

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Public Access Theses and Dissertations from
the College of Education and Human Sciences

Education and Human Sciences, College of
(CEHS)

Fall 11-2013

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

Wei Wang

University of Nebraska-Lincoln, wangwei.apply@gmail.com

Follow this and additional works at: <https://digitalcommons.unl.edu/cehsdiss>



Part of the [Educational Psychology Commons](#)

Wang, Wei, "Testing the Validity of GRE Scores on Predicting Graduate Performance for Engineering Students" (2013). *Public Access Theses and Dissertations from the College of Education and Human Sciences*. 192.

<https://digitalcommons.unl.edu/cehsdiss/192>

This Article is brought to you for free and open access by the Education and Human Sciences, College of (CEHS) at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Public Access Theses and Dissertations from the College of Education and Human Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

TESTING THE VALIDITY OF GRE SCORES ON PREDICTING GRADUATE
PERFORMANCE FOR ENGINEERING STUDENTS

by

Wei Wang

A THESIS

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Master of Arts

Major: Educational Psychology

Under the Supervision of Professor Kurt Geisinger

Lincoln, Nebraska

November, 2013

TESTING THE VALIDITY OF GRE SCORES ON PREDICTING GRADUATE PERFORMANCE FOR ENGINEERING STUDENTS

Wei Wang, M.A.

University of Nebraska, 2013

Advisor: Kurt F. Geisinger

The Graduate Record Examination (GRE), a set of standardized tests designed to determine the scholastic potential of graduate students, is widely used in graduate admissions in the United States. How GRE can predict graduate students' performance has crucial importance both for universities and for students. Numerous of research studies have examined the validity of GRE scores in predicting graduate success, however, some limitations and gaps still existed in previous studies. This study targeted a specific discipline of engineering, and investigated the validity of GRE scores in predicting graduate performance, as measured by graduate GPA (GGPA) for engineering students. The differences in the validity of GRE scores between American and international students and between master's and doctoral students were tested. The incremental predictive abilities of GRE over undergraduate GPA (UGPA) and TOEFL scores were also examined. Data of this study were obtained from 1083 students from the engineering programs in a large comprehensive midwestern university. Results of this study indicated that GRE was a useful predictor in predicting 1st-year, 2nd-year, and total GGPA of engineering students. The GRE-Verbal (GRE-V) and GRE-Quantitative (GRE-Q) scores had a different pattern in predicting graduate grades for master's and doctoral students. The GRE-V and GRE-Q scores explained more variance in graduate

performance for American students than for international students, but no statistically significant differences were found except when GRE-Q predicted GGPA total scores. UGPA was found to be a strong predictor, and TOEFL scores were also significantly correlated with the criterion variables. GRE scores were found to have significant incremental validity over UGPA and TOEFL scores in predicting graduate grades. These findings have implications for graduate admission decisions for engineering programs, and can suggest directions of future research, which were also discussed in this study.

Key Words: Graduate Record Examinations, Predictive Evidence of Validity, Graduate Performance, Language Proficiency

Copyright 2013, Wei Wang

ACKNOWLEDGEMENTS

It would not have been possible to complete this thesis without the help and support of the people around me. It is a pleasure to convey my gratitude to them all in my humble acknowledgement.

My first and sincere appreciation goes to my advisor, Dr. Kurt Geisinger for his continuous help and support in all stages of this thesis. This thesis would not have been possible without his encouragement and supervision. I would also like to thank him for his patience in guiding me academic writing, and for his attitude to research and his conscientiousness in supervision and teaching.

I gratefully acknowledge Dr. Anthony Albano for his advice, supervision, and review of the thesis. His thorough review of my thesis and his comments let me have a chance to change and supplement my work to be better.

In addition, I would like to give my great appreciation to the Institutional Research and Planning (IRP) for providing me the datasets that I requested for this thesis. Specifically, I would like to thank Mary Werner, the associate director in IRP, for her dedicated work in digging out the datasets from the vast database time after time.

I would like to acknowledge the Department of Educational Psychology in UNL where I learn much knowledge and skills which made the thesis possible. I also want to thank the faculties and staffs in the department for their assistance. Moreover, I would also like to thank my dear friends, to name just a few here like Yuchun Zhou, Chaorong Wu, Shanshan Deng, and Yaodong Yang, for their help and suggestions either technically or personally.

At last but not the least, I want to thank my dearest family, my father, my mother, and my brother for always believing in me and for their unconditional love and continuous supports. I would also like to express my deep gratitude to Y.M. Tang. With the encouragement and support from all the dear people, I have been learning to grow up and be brave in my life.

Contents

CHAPTER I.....	1
INTRODUCTION AND LITERATURE REVIEW	1
Literature Review	3
Theoretical Rationale of the Predictive Evidence of Validity of GRE.....	3
Defining the Criterion: Graduate Performance	4
How to Test the Predictive Validity	6
Specificity of Academic Disciplines in GRE Validity	7
Issue of Language Proficiency: American and International Students.....	12
Undergraduate Academic Performance in Graduate Admission.....	16
Difference in Validity by Degree Level	19
Role of Range Restriction in GRE Validation Research.....	19
Purpose Statement and Research Questions.....	22
Significance of the Study	24
CHAPTER II.....	26
RESEARCH METHODS	26
Sample.....	26
Procedures	26
Measures.....	28
Data Analysis	30
CHAPTER III	33
RESULTS	33
Descriptive Statistics	33
Difference of Descriptive Statistics of Variables by Alien Status and by Degree Level	36
Validity of GRE Scores on Predicting Graduate Performance	39

Differences in the Validity of GRE Scores by Alien Status and Degree Level	40
Incremental Validity of GRE over UGPA and TOEFL	43
CHAPTER IV	47
DISCUSSION AND CONCLUSIONS	47
Findings of the Study	48
Limitations of this Study	59
Implications and Recommendations	61
Conclusions	62
REFERENCES	64
APPENDIX.....	71

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Effective selection of graduate students is of critical importance for graduate programs. Poor decisions can result in large costs, such as inefficient use of resources in education, an overall weakening of the profession and academic quality for universities, and even time-consuming and financial burdens for students. In order to identify those applicants who best-fit the programs and who will excel in and enrich the field of study, admission committees consider multiple pieces of information about the applicants through the admission process. Standardized test scores and undergraduate grade-point averages (UGPA) are the objective information about applicants, which serve as primary screening devices. Some subjective information, including personal statements, writing samples, letters of recommendation and interviews, are also considered for admission. Many programs and universities post the requirements, such as average or explicit cutoff or minimum scores of admission tests in order to guide applicants in deciding whether the programs are appropriate to make applications.

The Graduate Record Examination (GRE), published by the Educational Testing Service (ETS), is a set of standardized tests designed to determine the scholastic potential of graduate students and is commonly used in admission decisions by many universities and institutions. According to Norcross, Hanych, and Terranova's (1996) review of the admission information of graduate psychology programs, applicants' GRE scores and UGPA are the two most heavily weighted numerical or objective pieces of information in graduate admission process. Based on the information from 458 institutions, 559 separate departments, and 2,023 individual graduate programs in psychology, the Norcross,

Hanych and Terranova review reported that 93% of doctoral departments and 81% of the master's departments required GRE general test scores. In addition, the Educational Testing Service (Graduate Records Examination Board, 2003) encourages all schools and departments to use GRE as a meaningful source of information to screen applicants or select fellowship awardees.

Having an opportunity to gain the access to graduate education is also very important for students because graduate degrees play an essential role in seeking for good employment opportunities and greater lifetime earnings. A first question is, can GRE scores accurately provide a good source of information for graduate admission committees to decide whether a student should be admitted or not? A related question is, can GRE scores really predict the success of graduate students? Whether to use GRE scores to screen graduate students for admission into graduate programs is a long-standing controversy. Proponents have argued that the GRE can predict graduate success well. Critics argue that the GRE may underpredict academic performance of marginalized groups, such as racial or ethnic minorities, women, and older graduate students, thus limiting their access and choice to graduate education (Stricker & Rock, 1995). 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 of the GRE validation have been conducted since GRE was created and administered by ETS in 1949. These studies have helped people to have more understanding of the predictive ability of the GRE. However, as stated in Kuncel, Hezlett and Ones' (2001) meta-analysis, there existed inconsistent results across studies and strong opinions of both sides on the usefulness of the GRE in predicting graduate performance. These inconsistent results resulted in some

contradictory ideas about how GRE scores can predict graduate performance. The limitations and gaps of previous studies are the impetus for further research about the validity of GRE scores.

Literature Review

Theoretical Rationale of the Predictive Evidence of Validity of GRE

The GRE was designed to test basic cognitive abilities that reflect the long-term learning of materials related to graduate performance. The GRE General Test measures the verbal reasoning, quantitative reasoning, critical thinking and analytical writing skills required for success in graduate and business school (Educational Testing Services, 2013). Specifically, the Verbal Reasoning subtest (GRE-V) measures the ability to analyze and evaluate written material and synthesize information obtained from it, and analyze relationships among component parts of sentences and recognize relationships among words and concepts. The Quantitative Reasoning subtest (GRE-Q) tests problem-solving ability, focusing on basic concepts of arithmetic, algebra, geometry, and data analysis. The Analytical Writing subtest (GRE-A) measures critical thinking and analytical writing skills, specifically the ability to articulate and support complex ideas clearly and effectively.

The belief that GRE scores can be used to predict graduate performance has been based on theoretical argument as well as empirical results. General cognitive ability has been said to be directly related with job knowledge, and job knowledge in turn is most strongly associated with school and job performance (Kuncel et al., 2001). Just as Kuncel, Crede and Thomas (2007) stated, the relationship between general cognitive ability and job performance is nearly fully mediated by job knowledge. According to a model of job

performance determinants from the field of industrial and organizational psychology (McColy, Campbell, & Cudeck, 1994), performance is conceptually a product of declarative knowledge (i.e., knowing what to do), procedural knowledge (i.e., knowing how to do), and motivation. The GRE measures verbal, quantitative and analytical abilities or skills. Test takers need to solve problems, synthesize information, and reason complex relationships between pieces of information. From these perspectives, Kuncel et al. (2001) argued that GRE scores would be correlated with the academic equivalent of job knowledge, and had an influence on graduate performance through declarative and/or procedural knowledge. Based on this analogy, new graduate students with higher GRE scores possess more “job” knowledge and skills, and these students should have a better graduate school performance when compared with those with less knowledge.

Kuncel et al. (2001) conducted a comprehensive meta-analysis of the predictive evidence of validity of the GRE. This 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 was generalizably a valid predictor of graduate performance.

Defining the Criterion: Graduate Performance

How to measure the criterion measure of graduate performance is important to be taken into account. In measuring performance, criterion relevance, accuracy, deficiency, and reliability are important considerations. Enright and Gitomer (1989) suggested that graduate performance was multidimensional, and Kuncel et al. (2001) identified eight different criteria that have been used as indicators of students’ success in graduate

programs. These criteria were graduate GPA (GGPA), 1st-year GPA, comprehensive examination scores, faculty ratings, number of publications-conference papers, number of times publications were cited, degree attainment, and time to attain degree. The predictive abilities of GRE scores were different on different indicators of success. According to the Kuncel et al. (2001) meta-analysis, the GRE was a better predictor of success for overall graduate GPA, first-year GPA, comprehensive exam scores, and faculty ratings than for research productivity, number of publication citations, time to degree attainment, and degree completion.

Among these indicators of student success, GPA perhaps best reflects the extent to which students master the material and acquire knowledge of the field of study, and thus GPA is suggested to be an indicator of both student ability and performance (Fenster, Markus, Wiedemann, Brackett, & Fernandez, 2001). As found in Morrison and Morrison's (1995) meta-analysis, among the 30 published articles examining the predictive evidence of validity of the GRE, 22 studies used GGPA as their criterion measure of academic success whereas others used diverse criteria such as faculty evaluations and levels of postgraduate productivity. Although GGPA has been pointed out to have limited value in terms of the predictive evidence of validity in some validation studies, GGPA (especially the 1st-year GGPA and cumulative GGPA) by far is the most widely used criterion of graduate school performance (Kuncel, Crede, & Thomas, 2007; Kuncel et al., 2001). For most of the specific research studies about the predictive evidence of validity of the GRE, either 1st-year grades or cumulative grades were used as the criterion measure of graduate success, while the performance of the second year was less frequently studied (Educational Testing Services, 2008a; Goldberg

& Alliger, 1992; Perez, 2011; Sternberg & Williams, 1997). The second year of graduate study is a continuous period after the first year in graduate school, so how students perform in the second year is also a reasonable criterion, as is overall grade-point average.

How to Test the Predictive Validity

Most of the research studies of the validity of GRE scores have used the correlations between GRE scores and a criterion and then present the percent of variance explained by the predictors to interpret the extent to which GRE predicts the graduate performance. However, Bridgeman, Burton and Cline (2009) argued that such ways to describe the validity are difficult to interpret, especially for non-technical audiences. It is possible that readers cannot understand what a variance means, for example, what 10% of variance means, and what an additional 10% of the variance explained by a second variable means, and so on. Correlations and variances are also likely to be misinterpreted. Despite the small percentages of variance accounted for by predictors, the predictors may actually be very important from a practical perspective. In order to display validity information easily for non-technical audiences, Bridgeman et al. (2009) employed expectancy tables to display the results. Students from different fields of study were divided into quartiles based on GRE scores and UGPA, and then the percentages of students earning the 1st-year GPA of 3.8 or higher (as a criterion of graduate success) were noted in both the top and the bottom quartiles. By comparing these percentages, one could compare the graduate success rates among students with high and low GRE scores and/or UGPA. Results from the expectancy tables demonstrated that the percent of students with high GRE scores falling in the top 25% of GPA was twice as high as the percent of students with low GRE scores. Also, by combining GRE and UGPA in one

graph, expectancy graphs could also provide information about the incremental validity of GRE over UGPA in predicting graduate success. This way, the expectancy tables gave a clear indicator of the potential value of test scores.

