 1st-year, 2nd-year and total GGPA in graduate program? What is the incremental validity of GRE over UGPA?
- 5) Specifically for international students, how does the language proficiency (measured by TOEFL) predict engineering students' 1st-year, 2nd-year and total GGPA in graduate program? What is the incremental validity of GRE over language proficiency?

Method

Sample

The sample for this study was from the various engineering programs in a large comprehensive midwestern university in the United States. Student records were obtained from the university's institutional research office. The data included all students who were enrolled in these engineering programs during the 11 academic years, from 2000 to 2011, with the number of these students totaling 1452. But only the students who had registered for classes for at least one and a half years or three academic semesters (in order to get the 1st-year and the 2nd-year GGPA) were retained for use in this study, which resulted in a total of 1096 students. Among the 1096 students, only 591 students had GRE scores, and 398 students had UGPA, 575 students had TOEFL scores. Thirteen students did not have scores for any of the three predictors and were hence excluded. Considering the existence of considerable missing or incomplete data for many students, and in order to keep as much information as possible, the students who had complete GPAs (1st-year, 2nd-year, and cumulative GGPA) and at least one predictor score (i.e., either GRE, TOEFL,

or UGPA) were retained. Finally, a total of 1083 students ($N=1083$, 79.8% were male and 20.2% were female) were available in the final database. Amongst these students, 39.6% of the students were originally from the United States (41.6% were non-alien), 24.6% were from China, 14.3% were from India, and the rest (21.5%) were from other 62 countries around the world. Among the international students in this sample, there were rarely students from such English-speaking countries as the United Kingdom and Australia.

Measures

Graduate Student Performance. Three measures were used as the indicators of graduate performance: the 1st-year GGPA, the 2nd-year GGPA, and the total GGPA (or cumulative GGPA). GGPA was evaluated on a 0-4.0 continuous scale. The 1st-year GGPA and the 2nd-year GGPA both covered one academic year. The total GGPA was different and it covered the period of time from the initial enrollment in graduate study until the end they graduated or until the last semester the data covered in the database (i.e., 2012 Fall). Considering the number of classes that students registered in each year differed individually, the corresponding credit hours for each student in each period of time were also kept.

UGPA. The UGPA is a cumulative grade-point average covering all the undergraduate coursework. It was also on a 0-4.0 scale. However, in this study, this measure was only available for 398 students who studied and got their bachelor's degrees from the targeted midwestern university. For other students (i.e., $n = 685$) who graduated from other universities or colleges, their UGPAs were not retained in the university system.

Graduate Record Examination. The GRE used in this study has three subtests that measure verbal reasoning (GRE-V), quantitative reasoning (GRE-Q), and analytical writing skills (GRE-A) (Educational Testing Services, 2013). At time of this study, the GRE-V and the GRE-Q had a possible score range from a minimum of 200 to a maximum of 800. The GRE-Total was the sum of the GRE-V and GRE-Q scores. In this study, only GRE-V, GRE-Q, and GRE

total scores were used in the analysis because the Analytic writing scores were not available for all students.

TOEFL. The TOEFL is a test to evaluate the English proficiency of people whose native language is not English. Since 2006, the Internet-based version of the TOEFL test (TOEFL iBT) had been phased in worldwide. Before 2006, TOEFL tests had two versions: paper-based (PBT) and computer-based (CBT) (Alderson, 2009). In this study, 53.1% ($n = 575$) of the total students had TOEFL scores but in different versions. Because the three versions have different scoring scales and in order to integrate the test scores, these three versions of scores were placed on a single scale according to the TOEFL Score Comparison Tables (Educational Testing Services, 2005). In this study, TOEFL iBT scores and PBT scores were transformed into scores using the scale of CBT, which has a scoring range from 0 to 300. For those students who took the TOEFL more than one time (only five students in this study), the highest score was used as his or her TOEFL score, as that was the admissions practice of this university.

Demographics. The term alien status in this study represented whether or not the students were the residents of the United States. In terms of the degree level, students who enrolled as a doctoral students or originally as master's students and then continued to the doctoral programs of the same university were noted as doctoral students. Students who enrolled as master's students and did not continue to the doctoral programs of the same university were noted as master's students. Table 1 shows some detailed demographic information of the sample.