Despite the possibility that correlation coefficients and the percent of the variance explained may be difficult to interpret or be misinterpreted, this approach has been widely used and recognized in the professional literature. In addition, results from some studies showed that the extent to which the test predicts typical graduate outcomes has varied widely, with some studies indicating less than 10% of the variance in an array of criteria (e.g., 1st-year GPA, cumulative GPA) accounted for by GRE scores (Goldberg & Alliger, 1992; Morrison & Morrison, 1995; Sternberg & Williams, 1997). In spite of the little variance accounted for, researchers suggested that even slight improvement in validity could be useful, and GRE scores were indicated to be a valid predictor of graduate success (Holt, Bleckmann, & Zitzmann, 2006; Kuncel et al., 2001). Also, many studies introduced in this review have found reasonably large variance accounted for by GRE. Thus, there is still enough evidence to demonstrate the importance of the use of correlation coefficients and the percent of the explained variance as the indicators of predictive validity.

Specificity of Academic Disciplines in GRE Validity

Although for all graduate students, there are many similarities in some of the fundamental tasks required, there are differences in the types of training and the demands of different academic areas. For example, the social science majors may require higher language proficiency of their students than mathematics, physics, or engineering majors. Also, the grading standards and evaluation criteria may differ from discipline to

discipline. Thus, it is good to consider the predictive evidence of validity of the GRE specifically for different disciplines.

Kuncel et al. (2001) categorized different disciplines into four different fields: social sciences, math-physical sciences, humanities, and life sciences. This 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 (2012) 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. They gave the ceiling effect as a possible explanation. In effect, once a certain level of intellectual functioning as measured by GRE scores is reached, the predictive power of the GRE may become weaker. House and Johnson (1993) also found that the relationships between predictor variables and degree completion varied by area of graduate study or academic background. The result showed that GRE-V entered the regression model first as the best predictor of degree completion for students in professional psychology but entered the regression model last for general/experimental psychology.

There are very few studies except meta-analytical studies that examined various disciplines or made a comparison among different disciplines in one study. Most of the individual studies only examined the validity within one discipline, or they did not differentiate various disciplines even if they used a mixed sample. In a study by Powers (2004), the validity of GRE scores was based on the context of veterinary medicine. The

1st-year GPAs from a sample of 16 veterinary medical colleges were predicted by GRE scores and undergraduate GPA (UGPA). Because of the highly selective nature of veterinary medical school admission, statistical correlations were applied to correct for the effects of range restriction in the test scores and UGPA and for the unreliability of the criterion, which resulted in a significant increase in the validity estimates. After the statistical corrections, the validity coefficients were 0.53 for the combination of all three GRE General Test scores, 0.59 for UGPA, and 0.71 for the combination of GRE scores and UGPA together.

Because of the difficulty to compare and contrast the validity of GRE scores among various disciplines in a single study, and in order to have a better understanding of the validity of GRE, it is necessary and important to integrate studies that targeted the same or closely related disciplines. For example, there were a few studies that examined the predictive validity of GRE for psychology students. Sternberg and Williams (1997) examined the validity of GRE in predicting different kinds of performance in graduate psychology program in Yale University. The different criteria included the 1st-year and the 2nd-year GPA, professors' ratings of students' dissertations, and professors' ratings of students' analytical, creative, practical, research and teaching abilities. In sum, the results showed that GRE scores were found to be modest predictors of the 1st-year GPA but not the 2nd-year GPA, but of limited or no use in predicting other aspects of performance. Only GRE-A scores were predictive of consequential evaluations of student performance, but only for men. In the discussion part of the study, the authors explained some objections that might be raised against the design of the study and the interpretation of their results, such as the restriction of range, the unreliability of faculty ratings, and that

Yale graduate students could not represent a typical sample and so on. Another study by Schmidt, Homeyer, and Walker (2009) targeted counseling graduate students. This study used GRE scores and UGPA as the predictors to predict the Counselor Preparation Comprehensive Examination scores (CPCE). The results indicated that UGPA, GRE-V, and GRE-Q scores were valid for predicting CPCE scores, accounting for 21% of the variation. Among the three predictors, GRE-V was the best predictor because GRE-V scores predicted not only CPCE total scores, but also each of the eight CPCE subscale scores, and also the probability of passing the CPCE on the first attempt. But the author mentioned that one of the limitations of this study was that the generalization of the results to other programs was not determined. In the study by Fenster, Markus, Wiedemann, Brackett, and Fernandez (2001), GRE scores were used to predict both GGPA and time to completion (TTC) for forensic psychology students. The results showed that all predictors were appreciably related to GGPA, but the ability to predict TTC was smaller. In addition, the regression model with separate verbal and quantitative subscores ($R^2 = 0.34$) was better than that with a combined GRE total score ($R^2 = 0.32$) in predicting GGPA, but the difference in prediction was not substantial and the results did not provide statistical support for one approach over the other. From these studies targeting only psychology students, it may be seen that the validities of GRE scores, even though for students in closely related majors, appear somewhat inconsistent.

Criminal justice is another discipline studied by researchers. McKee, Mallory, and Campbell (2001) studied how undergraduate GPA (UGPA) and GRE scores predicted graduate GPA for master students in criminal justice at a medium-sized southern university. They found that UGPA alone explained about 24% of the variance in GGPA.

When controlling UGPA, the GRE-V explained the most variance (9.9%), followed by GRE-A (8.9%) and last by GRE-Q (6.9%). A total of 40% of variance in graduate GPA was accounted for by the combination of UGPA and GRE scores. This study suggested all three GRE subscales should be considered in the graduate admission process in criminal justice programs. Stack and Kelley (2002) also studied the validity of GRE scores for master students in criminal justice. They found that GRE-V alone could explain 21% of the variance in GGPA, and it was a better predictor of GGPA than GRE-Q. Reisig and DeJong (2005) examined the validity of GRE scores and undergraduate GPA for both master's and doctoral students in criminal justice from Michigan State University. The study found that although the GRE scores had different predictive abilities on different measures of academic performance, the GRE was still a good reference by graduate committees during the screening and admission process.

Holt, Bleckmann, and Zitzmann's study (2006) examined the validity of GRE for students in engineering management program. Results indicated that GRE scores modestly predicted students' first-year and cumulative GPAs where GRE-V and GRE-Q scores accounted for a significant larger portion of variance than undergraduate GPA, but GRE-A added nothing to the validity. Another study by Ayers and Quattlebaum (1992) examining Asian master's students in engineering found that the GRE-Q was the best predictor of success in graduate study. Feeley, Williams and Wise (2005) tested the validity of GRE scores in predicting success only for communication students. It found that UGPA was a better predictor of GGPA than GRE for both master's and doctoral students. For master's students, GRE-V was positively correlated with GGPA and GRE-Q was positively related to earning a degree.

The studies reviewed above provided the information that the predictive evidence of validity of the GRE was valued across various disciplines. However, there is little understanding of the variation in the relationship across disciplines. These studies indicated that the validities of the GRE were not consistent across different disciplines or among closely related majors. Just as argued by researchers, with respect to graduate education, the relationship of GRE scores and graduate performance may be dependent on the particular discipline being investigated (House & Johnson, 1993; Stack & Kelley, 2002; Thornell & McCoy, 1985).

Issue of Language Proficiency: American and International Students

The relationship between GRE scores and graduate performance may be moderated by some variables. The first potential moderator is language proficiency. In the United States, international students make up a nonignorable proportion of the graduate population. According to Young and Brooks (2008), there is occurring a vast change in the demographics of the United States, as the number of racial and ethnic minorities is quickly becoming a greater proportion of the population, projected to account for more than 50% of the total population by 2050. For most international students, English is not their primary or preferred language. The GRE tests are focused upon students' cognitive abilities, which are expected to be similar for both native and non-native English speakers, but scores on these tests also reflect the language proficiency. This point of view can be demonstrated by Stricker's study (2004), which found high correlations of TOEFL total scores with GRE-V and GRE-A and its moderate correlation with GRE-Q. English language proficiency is a critical factor for the academic performance of non-native speaker students in a setting where English is used

for teaching and learning. For this reason, besides GRE scores, American colleges or universities usually request international students to reach a minimum level of language proficiency as a threshold of admission (Cho & Bridgeman, 2012; Wait & Gressel, 2009).

The Test of English as a Foreign Language (TOEFL) is a test that evaluates the English proficiency of people whose native language is not English. It is one of the most commonly used English proficiency tests that institutions of higher education consider in determining whether a prospective student has met the level of language proficiency. Many studies have showed that TOEFL scores played an important role on the academic success for international students. For example, Wait and Gressel (2009) found that there was a positive and statistically significant relationship between TOEFL scores and GPA for international engineering students. Increasing TOEFL scores was also related to an increasing probability of success indicated by the pass rate of comprehensive assessment examinations and graduation rate. However, Vinke and Jochems' (1993) study indicated there was a cut-off point in the relationship between English proficiency and academic success. They suggested that there was a range of TOEFL scores within which an increase in the score was related to an increase in the chance of academic success, while below or above the limits of the range, an improvement in TOEFL scores had little or no effect on the academic performance. And the upper and lower limits of this range may vary by different academic disciplines or institution specific.

In Cho and Bridgeman's study (2012), the validity of TOEFL iBT (Internet-based testing) scores was examined by predicting the 1st-year GPAs for non-native English speakers from 10 universities in the United States. The additional information accounted for by TOEFL beyond other admissions related tests (i.e., GRE, GMAT, SAT, etc.) was

also examined. Correlation-based analysis found that TOEFL scores only explained 3% of the variance of GPA for both graduate and undergraduate students, and the incremental validity of TOEFL scores was shown to be fairly small. The expectancy graphs were also used to provide useful information of the validity of TOEFL scores. For example, the expectancy graphs showed that the probability of being in the top 25% GPA category doubled when their TOEFL scores were in the highest 25%, compared with those in the bottom 25% TOEFL group. Also, the students with relatively low TOEFL scores had a higher chance to earning low GPA than students with relatively high TOEFL scores.

As indicated above, though studies found some significant correlations between TOEFL scores and academic performance, the variance explained by TOEFL scores was small. Cho and Bridgeman (2012) summarized that the research findings on the power of TOEFL scores to predict academic success were mixed and inconsistent. Such diverse findings made it difficult to make a definitive conclusion about the validity of TOEFL in predicting academic success. Cho and Bridgeman gave some reasons why it was difficult for a language proficiency test to predict academic success. One reason is that there is no definite logic of the relationship between language proficiency and academic success. Language proficiency is a necessary but not a sufficient condition for academic success. Language is a critical factor in learning, but it is only one of many factors, such as motivation, learning strategies, and so on. One example of the explanation is, though most native speakers have no problem with their English language skills, not all native speakers are successful in academia. However, Cho and Bridgeman stated that even a small correlation or a trivial amount of variance explained could indicate a meaningful relationship between variables.

A qualitative study by Mupinga and Mupinga (2005) explored the perceptions of international students toward the GRE. Seven international students from different countries and different graduate programs were interviewed to establish their perceptions toward the GRE. Their perceptions toward the GRE included the content and context, structure, and purpose of the exam. A conclusion of the study was that it was very hard for a test to measure all aspects of intelligence no matter how well it was developed. The results found that the content and context of the GRE test, especially the GRE verbal section, were considered to be biased against international students, and they also believed that in general the test did not measure the cognitive ability to perform well in graduate school. Moreover, Milner, McNeil and King (1984) found that when GRE was eliminated from the admission process and only UGPA of 3.0 or higher was used as the admission criterion, there was a big increase in the minority enrollment rate; it doubled from 9.85% to 17.56%, and the elimination of GRE in the admission did not appear to decrease the quality of the students.

Perez (2011) stated that the effects of factors such as country of origin and English as a Second Language (ESL) had not been explored extensively in studies, but these factors were very important for the vast number of immigrants from non-English speaking countries who were entering the United States. Pennock-Roman (2002) studied the relations between GRE scores and another Spanish language standardized test among Puerto Rican students. She found that Puerto Rican students performed better on the Spanish test than on the GRE. Perez (2011) pointed out that the negative stereotypes of races or minorities might be a factor that affected students' perceptions of achievement and consequently affected students' performance in high stakes tests where White men

were perceived to achieve higher levels of performance. This explanation could be demonstrated in Steel's (1997) study where African Americans were found to score lower in diagnostic tests of academic aptitude (such as admission tests) than in non-diagnostic tests. Moreover, when comparing the performance of Black and White students, the differences were narrowed significantly in non-diagnostic tests than in diagnostic conditions.

As stated above, nonnative speaking students take up a nonignorable amount of the population in graduate programs in the United States. It is good to examine the difference in the predictive abilities of the GRE in predicting graduate performance between American and international students. Further, the extent to which language proficiency predicts graduate performance, and the influence of language proficiency on the predictive validity of GRE, or, the incremental validity of the GRE over language proficiency are also deserved to examine.

Undergraduate Academic Performance in Graduate Admission

One indication of graduate students' performance and success is GGPA. Similarly, undergraduate GPA (UGPA) is usually a good indicator of students' academic knowledge, abilities and performance during undergraduate studies, and it is normally considered together with GRE scores by admission committees in the graduate admission process. Applicants' GRE scores and UGPAs are the two most heavily weighted numerical or objective information in graduate admissions process (Norcross et al., 1996). In research studies, UGPA has also been combined with GRE scores to predict graduate GPA, and it was found to act as a very important role in predicting graduate performance. The comprehensive meta-analysis by Kuncel et al. (2001) examined the validity of GRE

scores and undergraduate GPA on predicting graduate school performance. By integrating data from 1,753 independent samples and 82,659 graduate students, this meta-analysis found that the GRE and UGPA are generalizably valid predictors of most of the indicators of graduate performance.