Table 1
Frequency and Percentage of Master's and Doctoral Students by Alien Status

Alien Status	Master's Students		Doctoral Students		Total
	N	%	N	%	
US	370	55.7%	81	19.3%	451
Alien	294	44.3%	338	80.7%	632
Total	664	100%	419	100%	1083

Data Analysis

The data analysis for this study consisted of both descriptive and inferential statistics. To test the validity of GRE scores, 1st-year GGPA, 2nd-year GGPA, and total GGPA were regressed separately on GRE-V, GRE-Q, and GRE total scores. Hierarchical multiple regression analyses were conducted to examine the incremental validity of GRE scores over UGPA and TOEFL scores. To solve the problem of missing data, the method of Listwise deletion was applied.

Results

Table 2 shows the descriptive statistics for both predictor variables and criterion variables. The means of the three criteria, 1st-year GGPA ($M = 3.64$, $SD = .32$), 2nd-year GGPA ($M = 3.68$, $SD = .33$), and total GGPA ($M = 3.68$, $SD = .27$), were quite similar. The corresponding credit hours in each period of time, however, were different. The 1st-year mean hours were 20.24 ($SD = 6.21$), the 2nd-year hours were 17.14 ($SD = 7.64$), and the total mean hours were 50.21 ($SD = 27.24$). The GRE-V had a lower mean score, larger standard deviation, and larger range ($M = 435.25$, $SD = 119.42$, 200-800 range) than GRE-Q ($M = 731.74$, $SD = 73.32$, 320-800 range). The mean of UGPA was 3.41 ($SD = .36$) and the mean of TOEFL total scores was 236.14 ($SD = 27.30$).

Pearson correlations were computed among all variables. All correlations were statistically significant at the alpha level of either .01 or .05. The correlation between 1st-year and 2nd-year GGPA was .56,

and they had a spurious correlation with the total GGPA at .82 and .81, respectively. GRE-V correlated with GRE-Q at .34, and these two subtest scores had a spurious correlation with GRE total scores at .90 and .71, respectively. The correlations between the predictor variables (i.e., GRE and TOEFL scores) and the criterion variables ranged from .17 to .28. In contrast, the correlations between UGPA and the three criterion variables ranged from .52 to .64. GRE scores had correlations with UGPA, ranging from .37 to .51, and with TOEFL, ranging from .23 to .54.

Difference of Descriptive Statistics by Alien Status and by Degree Level

In order to see the differences between two groups (American students vs. international students; master's students vs. doctoral students), descriptive statistics of each group were given and independent sample *t* tests were conducted, as shown in Table 3. There were no significant differences between American students and international students in graduate performance as measured by 1st-year, 2nd-year, and total GGPA at the alpha level of 0.05. American students had higher GRE total scores than international students, but this difference was not statistically significant. For the subtests, American students had significantly higher GRE-V scores but lower GRE-Q scores than international students. Doctoral students had significantly higher scores than master's students in 1st-year GGPA, 2nd-year GGPA and total GGPA, also in GRE-Q scores and GRE total scores, and in UGPA. The differences in GRE-V and TOEFL were not found to be significant.

Table 2
Descriptive Statistics of the Predictors and the Criteria

Variables		N	M	SD	Minimum	Maximum
Criteria	1.GGPA-1	1083	3.64	.32	2.00	4.0
	2.GGPA-2	1083	3.68	.33	1.67	4.0
	3.GGPA-tot	1083	3.68	.27	2.44	4.0
Predictors	GRE-V	591	435.25	119.42	200	800
	GRE-Q	591	731.74	73.32	320	800
	GRE-tot	591	1166.99	159.95	650	1600
	UGPA	398	3.41	.36	2.49	4.0
	TOEFL	575	236.14	27.30	130	293

Table 3
Difference of Descriptive Statistics of Variables by Alien Status and by Degree Level

Variables	Alien status	N	M (SD)	t	Degree	N	M (SD)	t	
Criteria	GGPA-1	Alien	632	3.65 (.32)	.461	MS	664	3.58(.35)	9.78**
		US	451	3.64 (.33)		PHD	419	3.76(.24)	
	GGPA-2	Alien	632	3.69 (.31)	1.54	MS	664	3.63(.35)	6.14**
		US	451	3.66 (.34)		PHD	419	3.75(.27)	
	GGPA-tot	Alien	632	3.70 (.26)	1.61	MS	664	3.62(.28)	10.55**
		US	451	3.67 (.28)		PHD	419	3.78(.20)	
Predictors	GRE-V	Alien	461	421.52(121.97)	6.17**	MS	292	429.35(117.44)	1.19
		US	130	483.92 (95.47)		PHD	299	441.00(121.24)	
	GRE-Q	Alien	461	741.32 (66.06)	5.31**	MS	292	717.95(79.13)	4.59**
		US	130	697.77(86.78)		PHD	299	745.22(64.49)	
	GRE-tot	Alien	461	1162.84(159.48)	1.19	MS	292	1147.29(161.94)	2.98**
		US	130	1181.69(161.36)		PHD	299	1186.22(155.86)	
	UGPA					MS	342	3.39(.36)	3.88**
						PHD	56	3.59(.31)	
	TOEFL					MS	278	235.16(29.76)	.83
						PHD	297	237.05(24.80)	

* $p < .05$, ** $p < .01$

Validity of GRE Scores on Predicting Graduate Performance

As GRE-V, GRE-Q, and GRE total scores were considered separately, three simple linear regressions (SLR) were conducted on all criterion variables. As shown in Table 4, all standardized regression coefficients were significant, indicating the usefulness of the GRE-V, GRE-Q, and GRE total scores in predicting 1st-year, 2nd-year, and total GGPA. The variances in criterion variables that were explained by GRE scores, indicated by R^2 , ranged from 2.8% to 7.9%. Across the three criterion variables, more variance in total GGPA was explained by GRE scores (ranging from 4.6% to 7.9%) than that in 1st-year and 2nd-year GGPA (ranging from 2.8% to 5.2%).

Differences in the Validity of GRE Scores by Alien Status and Degree Level

Regression of graduate performance on GRE scores was conducted separately for American students and international students. As shown in Table 5, all the standardized coefficients were significant, which indicated the usefulness of GRE scores in predicting

Table 4
The Index (Standardized Regression Coefficient and R Square) of the Regression of GGPA on GRE scores

	GGPA-1		GGPA-2		GGPA-tot	
	β	R^2	β	R^2	β	R^2
GRE-V	.167**	.028	.185**	.034	.214**	.046
GRE-Q	.224**	.050	.180**	.032	.263**	.069
GRE-tot	.227**	.052	.221**	.049	.281**	.079

$p < .05$, ** $p < .01$.

graduate performance for both American students and international students. For American students, the variance in across the three criterion variables explained by GRE scores ranged from 10.5% to 22.2%. By contrast, for international students, the variance explained by GRE scores was much smaller, only ranged from 1.4% to 5.5%. Although the difference in the validity by Alien Status appeared to be large, however, in testing the interactions between Alien Status and GRE scores, the interactions were not found to be significant except the one between Alien Status and GRE-Q in predicting GGPA_tot ($F(23,$

531) = 1.66, $p < .05$). Namely, the statistically significant difference in the prediction between American students and international students was found only when GRE-Q was used in predicting GGPA total scores.

The same method was used to test the validity of GRE for masters' and doctoral students. As shown in Table 6, except the prediction by GRE-V on the 2nd-year GGPA and total GGPA for doctoral students, all other regressions were significant at either alpha level of .05 or .01. The differential validity by degree level differed among the three GRE scores. GRE-V explained relatively more variance for master's students (3.5%, 7.4%, 8.7%, respectively) than for doctoral students (1.7%, 0.7%, 1.2%, respectively) in 1st-year, 2nd-year, and total GGPA. GRE-Q explained relatively larger variance for doctoral students (6.3%, 3.6%, 8.3%, respectively) than for master's students (1.9%, 1.7%, 3.1%, respectively) in the three criteria. GRE total scores explained more variance for master's students ($R^2 = 6.8\%$ and 9.0% , respectively) than doctoral students ($R^2 = 2\%$ and 4.2% , respectively) in 2nd-year GGPA and total GGPA, but no difference in 1st-year GGPA. However, further statistical tests with both degree level and GRE scores in the regression model showed no significant interactions between degree and GRE scores except two: one interaction between degree and GRE-V in predicting 1st-year GGPA ($F(47, 487) = 1.46, p < .05$), and another one between degree and GRE-Q in predicting 2nd-year GGPA ($F(26, 528) = 1.57, p < .05$).