Reisig and DeJong (2005) examined the predictive abilities of GRE scores and prior GPA on multiple measures of performance (i.e., final graduate GPA, low grades, and incompletes) for master's and doctoral students of criminal justice from Michigan State University. This study found that GRE scores were not consistent in predicting different measures of academic performance. In contrast, undergraduate GPA was a consistently stronger predictor in predicting different measures of graduate student performance than GRE scores. The combination of GRE scores and undergraduate GPA was a fairly robust predictor of academic performance, and as expected this combination explained more variance in the dependent variable than either independent variable did by itself. This study suggested that both GRE scores and prior GPA should continue to be used in combination by graduate committees during the screening and admissions process. Powers (2004) found that when UGPA and GRE were used together to predict first year GGPA for students from veterinary medicine schools, they had a similar level of prediction, with the validity coefficients of 0.59 and 0.53, respectively, and with a multiple correlation coefficient of 0.71. McKee, Mallory, and Campbell (2001) studied master's students in criminal justice at a medium-sized southern university, and found that undergraduate GPA alone explained about 24% of the variance in GGPA, and a total of 40% of variance in GGPA was accounted for by the combination of undergraduate GPA and GRE scores. Williams and Wise (2005) tested the validity of GRE scores and

undergraduate GPA for communication students. It found that UGPA was a better predictor of GGPA than GRE for both master and Ph.D students. In Case and Richardson's (1990) study, based on the data from students in the Library and Information Science program at UCLA, undergraduate GPA was found to be the strongest predictor of GGPA. This study also found that GRE-V scores were more strongly correlated with GGPA than GRE-Q scores (i.e., $r = .354$ and $.285$, respectively). The results also showed that students with lower GRE scores and poor UGPAs were more likely to drop out of school. In Milner, McNeil and King's (1984) study, the minority enrollment rates were examined after GRE scores were eliminated from admission process but only using UGPA of 3.0 or higher as the admission criterion. This study found a significant increase in the minority enrollment rate when using this admission method, and UGPA was found to account for 9% of the variance in graduate class GPA, compared with 5.7% accounted for by the GRE.

However, in very few cases, the power of UGPA in predicting graduate success was not obvious. In the study by Smaby, Maddux, Richmond, Lepkowski, and Packman (2005), GRE scores and UGPA were used to predict counseling knowledge, counseling skills, and personal development of graduate students of counseling programs. Results indicated that GRE scores and UGPA were of limited value when used to predict success in counseling.

Based on the review of research about the validity of GRE scores, it appears that UGPA has at least as high value as GRE scores in predicting graduate student success. It is important to consider both GRE scores and UGPA to study their separate and combined influence on graduate performance.

Difference in Validity by Degree Level

Extensive research has examined the effectiveness of admissions tests for use in higher education, however, just as Kuncel, Wee, Serafin, and Hezlett (2010) summarized, the similarity of effectiveness for predicting performance at both the master's and doctoral levels had gone unexamined until recent years. There are differences in program complexity and structure by degree level, but would these differences have an impact on the validity of GRE? In order to answer this question and fill the research gap, the Kuncel et al. (2010) meta-analysis examined the difference in predictive evidence of the GRE by degree level. By integrating the results of about 100 studies involving about 10,000 students, this meta-analysis found that GRE scores had good validity for predicting the 1st-year GGPA, final GGPA, and faculty ratings for both master's and doctoral students, with differences ranging from small to zero. Perez (2011) studied the validity of GRE in predicting graduate success for students from a variety of disciplines at a Hispanic serving institution of higher education, and found that both the GRE-Q and GRE-V were good predictors of success for master's students, but the GRE-Q was not predictive of success for doctoral students.

Role of Range Restriction in GRE Validation Research

As noted above, validity evidence is usually expressed in terms of correlations. Correlational approaches face a problem if the data do not represent the full range of the population of interest, that is, information about how those who were not selected would have performed is missing. As a result, the correlation is often underestimated. Theoretically, the evaluations of the validity of the GRE should be based on all the students who took the test or who apply to the graduate school. In fact, however, the

validity of the GRE is based on individuals who have been admitted by graduate schools. Students with lower GRE scores than the minimum requirement of the schools are normally not admitted, so the range of GRE scores for graduate school incumbents is smaller than the range of graduate school applicants. This problem can attenuate the observed correlation between GRE scores and graduate school performance. Thus, due to the restriction of range of GRE scores among the selected students, the population value of the validity represented by the correlation between GRE scores and graduate performance criterion is almost always underestimated. This issue is range restriction. As Chernyshenko and Ones (1999) mentioned, two major reasons can result in the inconsistent results of GRE validation research: the criterion problem and the range restriction problem. The former identifies the limitations of all criteria of graduate performance. Moreover, besides the range restriction of GRE scores, graduate GPAs also have a range restriction. As Oldfield and Hutchinson (1997) pointed out, there are two kinds of range restrictions in graduate school research. One is from input variables resulting when students with low GRE scores and other admission variables are eliminated from the analysis, and another one is from output variables because most students receive an grade of A or B in courses which result in a narrow range of GPAs. The extreme problem of range restriction has been pointed out in some graduate admission validity studies (Oldfield and Hutchinson, 1997; Sterberg, & William, 1997), however, just as Kuncel et al. (2001) summarized, previous studies typically have not estimated the extent to which range restriction attenuates GRE validity coefficient.

To correct for range restriction, defining the population of interest is critical, and the ratios of selected group standard deviations to applicant pool standard deviations are

necessary (Kuncel et al., 2007). In the study by Chernyshenko and Ones (1999), the selection ratio was examined for 253 psychology Ph.D programs to estimate the effect of existing range restriction on GRE scores in validation. This study found that 11% of applicants were accepted and enrolled in the psychology programs, which means that the other 89% of the applicants were not fully available for validation research. After the correction of range restriction, GRE was found to be a valid predictor of graduate school performance. Chernyshenko and Ones (1999) argued that the controversy about the validity of GRE was the result of the reliance on small sample research and the lack of awareness of range restriction effects in graduate selection. The low selection ratio of graduate programs resulted in a restriction of variance of GRE scores, and consequently resulted in a decrease in observed validity coefficients. They also suggested that this problem produced erroneous results for previous investigations of GRE validity in which GRE was not found to be a valid predictor of graduate school success just based on observed correlations.

Kuncel et al. (2001) found considerably stronger relations among GRE scores and several criterion measures of graduate performance when correcting for range restriction. This review found the validity coefficients of GRE-V/GRE-Q with GGPA of 0.39/0.34, with faculty ratings 0.37/0.38, and with degree attainment 0.22/0.31. Power's (2004) study sampled from 16 veterinary medical schools which are also highly selective in admission. Thus, a restricted range of test scores and UGPA for enrolled students at each school was apparent. By applying statistical correlations to correct for the effects of range restriction in the predictors and for the unreliability of the criterion, it resulted in a significant increase in the validity estimates. Therefore, for future studies, the issue of

range restriction should be emphasized, and correction of range restriction should be considered if needed and applicable.

Purpose Statement and Research Questions

From the studies reviewed above, 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 results about the predictive evidence of validity of the GRE are 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 et al., 2001; Morrison & Morrison, 1995; Powers, 2004; Sternberg & Williams, 1997). A variety of conditions may have an impact on the validity, such as discipline specificity, the nature of the criteria, range restrictions, sampling errors, and some other uncontrolled factors. Thus, more research is needed to fill the gaps existing in previous studies. Moreover, as mentioned above, language minorities for whom English is not their first language are becoming an increasing number of the student population in graduate schools in the United States. In the present study, the difference in the validity of GRE scores between native English speakers and non-native English speakers was tested. The language issue was also taken into account to investigate the incremental validity of GRE scores over language proficiency.

Among all the various disciplines, engineering is an important graduate program in universities, and the importance of the engineering graduate degree is increasingly being recognized by the professional engineering community (Rogers & Goktas, 2010). As the National Academy of Engineering (2005) pointed out, the typical engineering bachelor degree cannot accommodate the academic development required for

professional engineers, and that the master's degree should be considered the first professional engineering degree. As a consequence of this awareness of the importance of the graduate engineering degree, 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). Facing the increasing number of applicants, admission committees for engineering programs are also facing the questions of how to identify the best students who fit the program through their admission process. The joint importance of admission issues and engineering education encourages more research on this field. However, from the literature review, it was found that the validity studies of GRE scores have not widely extended to the discipline of engineering. Thus, the present study intended to fill this gap and to target the population of engineering students to study the predictive validity of GRE scores in predicting graduate performance. In addition, as mentioned in the literature, since it was lacking of research differentiating degree level, this study also tested the difference of the validity between master's and doctoral students.

Specifically, the research questions were:

- 1) How do GRE scores predict engineering students' 1st-year, 2nd-year and total GGPA in the 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 master and doctoral students?
- 4) How does UGPA predict engineering students' 1st-year, 2nd-year and total GGPA in the 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 the graduate program? What is the incremental validity of GRE over language proficiency?

Significance of the Study

This study has both theoretical and practical values. Theoretically, it extends the research of the validity of GRE scores, and it contributes to an extensive knowledge of the effectiveness of GRE scores in predicting graduate success in graduate schools. Examining the predictions of 1st-year, 2nd-year, and total GGPA together in one study can reveal significant results about the predictive abilities of the GRE from both short-term and long-term perspectives. This extensively enriches the research on the predictive evidence of validity of the GRE, and also fills the gap of limited research on long-term prediction. Moreover, targeting engineering students fills a gap that limited previous research studied the validity of the GRE for this specific discipline. This study also provides several practical considerations. First, this study provides useful information for graduate admission decisions, important for both universities and applicants. Admission committees can make scientific decisions to recruit the right applicants to the graduate programs, and applicants can gain the appropriate opportunity to receive graduate education. The admission decisions can have a critical impact on the quality of the output of education, and on students' futures of life and careers. In addition, taking the issue of language proficiency into account will aid in the consideration of the equity in assessment. A test is considered to be biased if its predictive power is not equivalent for different subgroups (Johnson, Carter, Davison, & Oliver, 2001). As stated by Sandoval

and Durán (1998), fairness demands that extra care be taken with the growing population of language minorities. All tests normed on native speakers of English, to some extent, are measures of English competency and proficiency, and “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)”. By comparing the validities between American and international students, and examining the effect of TOEFL scores in predicting graduate performance can give a more comprehensive interpretation of the utility of GRE scores. It provides valuable implications about the possible different evaluations for American students and international students in the admission.

CHAPTER II

RESEARCH METHODS

Sample

The sample of this study was from the various engineering programs of a large comprehensive midwestern university in the United States. It covered a variety of programs of engineering, including Electrical Engineering, Manufacturing Engineering, Civil Engineering, Architectural Engineering, Environmental Engineering, Chemical Engineering, Mechanical Engineering, and some other engineering programs. Student records were obtained by submitting a data request from the university's office of Institution Research and Planning. The university kept the data of students since 2000. In this study, the data were collected from all students who were enrolled in these engineering programs during the 11 academic years, from 2000 to 2011. But only the students who had registered classes for at least one and a half years or 3 academic semesters (in order to get the 1st-year and the 2nd-year GGPA) were retained for use in this study.

Procedures

The data were obtained separately from different databases maintained by the university's office of Institution Research and Planning and the Graduate Studies. One was the admission file, a database containing the information from the applications submitted to the university, including GRE scores, UGPA, and TOEFL scores, and some other information. Another database contained the information about the demographics (e.g., gender, country of origin, etc.), the graduate status (e.g., enrolling terms, department, major, degree, etc.), and the records of coursework performance. The records

of coursework included the grades in each semester and the corresponding credit hours in each semester, and the cumulative grades and the corresponding cumulative credit hours at each point of semester. Then, the 1st-year GGPA and the total GGPA were obtained directly from the records, but the 2nd-year GGPA was obtained by averaging the grades of all the classes in the second year of graduate school. These databases were then organized and joined into a single database in order to obtain a more complete data profile for each student. Finally, only the data of interest were used, which included admission status, nationality, sex, degree level, department, major, entry term, exit term, GRE scores (Verbal, Quantitative, Analytical Writing, and total scores), TOEFL scores, UGPA, first-year GGPA, second-year GGPA, cumulative GGPA, and the corresponding credit hours.

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. Students with none of the three predictor scores were eliminated. Moreover, in synthesizing all the given records of grades and credit hours for each student, inconsistencies and incompletions in the data were found for some students. For example, by averaging the grades of classes in the first one or two years, the results for some students did not match with their given cumulative grades. This situation made it impossible to determine what the grades actually were for these students. Thus, these students with inconsistent or wrong records were eliminated. After meeting these selection criteria, a total of 1083 students ($N=1083$) 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. In terms of gender, 79.8% of the students were male and 20.2% were female. Regarding degree level, master's students consisted 61.3% of all the selected students, and doctoral students were the remaining 38.7%.

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). The three criterion measures were available for all the selected students ($N=1083$). GGPA was on a four-point scale ranging from 0 through 4.0. 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 beginning when students enrolled in graduate study until the end they graduated or to the last semester the data covered in the database (i.e., 2012 Fall). The number of years which total GGPA covered differed by individual. For example, doctoral students usually had more years of records than master's students. In addition, 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 in the database.

UGPA. UGPA is a cumulative grade point average covering all the undergraduate coursework. It was also on a four-point scale ranging from 0 to 4.0. However, in this study, this measure was only available for the students who had studied and got their bachelor's degrees from the targeted midwestern university. For other students who graduated from other universities or colleges, their UGPAs were not available because the university did not retain this record of admission.

Graduate Record Examination. GRE has three subtests that measure verbal reasoning, quantitative reasoning, and analytical writing skills (Educational Testing Services, 2013). Specifically, the Verbal Reasoning subtest (GRE-V) measures the ability to analyze and evaluate written material and synthesize information, and analyze relationships among component parts of sentences and recognize relationships among words and concepts. The Quantitative Reasoning subtest (GRE-Q) tests problem-solving ability, focusing on basic concepts of arithmetic, algebra, geometry, and data analysis. The Analytical Writing subtest (GRE-A) measures critical thinking and analytical writing skills, specifically the ability to articulate and support complex ideas clearly and effectively. The GRE-V and the GRE-Q had a possible score range from a minimum of 200 to a maximum of 800. GRE-A tests were changed during this period with test scores following two different scales: one was on a 7-point scale ranging from 0 to 6.0; another one had the same scale as GRE-V ranging from 200 to 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.