Incremental Validity of GRE over UGPA and TOEFL

The incremental predictive ability was analyzed by using hierarchical multiple regression. This method was used first to test the incremental validity of GRE over UGPA. Because of the reason that only a small portion of the students who had UGPA had GRE scores (65 out of 398), the predictive ability of UGPA was tested by simple linear regression (SLR) prior to MR so as to include all the 398 students. As shown in Table 7, the results of the SLR ($n = 398$) indicated that UGPA explained 40.8%, 27.4%, and 40.3% of variance (indicated by R^2) in 1st-year GGPA, 2nd-year

Table 5

The Index (Standardized Coefficient and R Square) of the Regression of GGPA on GRE scores by Alien Status

		GGPA-1		GGPA-2		GGPA-tot	
		β	R^2	β	R^2	β	R^2
GRE-V	Alien	.118*	.014	.154**	.024	.172**	.030
	US	.325**	.106	.324**	.105	.362**	.131
GRE-Q	Alien	.211**	.044	.144**	.021	.233**	.055
	US	.430**	.185	.369**	.136	.471**	.222
GRE-tot	Alien	.178**	.032	.177**	.031	.228**	.052
	US	.423**	.179	.390**	.152	.467**	.218

* $p < .05$, ** $p < .01$.

Table 6

The Index (Standardized Coefficient and R Square) of the Regression of GGPA on GRE scores by Degree Level

		GGPA-1		GGPA-2		GGPA-tot	
		β	R^2	β	R^2	β	R^2
GRE-V	MS	.187**	.035	.272**	.074	.294**	.087
	PhD	.131*	.017	.081	.007	.111	.012
GRE-Q	MS	.139*	.019	.131*	.017	.176**	.031
	PhD	.251**	.063	.189**	.036	.287**	.083
GRE-tot	MS	.204**	.041	.261**	.068	.299**	.090
	PhD	.206**	.042	.141*	.020	.205**	.042

* $p < .05$, ** $p < .01$.

GGPA, and total GGPA, respectively. In the hierarchical multiple regression with a smaller sample size ($n = 65$), UGPA alone explained 36.7%, 26.9%, and 40.3% of variance in 1st-year, 2nd-year, and total GGPA, respectively. After adding GRE scores into the model, the proportions of variance in three criterion variables increased significantly, indicated by R^2 and R^2 change, which indicated that GRE scores explained a significant additional proportion of variance over/beyond what UGPA explained. This incremental validity was presented more obviously for 1st-year GGPA and 2nd-year GGPA than total GGPA. Specifically, GRE scores explained additional 5% to 10% of variance beyond UGPA for 1st-year GGPA, additional 5.2% to 8.4% for 2nd-year GGPA, and additional 2.8% to 4.7% for total GGPA.

Table 7
Hierarchical Multiple Regression of GGPA on UGPA and GRE scores

		GGPA-1			GGPA-2			GGPA-tot			
		β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2	
SLR ($n=398$)	UGPA	.639**	.408		.524**	.274		.635**	.403		
MR ($n=65$)											
Model	1	UGPA	.605**	.367		.519**	.269		.635**	.403	
	2a	GRE-V	.240*	.416	.050*	.270*	.332	.063*	.180	.431	.028
	2b	GRE-Q	.359**	.462	.096**	.264*	.321	.052*	.227*	.441	.038*
	2c	GRE-tot	.365**	.466	.100**	.335**	.353	.084**	.251*	.450	.047*

Note. SLR = simple linear regression, MR = multiple regression.

* $p < .05$, ** $p < .01$.

The incremental validity of GRE scores over TOEFL scores was also examined. As shown in Table 8, the results of SLR ($n = 575$) showed that TOEFL scores explained 2.2%, 3%, and 4.9% of variance in 1st-year GGPA, 2nd-year GGPA, and total GGPA, respectively. However, after using the method of Listwise deletion in the hierarchical multiple regression ($n = 401$), TOEFL scores did not explain significant proportions of variance. After adding GRE scores into the model, the proportions of explained variance increased significantly for all the three criterion variables, which indicated that GRE scores explained a significant additional proportion of variance over/beyond TOEFL scores. Specifically, GRE scores explained additional 1.7% to 4.6% of variance beyond TOEFL scores for 1st-year GGPA, additional 1.4% to 2.7% for 2nd-year GGPA, and additional 2.7% to

5.3% for total GGPA. In addition, GRE-V explained less additional variance (at most 2.7%) than GRE-Q (at most 5.2%) over TOEFL.

Discussion

The purpose of this study was to investigate the predictive evidence of validity of GRE scores in predicting graduate performance for engineering students. The predictions for different demographic groups (American students vs. international students, and masters' vs. doctoral students) were compared. The language issue and undergraduate performance were taken into account to study their contributions to the graduate performance, as well as to examine the incremental validity of GRE scores beyond language proficiency and UGPA.

Table 8
Hierarchical Multiple Regression of GGPA on TOEFL and GRE scores

		GGPA-1			GGPA-2			GGPA-tot			
		β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2	
SLR($n=575$)	TOEFL	.149**	.022		.173**	.030		.221**	.049		
MR ($n=401$)											
Mode	1	TOEFL	.052	.003		.057	.003		.093	.009	
	2a	GRE-V	.154**	.019	.017**	.171**	.024	.021**	.197**	.036	.027**
	2b	GRE-Q	.220**	.049	.046**	.123*	.018	.014*	.234**	.060	.052**
	2c	GRE-tot	.230**	.042	.039**	.191**	.030	.027**	.268**	.062	.053**

Note. SLR = simple linear regression, MR = multiple regression.

* $p < .05$, ** $p < .01$.

In this study, no significant differences were found between international students and American students in graduate performance, as measured by 1st-year GGPA, 2nd-year GGPA, and total GGPA. American students had higher GRE-V and GRE total scores but lower GRE-Q scores than international students. This result, to some extent, corresponded to the findings of ETS (Educational Testing Services, 2008) that minority students usually received significantly lower GRE scores than White students, with the exception that Asian students usually got higher score on the GRE-Q section.

This study started from a comprehensive perspective and examined both short-term (as measured by 1st-year GGPA, as well as 2nd-year GGPA) and long-term performance (as measured by total GGPA). The result of testing the general predictive ability of GRE scores indicated that GRE was a valid predictor in predicting all three criterion variables. Across the three criterion variables, relatively more variance in total GGPA was explained than that in 1st-year and 2nd-year GGPA. One possible explanation could be that total GGPA (which covered more years of grades) is more reliable than the one-year GGPA. Moreover, the subtests of GRE differed in the prediction (GRE-Q and GRE total scores were found to have higher predictive ability than GRE-V), which suggested the right choice to consider GRE-V and GRE-Q separately in the analysis. Kuncel et al. (2001) also found GRE-Q had higher predictive ability than GRE-V (i.e., operational validity coefficients = .31, and .26, respectively) for students in STEM fields in which engineering students were included. In the discipline of engineering, students are usually required to have high abilities in Math, statistics, numerical logic, and some other advanced quantitative skills. Not surprising, the quantitative abilities appear more important than verbal abilities for engineering students. From this logic, GRE-Q would be likely to have a higher correlation with graduate performance than GRE-V for engineering students.

In testing the differences in the prediction of GRE scores between different groups, this study found that in general GRE scores explained more variance in graduate performance for American students

(explained 10.5% to 22.2% variance) than for international students (explained 1.4% to 5.5% of variance), although statistical tests did not find all significant differences. There were very few studies in literature that examined the GRE validity for international students, or differentiated GRE validity by alien status. Thus, the findings of the present study were not comparable to many previous studies.

Regarding degree level, GRE scores significantly predicted the three criterion variables for both masters' and doctoral students. For masters' students, GRE-V and GRE total explained larger variance than GRE-Q scores across 1st-year, 2nd-year, and total GGPA. For doctoral students, by contrast, GRE-Q and GRE total scores better predicted the three criteria than GRE-V. This finding was consistent to the results in Kuncel et al. (2010) meta-analysis which found that for master's students, GRE-V had a slight larger operational validity than GRE-Q ($\rho = .38$ and $.35$ for GRE-V, $\rho = .30$ and $.28$ for GRE-Q in final GGPA and 1st-year GGPA, respectively); and for doctoral students, GRE-Q had a slight larger operational validity than GRE-V ($\rho = .28$ and $.33$ for GRE-Q, and $\rho = .27$ and $.29$ for GRE-V in final GGPA and 1st-year GGPA, respectively). In the present study, although the values of R^2 in regression models were different by degree level, results did not indicate statistically significant differences. As reported above, doctoral students had higher GGPA scores and higher GRE scores than masters' students, which made the distribution of the scores of doctoral students tend to be in the higher end. This range restriction may make it hard to test the difference even if the difference existed.