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, testing listening, reading, speaking and writing skills. Before 2006, TOEFL tests had different versions - paper-based TOEFL (PBT) and computer-based TOEFL (CBT), testing listening, structure/writing, and reading skills, no speaking (Alderson, 2009). These three versions have different scoring scales. The TOEFL Score Comparison Tables are available to show the relationship between the new TOEFL iBT scores and the scores from the CBT and PBT versions of

the TOEFL tests (Educational Testing Services, 2005). Institutions may choose to set TOEFL iBT scores that are comparable to the CBT and PBT for the total score, and for the Reading, Writing and Listening scores. In this study, 53.1% ($N = 575$) of the total students had TOEFL scores. Among these students, 211 students took TOEFL PBT, 176 iBT, and 188 CBT. In order to integrate the TOEFL scores, these three versions of scores were placed on one scale. Based on the characteristics of the scores on the TOEFL Score Comparison Tables, this study transformed TOEFL iBT scores and PBT scores into the CBT scores. The TOEFL CBT has a scoring range of 0 to 300. Since the difference of the subtests of each version (i.e., TOEFL iBT has a subtest which tests the speaking skills but CBT and PBT not), only the total scores were used and transformed. For those students (i.e. only five students in this study) who took the TOEFL more than one time, the highest score was used as his or her TOEFL score.

Demographics. The term alien status in this study represented whether or not the students were the residents of the United States. In the final sample, 41.6% of the students were US residents, and the remaining 58.4% were aliens. 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 midwestern university were noted as doctoral students. Students who enrolled as master's students and did not continued to the doctoral programs of the same university were noted as master's students.

Data Analysis

The data analysis for this study consisted of both descriptive and inferential statistics. Some demographic information was presented at the beginning, such as the frequencies and percentages of different groups (e.g., alien status, degree, gender). The

descriptive statistics were computed for all three predictor variables (GRE scores: Verbal, Quantitative, total; UGPA, and TOEFL) and three criterion variables (1st-year, 2nd-year, and cumulative GPA). In addition, the descriptive statistics were also computed for different groups (i.e., American vs. international students; master's vs. doctoral students). The group differences were further tested by independent *t* test. The difference of variables by alien status and degree level would help us to explain the following differential prediction of GRE scores between different subgroups.

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. The standardized regression coefficient provided the information about the direction and the strength of the prediction, and *R* square informed us how much variance was explained by the predictor. The corresponding *F*-test tested the significance of the prediction. To test the difference in the validity of GRE scores by alien status and degree level, separate regressions were conducted for American and international students, and for master's and doctoral students.

Hierarchical multiple regression analyses were conducted to examine the incremental validity of GRE scores over UGPA in predicting 1st-year, 2nd-year, and total GGPA. This analysis evaluated whether UGPA was a good predictor of graduate performance, and whether adding GRE scores could improve the prediction and how much more variance in GGPA could be explained by GRE scores over/beyond UGPA. In the analysis, the UGPA was entered into the hierarchical model first (i.e., model 1), then GRE scores (GRE-V, GRE-Q, GRE-tot, separately) into the model (i.e., model 2). Hierarchical multiple regression analysis provides the proportion of variance in the

criterion that is explained by the predictor variables in each model, indicated by R^2 , and the change in R^2 in the latter model over the former model. In this case, the R^2 in model 1 was the variance explained by the UGPA, then the R^2 in model 2 was the variance explained by both UGPA and GRE scores, and the change in R^2 was the additional variance explained by GRE score but not by UGPA. In addition, by using F -test, the significances of R^2 and R^2 change were tested to determine whether the explained variances were significant and whether GRE scores added significant more variance over UGPA. Similarly, to test the incremental validity of GRE scores over TOEFL scores, this hierarchical multiple regression analysis was used in the same way.

As mentioned in the section of research methods, because of incomplete records of many students, only the students who had full records of the three criterion variables and the record of at least one of the three predictors (GRE, TOEFL, UGPA) were retained in the final database. Finally, 1083 students were retained, however, GRE scores were available for 591 of these students, UGPA for 398, and TOEFL scores for 575 students. Among these students, only 401 students had both GRE and TOEFL scores (and of course these students were all international students), and 65 students had both GRE scores and UGPA, however, only 5 students had all the three predictor scores. Among these numbers, the overlapped students (i.e., those students who had the scores of two predictors) were different. To solve the problem of incomplete data, this study applied the Listwise deletion to deal with the missing data. As a consequence, in answering each research question, different samples and the corresponding different sample sizes were used in the analysis, as the variables involved in each question differed from each other.

The software SPSS 16.0 was utilized to conduct all the analyses.

CHAPTER III

RESULTS

This chapter presents the results of this study that investigated the validity of GRE scores in predicting graduate performance for engineering students. Descriptive statistical analyses were conducted first for the demographic and research variables. To examine the validity of GRE scores specified in each research question, the methods of Pearson's correlations, simple linear regression, and hierarchical multiple linear regression were utilized.

Descriptive Statistics

The database contained the information for a total of 1083 students ($N = 1083$). Among these students, 41.64 % ($N = 451$) were non-alien (i.e., American students) and 58.36% ($N = 632$) were Alien (i.e., international students); 61.31% ($N = 664$) of the students were master's students, and 38.69% ($N = 419$) were doctoral students. As shown in Table 1, amongst master's students, American students and international students constituted 55.7% and 44.3%, respectively. Amongst doctoral students, international students composed a much larger percentage (80.7%) than American students (19.3%). American students for most part ($N = 370$, percent = 82.0%) were master's students, compared with a much smaller number of doctoral students ($N = 81$, percent = 18%). By contrast, for international students, a relatively close proportion of students distributed between master's and doctoral students (i.e., 46.5% and 53.5%, respectively).

Table 1

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

Master's Students	Doctoral Students
-------------------	-------------------

Alien Status	<i>N</i>	%	<i>N</i>	%	Total
US	370	55.7%	81	19.3%	451
Alien	294	44.3%	338	80.7%	632
Total	664	100%	419	100%	1083

As for academic characteristics, Table 2 presents the means, standard deviations, and minimums and maximums of all criterion variables and predictor 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. They were all on a 0-4 scale. The corresponding credit hours in each period of time were different, and the ranges of hours were large. The mean credit hours was 20.24 ($SD = 6.21$) for the 1st-year GGPA, 17.14 ($SD = 7.64$) for the 2nd-year GGPA, and 50.21 ($SD = 27.24$) for total GGPA. The distribution of credit hours for total GGPA was quite varied compared to 1st-year and 2nd-year GGPA. One possible reason for the differences in credit hours was that the sample included both masters' and doctoral students, and doctoral students usually studied for more years and earned more credits than masters' students. Another possible reason was that some students had completed their course work but some others not

The three predictor variables in this study were GRE scores, UGPA, and TOEFL scores of international students. This study considered the GRE Verbal and Quantitative and total scores separately. The mean of GRE total score was 1166.99 ($SD = 159.95$, ranging from 650 to 1600). The GRE-V had a lower mean score, larger standard deviation, and larger range ($M = 435.25$, $SD = 119.42$, ranging from 200 to 800) than GRE-Q ($M = 731.74$, $SD = 73.32$, ranging from 320 to 800). Both GRE-V and GRE-Q

scores were on a scoring scale of 200 to 800. The mean UGPA was 3.41 ($SD = .36$) on a 0-4 scale. The mean of TOEFL total score was 236.14 ($SD = 27.30$), with a range of 130 to 293 on a 0-300 scale.

Table 2

Descriptive Statistics of the Predictors and the Criteria

Variables		<i>N</i>	<i>M</i>	<i>SD</i>	Minimum	Maximum
Criteria	1.GGPA-1	1083	3.64	.32	2.00	4.0
	(1 st -year hours)		(20.24)	(6.21)	(5)	(48)
	2.GGPA-2	1083	3.68	.33	1.67	4.0
	(2 nd -year hours)		(17.14)	(7.64)	(1)	(60)
	3.GGPA-tot	1083	3.68	.27	2.44	4.0
	(total hours)		(50.21)	(27.24)	(9)	(174)
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

Pearson correlations were computed among all variables. As shown in Table 3, all correlations were significant at an $\alpha = .01$ level except the one between TOEFL and UGPA which was significant at an $\alpha = .05$ level. The correlation between 1st-year and 2nd-year GGPA was .56, and the two one-year GGPAs both correlated with total GGPA at .82 and .81, respectively, although these correlation coefficients were spurious. GRE-V

correlated with GRE-Q at .34, and these two subtest scores had a spurious correlation with GRE-tot at .90 and .71, respectively. The correlations between predictor variables, GRE and TOEFL scores, and 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.

Table 3

Correlations among Variables

	GGPA-1	GGPA-2	GGPA-tot	GRE-V	GRE-Q	GRE-tot	UGPA	TOEFL
GGPA-1	1							
GGPA-2	.56**	1						
GGPA-tot	.82**	.81**	1					
GRE-V	.17**	.19**	.21**	1				
GRE-Q	.22**	.18**	.26**	.34**	1			
GRE-tot	.23**	.22**	.28**	.90**	.71**	1		
UGPA	.64**	.52**	.64**	.37**	.51**	.50**	1	
TOEFL	.15**	.17**	.22**	.54**	.23**	.51**	.50*	1

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

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

The means and standard deviations of variables for American and international students were presented in Table 4. In order to see the significance of the difference in means between two groups, independent sample t tests were conducted. As mentioned before, because the majority of students who had UGPA were American students and

TOEFL scores were only for international students, this descriptive analysis did not include UGPA and TOEFL. There were no significant differences between American students and international students in 1st-year, 2nd-year, and total GGPA ($p > .05$).

American students had higher GRE total scores than international students ($M_{US} = 1181.69$, $M_{Alien} = 1162.84$, respectively), but the result of t -test showed that this difference was not significant ($t = 1.19$, $p > .05$). The differences in GRE-V and GRE-Q were found to be significant in t -test ($p < .01$). Specifically, American students had higher GRE-V scores than international students ($M_{US} = 483.92$, $M_{Alien} = 421.52$, respectively; $t = 6.17$, $p < .01$), but international students had higher GRE-Q scores than American students ($M_{Alien} = 741.32$, $M_{US} = 697.77$, respectively; $t = 5.31$, $p < .01$).

Table 4

Descriptive Statistics of GGPA and GRE scores for Alien and American Students

Variables		Alien status	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>
Criteria	GGPA-1	Alien	632	3.65	.32	.461
		US	451	3.64	.33	
	GGPA-2	Alien	632	3.69	.31	1.54
		US	451	3.66	.34	
	GGPA-tot	Alien	632	3.70	.26	1.61
		US	451	3.67	.28	
Predictors	GRE-V	Alien	461	421.52	121.97	6.17**
		US	130	483.92	95.47	
	GRE-Q	Alien	461	741.32	66.06	5.31**
		US	130	697.77	86.78	
	GRE-tot	Alien	461	1162.84	159.48	1.19
		US				

US	130	1181.69	161.36
----	-----	---------	--------

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

The descriptive statistics of variables are also presented for masters' and doctoral students, and the significance of the differences between the two groups were analyzed by using independent sample t tests. As shown in Table 5, the differences in means were found to be significant in 1st-year, 2nd-year and total GGPA, and also in GRE-Q, GRE-tot, and UGPA ($p < .01$), but not in GRE-V and TOEFL scores ($p > .05$). Specifically, doctoral students had higher scores than master's students in graduate performance, as measured by 1st-year GGPA ($M_{Doc} = 3.76$, $M_{Ma} = 3.58$; $t = 9.78$, $p < .01$), 2nd-year GGPA ($M_{Doc} = 3.75$, $M_{Ma} = 3.63$; $t = 6.14$, $p < .01$), and total GGPA ($M_{Doc} = 3.78$, $M_{Ma} = 3.62$; $t = 10.55$, $p < .01$). Doctoral students had higher scores than master's students in GRE-Q ($M_{Doc} = 745.22$, $M_{Ma} = 717.95$; $t = 4.59$, $p < .01$), in GRE total scores ($M_{Doc} = 1186.22$, $M_{Ma} = 1147.29$; $t = 2.98$, $p < .01$), and in UGPA ($M_{Doc} = 3.59$, $M_{Ma} = 3.39$; $t = 3.88$, $p < .01$). On the whole, these results indicated that doctoral students had better performance in the GRE, UGPA, and graduate performance than master's students.

Table 5

Descriptive Statistics of Variables for Master's and Doctoral Students

Variables		Degree	N	M	SD	t
Criteria	GGPA-1	MS	664	3.58	.35	9.78**
		PHD	419	3.76	.24	
	GGPA-2	MS	664	3.63	.35	6.14**
		PHD	419	3.75	.27	
	GGPA-tot	MS	664	3.62	.28	10.55**

		PHD	419	3.78	.20	
Predictors	GRE-V	MS	292	429.35	117.44	1.19
		PHD	299	441.00	121.24	
	GRE-Q	MS	292	717.95	79.13	4.59**
		PHD	299	745.22	64.49	
	GRE-tot	MS	292	1147.29	161.94	2.98**
		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 6, all standardized regression coefficients were significant ($p < .01$), indicating the usefulness of the GRE-V, GRE-Q, and GRE total scores in predicting 1st-year, 2nd-year, and total GGPA. Increases in GRE scores generally indicated increases in GGPA. The variances in criterion variables that were explained by GRE scores, indicated by R^2 , ranged from 2.8% to 7.9%. The GRE-V explained relatively less variance ($R^2 = 2.8\%$, 3.4%, 4.6%, respectively for 1st-year, 2nd-year, and total GGPA) than GRE-Q ($R^2 = 5\%$, 3.2%, and 6.9%) and GRE total scores ($R^2 = 5.2\%$, 4.9%, 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%). On the whole, these results indicated that GRE scores predicted total GGPA better than 1st-year and 2nd-year GGPA, and GRE-Q and GRE total scores explained more variance in graduate performance than GRE-V scores.

Table 6

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

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

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

Regression of graduate performance on GRE scores was conducted for American students and international students. As shown in Table 7, all the standardized coefficients were significant ($p < .01$) which indicated the usefulness of the predictive ability of GRE scores for both American students and international students. For American students, the variance explained by GRE scores ranged from 10.5% to 22.2%. Among the three GRE scores, GRE-Q and GRE total scores explained more variance than GRE-V across three criterion variables. Specifically, GRE-Q explained 13.6% to 22.2% of the variance, and GRE total explained 15.2% to 21.8%, but GRE-V explained 10.5% to 13.1%. For international students, the variance explained by GRE scores ranged from 1.4% to 5.5%.

Similar to American students, GRE-Q and GRE total also explained more variance than GRE-V for international students. Specifically, GRE-Q explained 2.1% to 5.5% of the variance, and GRE total explained 3.1% to 5.2%, but GRE-V explained 1.4% to 3%. As a whole, across the three criterion variables, GRE scores explained more variance for American students than for international students. Although there appeared to be difference in the validity by Alien Status, however, in testing the statistical significance of the difference, the interactions between Alien Status and GRE scores 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 by Alien Status was found only when GRE-Q was used in predicting GGPA total scores.

Table 7

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

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

To test the validity of GRE for students in different degree levels, regression of graduate performance on GRE scores was conducted for master's and doctoral students. As shown in Table 8, except the prediction by GRE-V on the 2nd-year GGPA and total GGPA for doctoral students, all other regressions were useful, indicated by the statistical significance of the standardized regression coefficients ($p < .05$, or $p < .01$). 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 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 1st-year, 2nd-year, and total GGPA. 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. Although there appeared to be difference in the validity by degree level, however, further statistical tests with both degree level and GRE scores in the regression model showed that the interactions between degree and GRE scores were not found to be significant 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 ($F(26, 528) = 1.57, p < .05$).

Table 8

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

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

Incremental Validity of GRE over UGPA and TOEFL

The incremental predictive ability was analyzed by using hierarchical multiple regression. As the three GRE values were considered separately, three hierarchical multiple regressions were conducted for all the criterion variables. 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 SLR prior to MR so as to include all the 398 students who had UGPA.

As shown in Table 9, 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 GGPA, and total GGPA, respectively. In the hierarchical multiple regression ($N = 65$), 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, indicated by R^2 and R^2 change (i.e., ΔR^2). The significance of R^2 change indicated that GRE scores explained a significant

additional proportion of variance over/beyond what UGPA explained. Specifically, for the 1st-year GGPA, GRE-V explained additional 5% of variance beyond UGPA ($p < .05$), GRE-Q 9.6% ($p < .01$), and GRE total 10% ($p < .01$). For the 2nd-year GGPA, GRE-V explained additional 6.3% of variance over UGPA ($p < .05$), GRE-Q 5.2% ($p < .05$), and GRE total 8.4% ($p < .01$). For total GGPA, GRE-V increased the explained variance by 2.8% over UGPA ($p > .05$), GRE-Q 3.8% ($p < .05$), and GRE total 4.7% ($p < .05$). On the whole, GRE-Q and GRE total scores had larger incremental predictive abilities over UGPA than GRE-V in this population. The incremental validity of GRE scores over UGPA was presented more obviously for 1st-year GGPA and 2nd-year GGPA than for total GGPA.

Table 9

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		UGPA	.639**	.408		.524**	.274		.635**	.403	
(N=398)											
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-	.365**	.466	.100**	.335**	.353	.084**	.251*	.450	.047*
tot											

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

Similar as the test of the incremental validity of GRE scores over UGPA, the same method was used to examine the incremental validity of GRE scores over TOEFL scores. There were 401 students presenting both GRE and TOEFL scores out of 575 students who had TOEFL scores. The predictive ability of TOEFL scores was tested by SLR prior to MR.

As shown in Table 10, the results of SLR ($N = 575$) showed that TOEFL scores explained 2.2% ($p < .01$), 3% ($p < .01$), and 4.9% ($p < .01$) of variance in 1st-year GGPA, 2nd-year GGPA, and total GGPA, respectively. However, in the hierarchical multiple regression ($N = 401$), TOEFL scores did not explain significant proportions of variance ($p > .05$). After adding GRE scores into the model, the proportions of explained variance, indicated by R^2 and R^2 change, increased across the GRE-V, GRE-Q and GRE total scores for all the three criterion variables. The significance of R^2 change indicated that GRE scores explained a significant additional proportion of variance over/beyond what TOEFL scores explained. Specifically, for the 1st-year GGPA, GRE-V explained additional 1.7% of variance beyond TOEFL ($p < .01$), GRE-Q 4.6% ($p < .01$), and GRE total 10% ($p < .01$). For the 2nd-year GGPA, GRE-V explained 2.1% of variance over UGPA ($p < .01$), GRE-Q 1.4% ($p < .05$), and GRE total 2.7% ($p < .01$). For total GGPA, GRE-V explained 2.7% of variance over TOEFL ($p < .01$), GRE-Q 5.2% ($p < .01$), and GRE total 5.3% ($p < .01$). On the whole, GRE-Q and GRE total scores generally added more variances over TOEFL than GRE-V scores. The incremental predictive ability of GRE scores over TOEFL was shown more obviously in 1st-year GGPA and total GGPA than in 2nd-year GGPA.

Table 10

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	TOEFL		.149**	.022		.173**	.030		.221**	.049	
(N=575)											
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-	.230**	.042	.039**	.191**	.030	.027**	.268**	.062	.053**
tot											

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

CHAPTER IV

DISCUSSION AND CONCLUSIONS

This chapter provides a brief summary and interpretation of the findings as they related to each of the research questions. These findings are analyzed and contrasted with the results of previous research studies. Implications derived from the findings are also analyzed in this part. The limitations of the study are reported, and some suggestions for future research are also provided.

The purpose of this study was to investigate the predictive evidence of validity of GRE scores on predicting short-term and long-term graduate performance for engineering students. Considering UGPA was also an important admission variable, this study tested the predictive ability of UGPA and the incremental validity of GRE over UGPA--whether GRE added significantly more power beyond UGPA in the prediction of graduate performance. Moreover, considering that international students are composing a considerable portion of the graduate population and that the issue of the fairness of assessments for subgroups is receiving considerable attention, the difference in the validity of GRE scores between American students and international students was examined. The majority of international students were language minorities, so the language proficiency was taken into consideration in this study, and the incremental validity of GRE over language proficiency was tested. As there were few amount of studies taking into account the impact of degree level on the prediction (Kuncel et al., 2010), this study also compared the validities of GRE between master's students and doctoral students.

The data in this study were obtained from 1083 students from the graduate engineering programs in a large comprehensive midwestern university. Since not all students had complete data for all variables, this study applied the Listwise deletion to deal with the missing data in the data analysis. As a consequence, in answering each research question, different samples and the corresponding different sample sizes were used. The simple linear regression (SLR), and hierarchical multiple regression statistical techniques were utilized to answer the research questions. Descriptive analyses and independent sample *t* tests were also used to provide general information about students in terms of the scores of the predictors and criteria between different subgroups. The GRE-V, GRE-Q, and GRE total scorers were considered separately, and the analyses were also conducted separately for the three criterion variables (i.e., 1st-year, 2nd-year, and total GGPA).

Findings of the Study

The descriptive statistics for the predictor variables (GRE, UGPA, and TOEFL) and criterion variables (1st-year, 2nd-year, and total GGPA) were first obtained for different subgroups of alien status and degree level. As for alien status, there were no significant differences between international students and American students in terms of the graduate performance, as measured by 1st-year GGPA, 2nd-year GGPA, and total GGPA. In terms of GRE scores, international students and American students had no significant differences in GRE total scores; however, American students had significantly higher GRE-V scores but lower GRE-Q scores than international students. The comparison by alien status was not available for UGPA as only a very small number of international students had UGPA. As for degree level, the differences between master's

and doctoral students were obvious in most of the variables. Doctoral students had significant higher graduate performance (i.e., 1st-year GGPA, 2nd-year GGPA, and total GGPA), higher GRE-Q and GRE total scores, and higher UGPA than master's students. Though doctoral students also had higher GRE-V and TOEFL scores than master's students, this difference was not found to be statistically significant.

The distinct scores in GRE-V and GRE-Q between different subgroups in this study confirmed the right choice to separate the consideration of GRE-V and GRE-Q so as to have a better understanding of the characteristics. The difference in GRE scores between American students and international students, to some extent, corresponded to the findings of ETS (Educational Testing Services, 2008b) that minority students usually received significant lower GRE scores than White students, with the exception that Asian students usually got higher score on the GRE-Q section. This finding was easy to understand. The GRE is tested in the English language environment which American students are more familiar with than those international students for whom English is not their first language. This difference was evident especially in GRE Verbal section of the test that requires a relatively high level of English proficiency to get a high score in this subtest. Mupinga and Mupinga's (2005) qualitative study provides a possible explanation. They found that the content and context of the GRE test, especially the GRE Verbal section, were very difficult and thus likely to be perceived as biased against international students. It was possibly because the Verbal subtest requires a difficult vocabulary and measures the ability to analyze and evaluate written material and synthesize information, and analyze relationships among component parts of sentences and recognize relationships among words and concepts (Educational Testing Services, 2013).

In terms of the discrepancies between master's and doctoral students, there were some possible reasons. One possible explanation could be the different career goals of master's and doctoral students. Doctoral students need to study more years in graduate school, to obtain more academic knowledge and skills, and to receive more academic training in order to get a degree of Ph.D, and finally they are more likely to seek a career related to research. In contrast, master's students are usually starting their careers after only two or three years of graduate study, and are more likely to start their career in practical settings where the requirements in research are less needed. The second possible reason could be that, because of the different career goals, the concentration and the amount of time and energy focused on study and the motivation of study would be quite different for master's students and doctoral students. The grades in graduate school seem to be more important for doctoral students than for master's students.

Research question 1: How do GRE scores predict engineering students' 1st-year, 2nd-year and total GPA in the graduate program?

The general predictive ability of GRE scores was first tested by using simple linear regression (SLR). The significant regression coefficients of the SLR indicated that there were linear relationships between GRE scores and GGPA. Students with higher GRE-V, GRE-Q, and GRE total scores tended to have higher 1st-year, 2nd-year, and total GGPA. The variances in the criterion variables that were explained by GRE scores ranged from 2.8% to 7.9%. In general, GRE-Q and GRE total scores had larger predictive ability (explaining 5% to 7.9% of variance) than GRE-V (explaining 2.8% to 4.6% of variance). Across the three criterion variables, the variance in total GGPA was explained

(4.6% ~7.9%) more than that in 1st-year and 2nd-year GGPA (2.8% ~ 5.2%). On the whole, GRE was a valid predictor of graduate performance, as measured by 1st-year, 2nd-year, and total GGPA.

The findings of the general analysis of the validity of the GRE corresponded with many studies which found that GRE scores are generally valid predictors of students' performance during their studies in graduate school (Case & Richardson, 1990; Educational Testing Services, 2008a; Hyun, 2012; Kuncel et al., 2001; Perez, 2011). Most of the previous research about how GRE scores predicted graduate success were limited to only the first-year grades or only final grades, but the second-year was less frequently studied, nor were simultaneously two or more criterion variables considered in one study (Educational Testing Services, 2008a; Goldberg & Alliger, 1992; Perez, 2011; Sternberg & Williams, 1997). 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 GRE was found to predict long-term performance better than short-term performance in this study. One possible explanation could be that total GGPA (which covered more years of grades) is more reliable than the one-year GGPA. Kuncel et al. (2001) provided the most definitive evidence regarding the predictive validity of GRE scores, and they found that GRE was a valid predictor of final GPA and 1st-year GPA, but the difference about how final GPA and 1st-year GPA were predicted was not examined. In addition, Kuncel et al. (2001) found that the predictive validity coefficients for GRE-V and GRE-Q scores were very similar (operational validity coefficients = .34, and .33, respectively) when different disciplines were combined together. However, when separating different discipline areas,

they found that 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). Corresponding with the results of Kuncel et al., this study found that GRE-Q as well as GRE total scores explained more variance in graduate performance than GRE-V scores for engineering students. Other studies found different results, but the samples were from disciplines other than engineering. For example, Stack and Kelley's (2002) examined the validity of GRE scores for master's students in criminal justice, and found that GRE-V (alone explained 21% variance) was a better predictor of GGPA than GRE-Q. Case and Richardson (1990) found that GRE-V scores were more strongly correlated with GGPA than GRE-Q scores for students in Library and Information science program.

The different prediction of GRE-V and GRE-Q and the lack of correspondence with some of the previous studies are easy to explain if the discipline features are taken into account. In the discipline of engineering, students are usually required to have much knowledge and high abilities in Math, statistics, numerical logic, and some other advanced quantitative skills. The quantitative abilities appear more important than verbal abilities for engineering students. In contrast, for some other disciplines, like social science and humanities, however, the verbal abilities may be more critical than quantitative abilities. From this logic, GRE-Q would be likely to have a higher correlation with graduate performance than GRE-V for engineering students, but would be opposite for students in social science and humanity.

Research question 2: What are the differences in the validities of GRE scores between American students and international students?

Research question 3: What are the differences in the validities of GRE scores between master's and doctoral students?

The differences in the prediction of GRE scores between different groups of alien status and degree level were examined. Separate regressions were firstly conducted for different subgroups (i.e. American and international students, and master's and doctoral students) to test the validities for subgroups, and further statistical tests with group variable added into the regression model were also conducted to test the significance of the difference in the validity. Generally, GRE scores explained more variance in graduate performance as measured by grades for American students (the explained variance ranged from 10.5% to 22.2%) than for international students (the explained variance ranged from 1.4% to 5.5%). However, statistical tests did not indicate that the differences by alien status were significant (except one significant difference, i.e. GRE-Q predict GGPA total scores better for American students than for international students). In this study, because of the incomplete or missing data, the representativeness of the data used in the statistical test could not be tested. As indicated above, more than 1/3 of international students and more than 2/3 of American students in the sample did not have GRE scores, but the reason why these students did not have these scores was unknown. As a result, the number of American students who had GRE scores was much less than that of international students, and accordingly, much less American students than international students were in the regression models (see Appendix). In order to

determine whether the differences in the prediction by subgroups actually exist, further examination in the future is needed.