In terms of the contributions of undergraduate performance in predicting graduate performance, the UGPA explained 40.8%, 27.4%, and 40.3% of the variance in 1st-year, 2nd-year, and total GGPA, respectively. This result was consistent with some previous findings that also indicated UGPA was a strong predictor of graduate performance (Kuncel et al., 2001; McKee, Mallory, & Campbell, 2001; Powers, 2004; Reisig & DeJong, 2005). GRE scores were found to have significant incremental validity over UGPA, increasing the explained variance by 3.8% to 10%. In

addition, corresponding to some previous studies, this study also found that the amount of variance in graduate performance that explained by GRE alone was less than that by UGPA though GRE explained some additional variance that UGPA did not explain (Milner, McNeil & King, 1984; Reisig & DeJong, 2005).

The GRE scores also had a significant incremental predictive power over TOEFL in predicting graduate performance, which confirmed that GRE was more than a test of language skills. The GRE explained a significant additional amount of variance that language proficiency test (as measured by TOEFL) could not explain. As to the two GRE subtests, GRE-V had lower incremental ability than GRE-Q subsequent to the use of TOEFL. This difference can be explained by the specific purpose and content of each subtest. GRE-V is more related to reasoning ability and language skills, while GRE-Q is more related to quantitatively problem-solving ability (e.g., arithmetic, algebra, geometry and data analysis), thus the GRE-V was found to add less power in predicting graduate performance over TOEFL compared to the GRE-Q.

Limitations and Implications of this Study

There are various limitations in this study. The first one is both a limitation and an advantage. On one hand, this study only targeted students from engineering programs, so the findings and implications can only be directly applied to this specific discipline. Generalization of the findings of this study to other disciplines should be with caution. On the other hand, focusing on the discipline of engineering can draw accurate conclusions and implications for this specific discipline. The second limitation concerns the missing data in this study. In the database in university's system, many students did not have complete data on all variables and the reasons for these deficiencies were unknown, so the representativeness of the data could not be tested. Thus, the results should be interpreted with some caution. Another limitation concerns the restriction of range of scores and especially, graduate grades, as is so often the case. Correction for range restriction was recommended because it was found that the corrected correlation between

predictor variables and criterion variables improved (Chernyshenko & Ones, 1999; Kuncel et al., 2001; Powers, 2004). As Kuncel et al. (2001) mentioned, to correct for the restriction of range, the definition of the interested population is critical, and the standard deviations of both sample and population should be known. However, this information was not available for this study, so the correction of range restriction was not possible, which may make it hard to determine the exact relationships between variables.

Above all, this study confirmed the usefulness of the GRE in predicting graduate performance, as measured by 1st-year, 2nd-year, and total GGPA, for engineering students. The test of the incremental validity suggested combining GRE with UGPA and TOEFL in consideration in graduate admission decisions. Although the differences in the validity by alien status and degree level were not found to be statistically significant in all cases, the absolute values of the explained variance, to some extent, showed some difference. In order to use GRE scores more wisely in graduate admission, this possible difference should be taken into consideration. In addition, a great amount of variance in the criterion variables remains unexplained by the predictors in this study, so there may be room and a need to conduct more research to study the unexplained portion of variance in graduate performance. As suggested by Kuncel et al. (2010), multiple aspects of student performance should be considered to have a more comprehensive picture about students' performance. The criteria may include such information as faculty ratings, degree attainment, degree completion, and research productivity. Moreover, as to the factors that may impact the graduate performance, not only cognitive abilities should be considered, but also some noncognitive characteristics of the applications, such as motivation, interest, personality, and some other characteristics.

References

1. Alderson, J. C. (2009). Test review: Test of English as a Foreign Language: Internet-based Test (TOEFL iBT). *Language Testing*, 26, 621–631.
2. Bell, N.E. (2008). *Graduate enrollment and degrees: 1997 to 2007*. Washington, DC: Council of Graduate Schools.