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. When averaging all previous studies where GRE was found to be a valid predictor of graduate performance, international students were usually a very small part of the student population. As Educational Testing Services stated (2008a), the data derived from samples with small number of minorities may not account for potential differences in cultures, linguistic background, and educational experiences for these populations. However, in the sample of the present study, international students constituted a considerable proportion (58.36%) of the total sample size. The results of the present study could have a good power of explanation for the difference between American students and international students.

Regarding degree level, GRE scores were found to predict the three criterion variables significantly for both master's and doctoral students. In predicting the school performance of master's and doctoral students, the patterns of the prediction by the GRE-V and GRE-Q were different as indicated by R^2 in the regression model. For master's students, GRE-V and GRE total explained larger variance than GRE-Q scores across 1st-year, 2nd-year, and total GGPA. GRE-V and GRE total scores explained 3.5% to 9% of the variance in the criteria, but GRE-Q explained the variance of 1.7% to 3.1 %. For doctoral students, by contrast, GRE-Q and GRE total scores better predicted 1st-year, 2nd-year, and total GGPA than GRE-V. GRE-Q scores explained the variance of 3.6% to 8.3%, but GRE-V only explained 0.7% to 1.7%.

Most GRE validation studies were conducted based on total graduate student population, without differentiating degree levels, so the findings of the present study also were not directly comparable to the previous studies. However, by comparing with some meta-analytic studies, the difference between master's and doctoral students in the present study was consistent with their findings. For example, ETS (Educational Testing Services, 2008a) examined the correlations between GRE General test and GGPA: for master-level students, $r = .32$ in Verbal section, and $r = .26$ in Quantitative section; while for doctoral level students, $r = .27$ in Verbal section, and $r = .30$ in Quantitative section. In Kuncel et al. (2010) meta-analysis, by integrating 100 studies in the meta-analysis, researchers found that GRE predicted well for both master's and doctoral students, but the differences ranged from small to zero. Though generally small to zero, it still showed discrepancies in the validity across master's and doctoral programs. Kuncel et al. (2010) 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 between masters' students and doctoral students, the further tests did not indicate that the differences by degree level were statistically significant. 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 (see Appendix). This range restriction may make it hard to test the difference even if the

difference existed. Study in the future is suggested to take the restriction of range into account.

Research question 4: How does UGPA predict engineering students' 1st-year, 2nd-year and total GPA in the graduate program? What is the incremental validity of GRE over UGPA?

The validity of UGPA was tested. It 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 previous findings about the predictive ability of UGPA that also found UGPA was a strong predictor of graduate performance (Kuncel et al., 2001; McKee, Mallory, & Campbell, 2001; Powers, 2004; Reisig & DeJong, 2005). After adding GRE scores, the total variance accounted for by the combination of GRE and UGPA was higher than that explained by UGPA alone, and the explained variance increased by 3.8% to 10%. This finding indicated that the GRE had significant incremental validity over UGPA in predicting graduate performance. This result corresponded with some previous studies which found that the combination of GRE scores and UGPA was fairly a robust predictor of academic performance (McKee, Mallory, & Campbell, 2001; Reisig & DeJong, 2005). McKee, Mallory, and Campbell (2001) found that UGPA alone explained about 24% of the variance in GGPA, and a total of 40% of variance was accounted for by the combination of UGPA and GRE. In addition, though GRE added more power in predicting graduate performance in the present study, however, if looking at the results of the general predictive ability of the GRE in the above analysis (i.e., GRE alone only explained 2.8% to 7.9% of the variance), it was easy to find that the variance explained

by UGPA was much larger than explained by GRE. Similar findings were found in some previous studies. Reisig and DeJong (2005) found that UGPA had a consistently stronger correlation with graduate student performance than GRE scores. Williams and Wise (2005) concluded that UGPA was a better predictor of GGPA than GRE for both master's and Ph.D students. Milner, McNeil and King (1984) also found UGPA accounted for more variance than GRE, with a variance of 9% explained by UGPA and 5.7% by GRE scores.

Research question 5: Specifically for international students, how does the language proficiency (measured by TOEFL) predict engineering students' 1st-year, 2nd-year and total GPA in the graduate program? What is the incremental validity of GRE over language proficiency?

For international students, their language proficiency was tested by TOEFL. In this study, TOEFL scores were significantly correlated with the 1st-year, 2nd-year, and total GGPA, and explained 2.2%, 3%, and 4.9% of the variance, respectively. The findings were similar to Wait and Gressel's (2009) study which found that there was a positive and statistically significant relationship between TOEFL scores and GPA for international students. In their study, increases in TOEFL scores were related to increases in passing rate of the comprehensive examination and the rate of graduation. Significant correlations were found between TOEFL scores and GGPA in this study, however, the absolute values of the correlation coefficients were small, and the explained variance was only 2.2% to 4.9%. This finding was similar to that of Cho and Bridgeman's study (2012) which found 3% of variance was explained by TOEFL scores. The possible reason of the

small correlations between TOEFL scores and GPA may be derived from Vinke and Jochems' (1993) explanation. Vinke and Jochems suggested that the predictive ability of TOEFL scores was good when TOEFL scores were in a wide range whereby an increase in the scores was related to an increase in the chance of academic success. However, when the range of scores is restricted, an improvement in TOEFL scores had nearly no effect on the academic performance. In graduate admission, a minimum requirement of TOEFL scores was set. Applicants with lower than minimum TOEFL scores were usually out of the consideration of admission, so the TOEFL scores of the students who were admitted might be possible over or outside of the range limit as mentioned above, and as a result, the effect of TOEFL scores based on sample statistics was lowered.

The incremental validity of GRE over TOEFL was also investigated. In hierarchical multiple regression analyses, after adding GRE scores, the explained variance in criterion variables increased by 1.4% to 5.3%, which indicated that GRE scores had a significant incremental predictive power over language proficiency in predicting graduate performance. GRE-Q and GRE total scores generally added more variances over TOEFL than GRE-V scores. The GRE, to some extent, can also be looked as a test of language proficiency besides only a test of general cognitive abilities. However, there were no previous studies testing the incremental predictive ability of GRE scores over language proficiency, at least within the extent of the literature review of this study. So it was not available to make any comparisons with any previous studies, but this study provides evidence in this regards. As it is known, GRE measures verbal and quantitative reasoning skills and analytical writing abilities. GRE test takers need to solve problems, synthesize information, and reason complex relationships between pieces

of information. From the results of this study, it was found that the GRE had significant incremental predictive ability over TOEFL, which confirmed that GRE was more than a test of language skills. The GRE explained a significant 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 the TOEFL. This difference can be explained by the specific purpose and content of each subtest. The GRE-V measures the ability to analyze and evaluate written material and synthesize information, and to analyze relationships among component parts of sentences and recognize relationships among words and concepts. The GRE-Q tests problem-solving ability, focusing on basic concepts of arithmetic, algebra, geometry and data analysis. It is clear that GRE-V is more related to language skills, and thus the GRE-V was found to add less power in predicting graduate performance over TOEFL compared to the GRE-Q.

Limitations 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. The findings of this study explained the phenomena and conclusions of engineering, so the implications can only be directly applied to this specific discipline. Generalization of the findings of this study to other disciplines should be with cautions and careful considerations. On the other hand, focusing on the discipline of engineering can draw accurate conclusions and implications for this specific discipline. The findings will have more power in explanation of the validity of GRE for engineering students.

The second limitation concerns the missing data in this study. Though students who had complete data of the three criterion variables were retained in the database,

many of them did not have complete data on all the three predictors, and the reasons why data were lost were unknown. The method of Listwise deletion was applied in the data analysis, but the representativeness of the remaining data could not be tested. Since the overlapped students (i.e., those students who had data of two predictor variables) differed in each research question, the results of each research question were based on different groups of students. Thus, the results should be interpreted with care and critical thinking. For example, the UGPA in this study was only available for students who received their bachelor's degree from the same university as their graduate school, and almost all of these students were American students, and most of these students did not have GRE scores. So when using SLR to obtain the validity of UGPA and using MR to analyze the incremental validity of GRE over UGPA, the corresponding groups of students were different. Only students who had both UGPA and GRE scores were included in the analysis of the incremental validity of GRE over UGPA, but the representativeness of this group of students was not guaranteed. However, this problem resulted from the real situation and reflected the reality of the database in which the university kept in their records, so it was not in control of the researcher. Though the missing data was a big problem, the large sample size helped deal with the flaws, with the smallest group sample size of 65 ($N = 65$, in MR) when testing the incremental validity of GRE over UGPA.

Another limitation is that the relationships between the predictors and the criteria were not corrected for range restriction. 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, and thus the correction of range restriction was not possible.

Implications and Recommendations

This study has some implications for the graduate admission decisions for engineering programs. Recommendations for future research are also suggested. First, this study confirmed the usefulness of the GRE in predicting graduate performance, as measured by 1st-year, 2nd-year, and total GGPA, so the GRE is still suggested as a good predictor to be employed in the graduate admission. 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. As found in this study, the GRE scores of American students and international student differed significantly, but their graduate performance, indicted by GGPA, did not found to be significantly different. It can be argued that though some international students have lower GRE scores than American students, yet they seem to earn the same level of graduate performance as American students. There are probably some other factors that are more important than GRE scores in predicting international students' success, such as education background, motivation, perseverance, and hard-working, and so on. These factors probably should be considered as much as possible in different ways, such as personal statements, letters of recommendation, interviews, and assessment of personality. In addition, the difference by degree level suggests that the differences in career orientations and career goals between

master's and doctoral students should also be taken into consideration. Second, since GRE-Q scores were found to have higher predictive power than GRE-V scores for engineering students, it is suggested that GRE-Q be given more weight than GRE-V in deciding which applicants to be admitted into engineering programs. Third, because of a large proportion of variance was found to be explained by UGPA, undergraduate academic performance or educational background should be given a large weight in deciding whether a student should be admitted or not, and UGPA and GRE should be considered together in the admission decisions so as to better predict students' performances during graduate studies.

A great amount of variance in the criterion variables remains unexplained by the predictors in this study. There may be room and a need to conduct more research to study the unexplained portion of variance in graduate performance. In this study, graduate performance was examined only in the form of GGPA. As suggested by Kuncel et al. (2010), multiple aspects of student performance should be considered so as to have a more comprehensive picture about students' performance. The criteria may include the information like faculty ratings, degree attainment, degree completion, and research productivity. Moreover, as to the admission criteria, committees should admit students based on not only cognitive abilities but also some noncognitive characteristics of the applications, such as motivation, interest, personality, and some other characteristics.

Conclusions

As stated by Bachman and Palmer (1996), one of the most important considerations in designing and developing an assessment is the use for which the assessment is intended, and the effectiveness of the assessment in achieving its purpose

determines its usefulness. The GRE is a set of standardized tests designed to determine the scholastic potential of graduate students, and it is widely used in graduate admission decisions by many universities and institutions in the United States. How well GRE scores can predict students' graduate performance is a crucial factor to use to evaluate the effectiveness of the GRE. This study investigated the predictive ability of GRE scores in predicting graduate performance for students from a specific discipline of engineering at a large midwestern university. In general, GRE was found to be a useful predictor in predicting 1st-year, 2nd-year, and total GGPA. GRE-V and GRE-Q scores had a different pattern in predicting graduate grades of masters' and doctoral students. GRE-V and GRE-Q scores explained more variance in graduate performance for American students than for international students, but no statistically significant differences were found except when GRE-Q predicted GGPA total scores. UGPA was found to be a strong predictor, and TOEFL scores were also significantly correlated with GGPA, but GRE scores still have significant incremental validity over both UGPA and TOEFL scores in predicting graduate grades. These findings have some implications for graduate admission decisions, and can suggest directions of future research.

REFERENCES

- Alderson, J. C. (2009). Test review: Test of English as a Foreign Language: Internet-based Test (TOEFL iBT). *Language Testing*, 26(4), 621–631.
- Ayers, J. B., & Quattlebaum, R. F. (1992). Toefl Performance and Success in a Masters Program in Engineering. *Educational and Psychological Measurement*, 52, 973-975.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bell, N.E. (2008). *Graduate enrollment and degrees: 1997 to 2007*. Washington, DC: Council of Graduate Schools.
- Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education*, 22, 109-119.
- Case, D. O., & Richardson, J. V. (1990). Predictors of student performance with emphasis on gender and ethnic determinants. *Journal of Education for Library & Information Science*, 30(3), 163-182.
- 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 (6), 951-961.
- Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, 29(3), 421-442.
- 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_Table_s.pdf.

Educational Testing Services. (2008a). Graduate Record Examination: Guide to the use of GRE scores 2008-2009.

Educational Testing Services. (2013). Graduate Record Examination: Guide to the use of GRE scores 2013-2014.

Enright, M. K., & Gitomer, D. (1989). *Toward a description of successful graduate students* (GRE Board Research Report 85-17R). Princeton, NJ: Educational Testing Service.

Feeley, T. H., Williams, V. M., & Wise, T. J. (2005). Testing the predictive validity of the GRE Exam on communication graduate student success: A case study at university at Buffalo. *Communication Quarterly*, 53(2), 229-245.

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(2), 336-348.

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.

Graduate Records Examination Board. (2003). *Guide to the use of scores*, Education Testing Services.

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.

- 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.
- Hyun, M. S. (2012). *Examining the validity of the GRE General Test scores and undergraduate GPA for predicting success in graduate school at a large racially and ethnically diverse public university in Southeast Florida*. FIU Electronic Theses and Dissertations, Paper 797.
- 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.
- 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(1), 51–68.
- 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(1), 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(2), 340-352.
- McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology*, 79, 493-505.