3. Chernyshenko, O. S., & Ones, D. S. (1999). How selective are psychology graduate programs? The effect of the selection ratio on GRE score validity. *Educational and Psychological Measurement*, 59, 951-961.
4. Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, 29, 421-442.
5. Educational Testing Services. (2005). TOEFL Interest-based Test: Score Comparison Tables. Online version available at http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_iBT_Score_Comparison_Tables.pdf.
6. Educational Testing Services. (2008). Graduate Record Examination: Guide to the use of GRE scores 2008-2009.
7. Educational Testing Services. (2013). Graduate Record Examination: Guide to the use of GRE scores 2013-2014.
8. Fenster, A., Markus, K. A., Wiedemann, C. F., Brackett, M. A., & Fernandez, J. (2001). Selecting tomorrow's Forensic Psychologists: a fresh look at some familiar predictors. *Educational and Psychological Measurement*, 61, 336-348.
9. 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.
10. House, J. D., Johnson, J. J. (1993). Graduate record examination scores and academic background variables as predictors of graduate degree completion. *Educational and Psychological Measurement*, 53, 551-556.
11. Johnson, J. W., Carter, G. W., Davison, H. K., & Oliver, D. H. (2001). A synthetic validity approach to testing differential prediction hypotheses. *Journal of Applied Psychology*, 86, 774-780.
12. Kuncel, N. R., Crede, M., & Thomas, L. L. (2007). A meta-analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and undergraduate grade point average (UGPA) for graduate student academic performance. *Academy of Management Learning & Education*, 6, 51-68.
13. 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.
14. 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.
15. McKee, A. J., Mallory, S. L., & Campbell, J. (2001). The Graduate Record Examination and undergraduate grade point average: Predicting graduate grade point averages in a criminal justice graduate program. *Journal of Criminal Justice Education*, 12, 311-317.
16. 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.
17. Morrison, T., & Morrison, M. (1995). A meta-analysis 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, 309-316.
18. Mupinga, E. E., & Mupinga, D. M. (2005). Perceptions of international students toward GRE. *College Student Journal*, 39, 402-408.
19. National Academy of Engineering. (2005). *Educating the engineer of 2020: Adapting engineering education to the new century*. Washington, DC: The National Academies Press.
20. Norcross, J. C., Hanych, J. M., & Terranova, R. D. (1996). Graduate study in psychology: 1992-1993. *American Psychologist*, 51, 631-643.
21. Pennock-Roman, M. (2002). Relative effects of English proficiency on general admissions tests versus subject tests. *Research in Higher Education*, 43, 601-623.
22. Perez, K. (2011). *GRE as a predictor of graduate student success at a Hispanic servicing institution of higher education*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. UMI 3472034)
23. Powers, D. E. (2004). Validity of Graduate Record Examinations (GRE) General Test scores for admissions to Colleges of Veterinary Medicine. *Journal of Applied Psychology*, 89, 208-219.
24. Reisig, M. D., & DeJong, C. (2005). Using GRE scores and prior GPA to predict academic performance among criminal justice graduate students. *Journal of Criminal Justice Education*, 16, 37-59.
25. Rogers, S. W., & Goktas, R. K. (2010). Exploring engineering graduate student research proficiency with student surveys. *Journal of Engineering Education*, 99, 263-278.
26. Sandoval, J., & Durán, R. (1998). *Language*. In Sandoval, J., Frisby, C. L., Geisinger, K. F., Scheuneman, J., & Grenier, J. R. (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (181-211). Washington, DC: APA.

27. Smaby, M. H., Maddux, C. D., Richmond, A. S., Lepkowski, W. J., & Packman, J. (2005). Academic admission requirements as predictors of counseling knowledge, personal development, and counseling skills. *Counselor Education & Supervision, 45*, 43-57.
28. Stack, S. & Kelley, T. (2002). The graduate record examination as a predictor of graduate student performance: The case of criminal justice. *Journal of Criminal Justice Education, 13*, 335-349.
29. Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? *American Psychologist, 52*, 630.
30. Stricker, L. J. (2004). The performance of native speakers of English and ESL speakers on the computer-based TOEFL and GRE General Test. *Language Testing, 21*, 146-173.
31. Thornell, J. G., & McCoy, A. (1985). The predictive validity of the Graduate Record Examination for subgroups of students in different academic disciplines. *Educational and Psychological Measurement, 45*, 415-419.
32. Wait, I. W., & Gressel, J. W. (2009). Relationship between TOEFL score and academic success for international engineering students. *Journal of Engineering Education, 98*, 389-398.