- 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.
- 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.
- 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.
- Mupinga, E. E., & Mupinga, D. M. (2005). Perceptions of international students toward GRE. *College Student Journal, 39*(2), 402-408.
- National Academy of Engineering. (2005). *Educating the engineer of 2020: Adapting engineering education to the new century*. Washington, DC: The National Academies Press.
- Norcross, J. C, Hanych, J. M., & Terranova, R. D. (1996). Graduate study in psychology: 1992-1993. *American Psychologist, 51*, 631-643.
- Oldfield, K., & Hutchinson, J. R. (1997). Predictive validity of the Graduate Record Examination with and without range restraints. *Psychological Reports, 81*, 211-220.
- Pennock-Roman, M. (2002). Relative effects of English proficiency on general admissions tests versus subject tests. *Research in Higher Education, 43*, 601-623.

- 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)
- 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(2), 208-219.
- 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(1), 37-59.
- Rogers, S. W., & Goktas, R. K. (2010). Exploring engineering graduate student research proficiency with student surveys. *Journal of Engineering Education*, July, 263-278.
- 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.
- Schmidt, E. A., Homeyer, L. E., & Walker, J. L. (2009). Predictors of success on the preparation comprehensive examination. *Counselor Education & Supervision*, 48, 226-238.
- 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.

- 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*(2), 335-349.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613-629.
- Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists. *American Psychologist, 52*(6), 630.
- Stricker, L. J., & Rock, D. A. (1995). Examinee background characteristics and GRE general test performance, *Intelligence, 21*, 49-67.
- 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*(2), 146-173.
- Thomell, 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.
- Vinke, A. A., & Jochems, W. M. G. (1993). English proficiency and academic success in international postgraduate education. *Higher Education, 26* (3), 275-285.
- Wait, I. W., & Gressel, J. W. (2009). Relationship between TOEFL score and academic success for international engineering students. *Journal of Engineering Education, (October)*, 389-398.
- Young, M. D., & Brooks, J. S. (2008). Supporting graduate students of color in educational administration preparation programs: Faculty perspectives on best

practices, possibilities, and problems. *Educational Administration Quarterly*, 44, 391-423.

APPENDIX

Scatter Plots Showing the Relations between GGPA and GRE

The following scatter plots show the two-dimensional relations between GGAP (GGPA-1, GGPA-2, GGPA-tot, respectively) and GRE scores (GRE-V, GRE-Q, GRE-tot, respectively), separated by alien status and degree level.

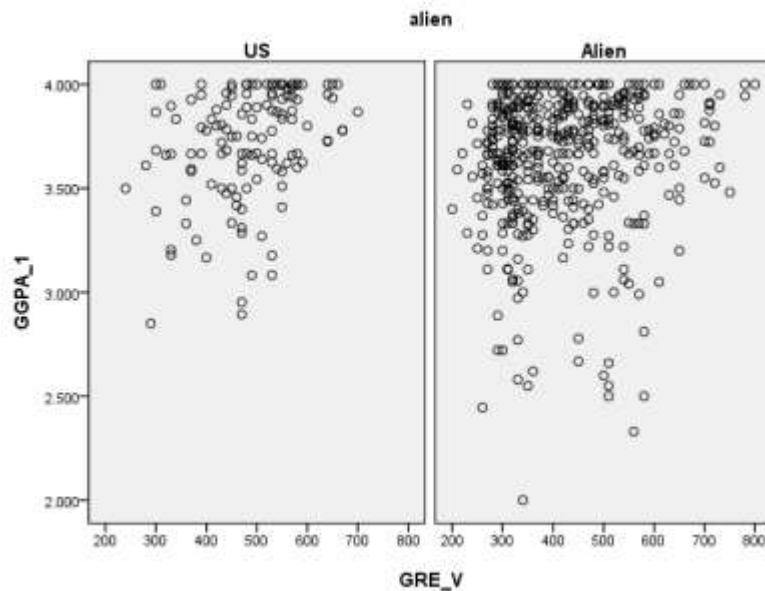


Figure 1. Scatter Plots between GRE-V and GGPA-1 Separated by Alien Status
Note: US: American students; Alien: International students. (the same for the followings)

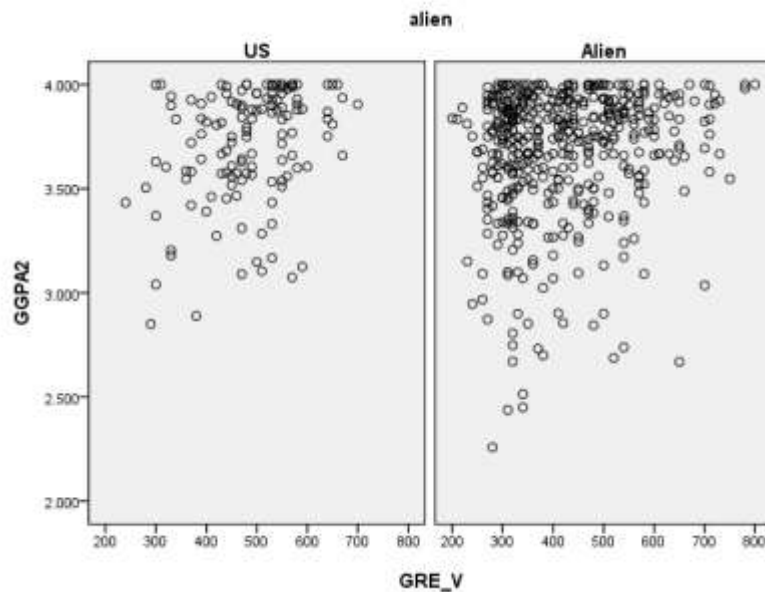


Figure 2. Scatter Plots between GRE-V and GGPA-2 Separated by Alien Status

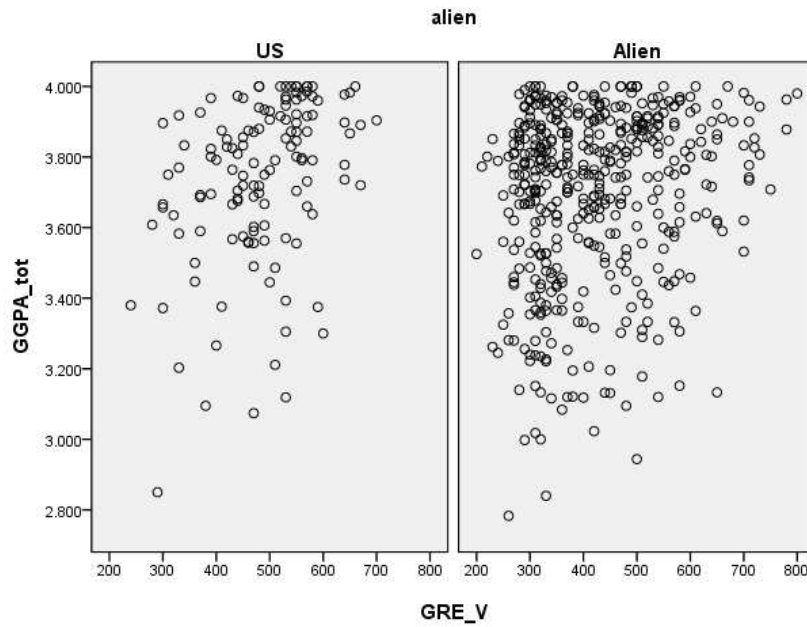


Figure 3. Scatter Plots between GRE-V and GGPA-tot Separated by Alien Status

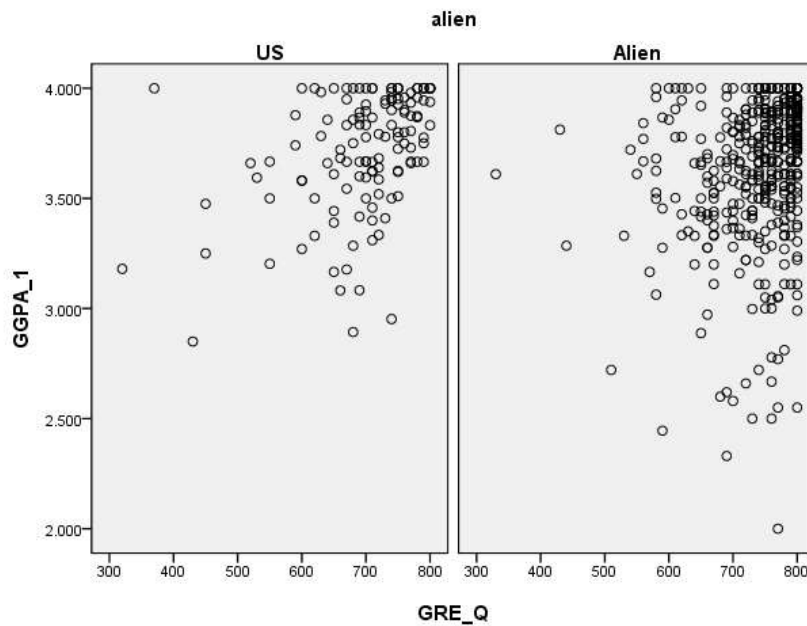


Figure 4. Scatter Plots between GRE-Q and GGPA-1 Separated by Alien Status

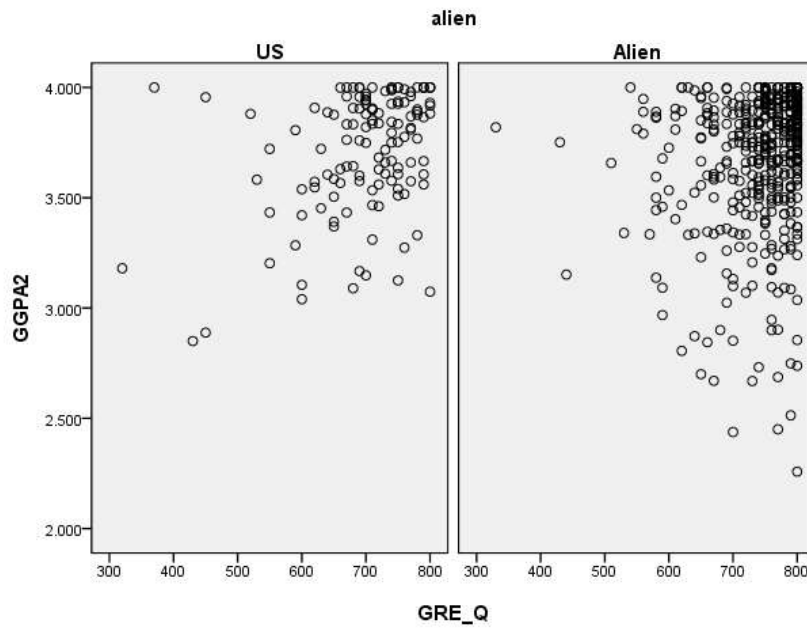


Figure 5. Scatter Plots between GRE-Q and GGPA-2 Separated by Alien Status

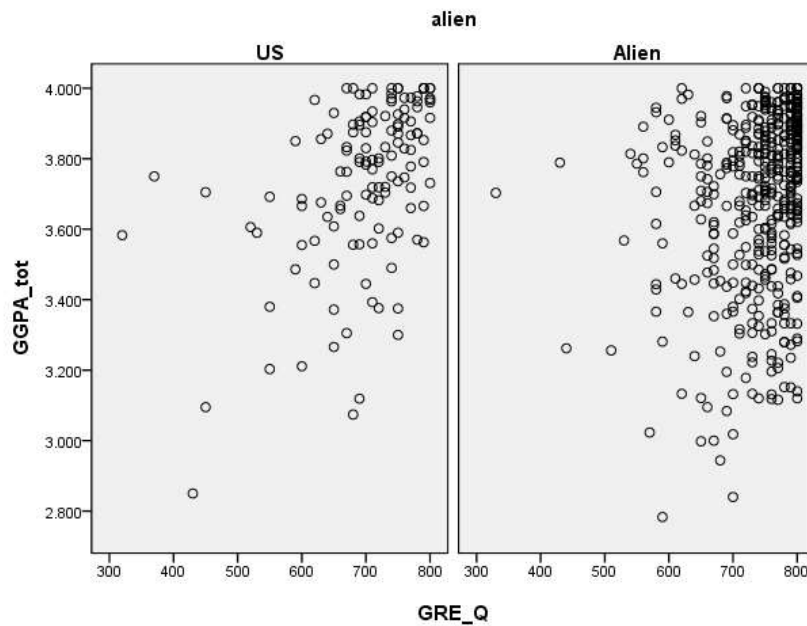


Figure 6. Scatter Plots between GRE-Q and GGPA-tot Separated by Alien Status

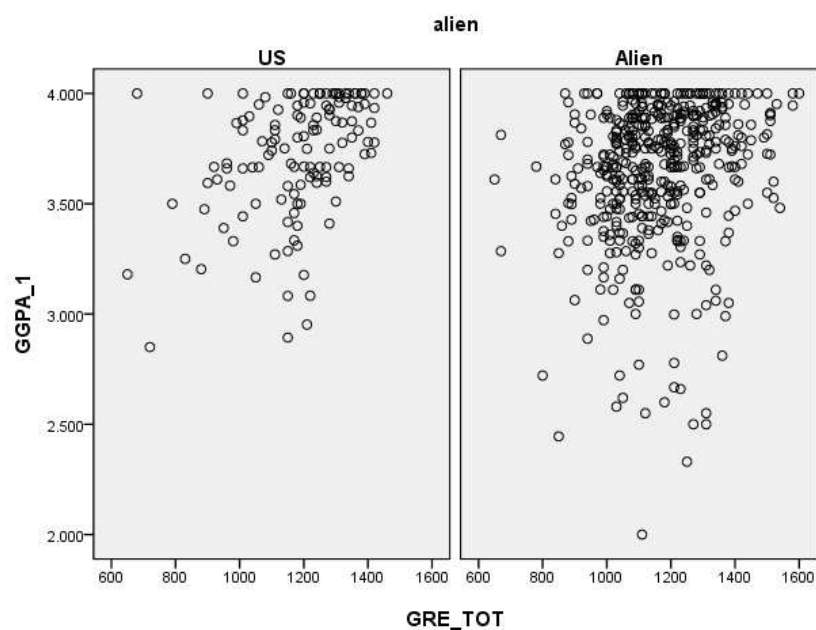


Figure 7. Scatter Plots between GRE-tot and GGPA-1 Separated by Alien Status

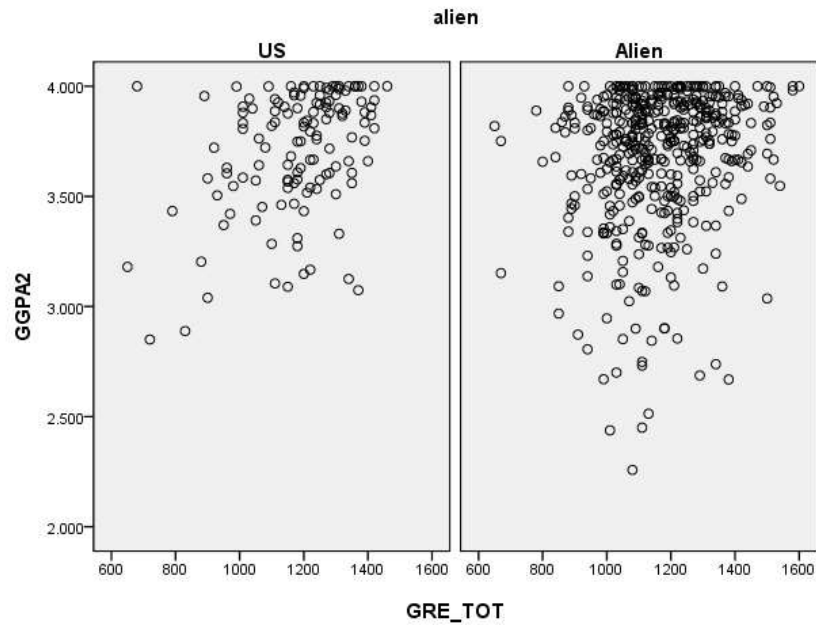


Figure 8. Scatter Plots between GRE-tot and GGPA-2 Separated by Alien Status

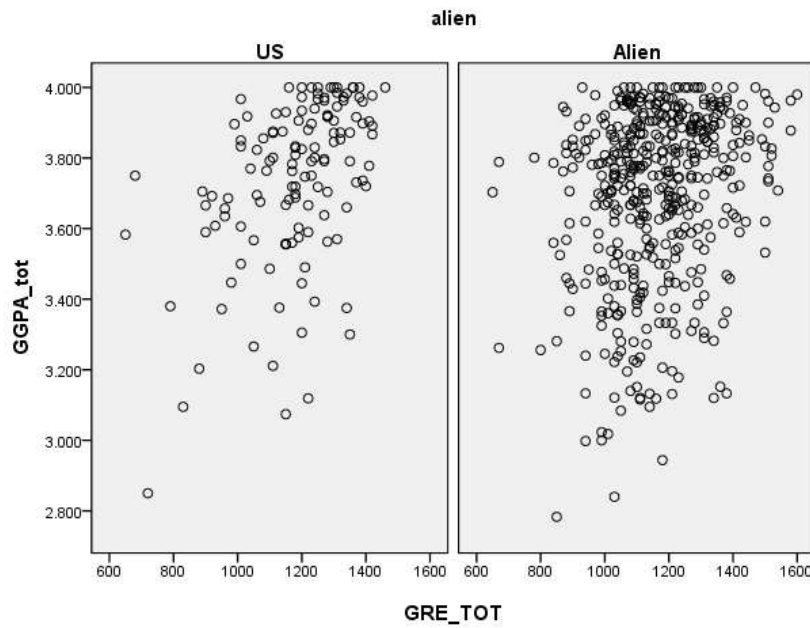


Figure 9. Scatter Plots between GRE-tot and GGPA-tot Separated by Alien Status

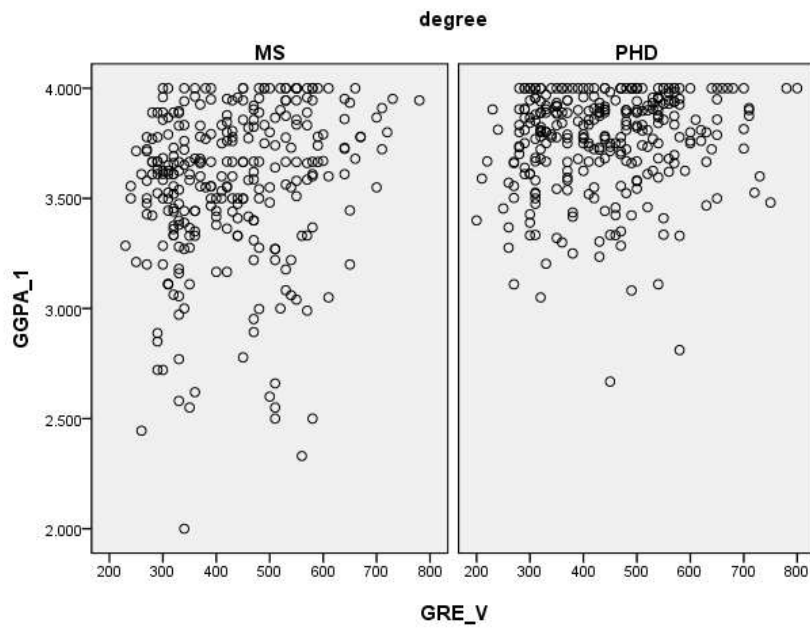


Figure 10. Scatter Plots between GRE-V and GGPA-1 Separated by Degree Level

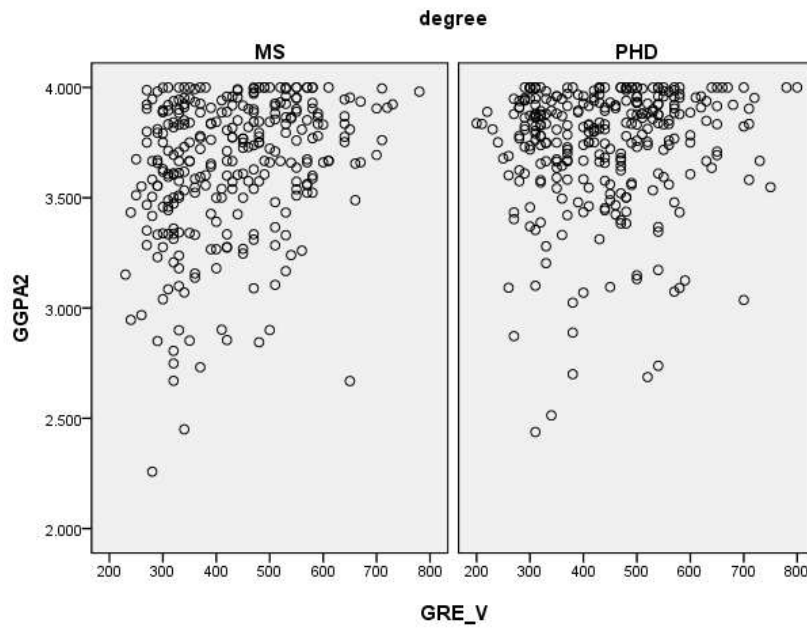


Figure 11. Scatter Plots between GRE-V and GGPA-2 Separated by Degree Level

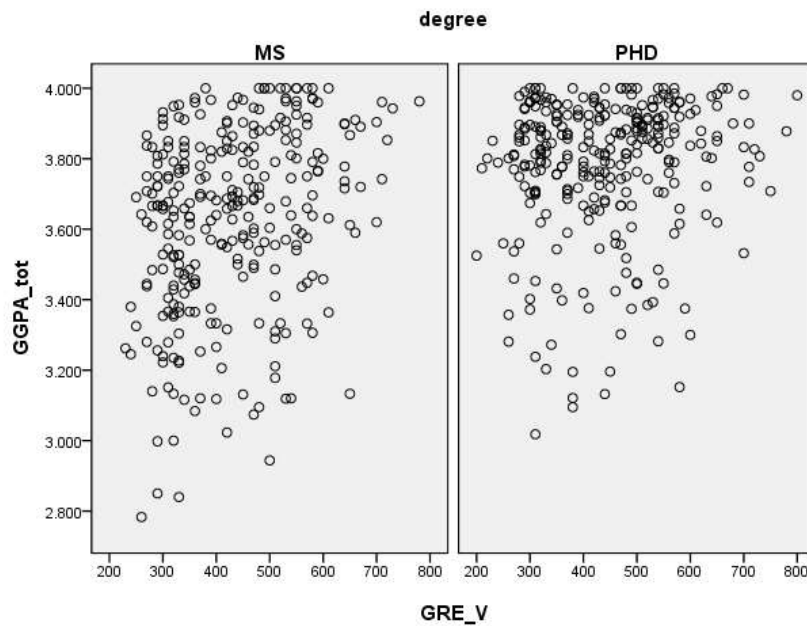


Figure 12. Scatter Plots between GRE-V and GGPA-tot Separated by Degree Level

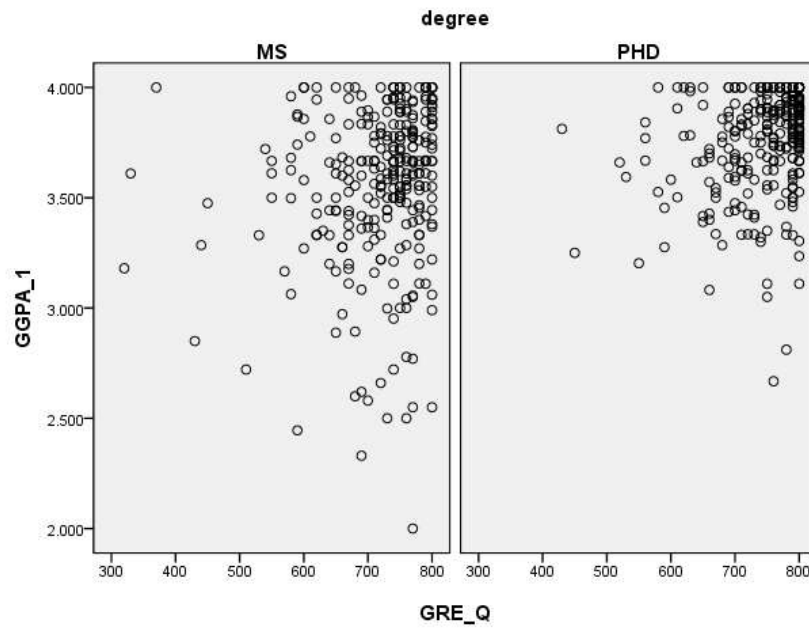


Figure 13. Scatter Plots between GRE-Q and GGPA-1 Separated by Degree Level

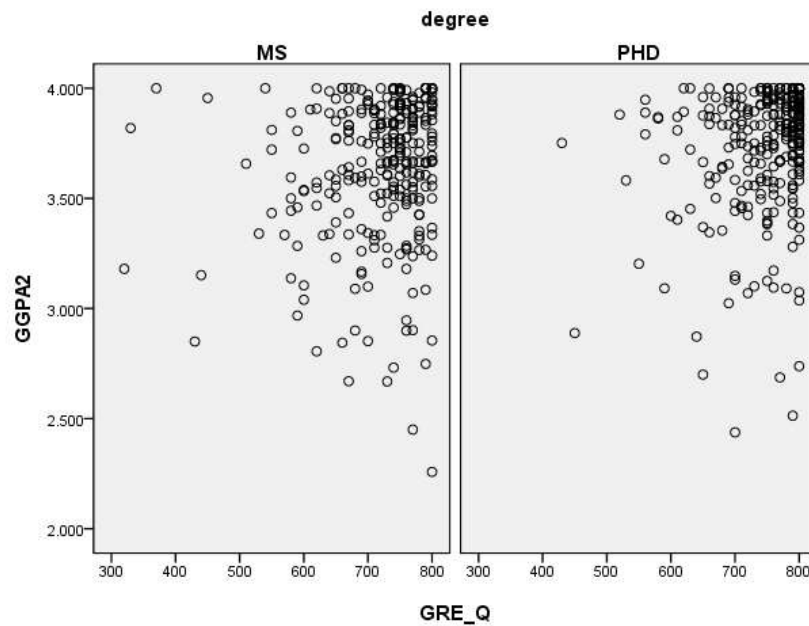


Figure 14. Scatter Plots between GRE-Q and GGPA-2 Separated by Degree Level

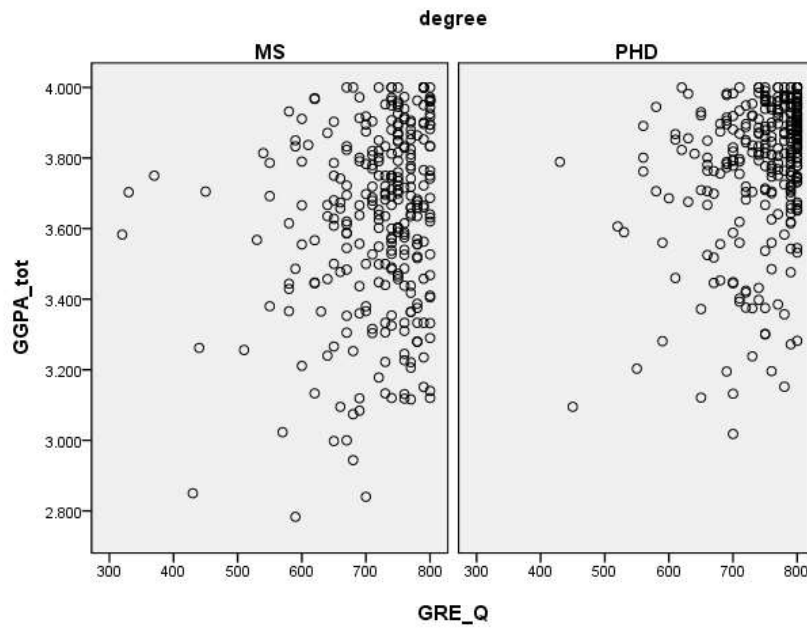


Figure 15. Scatter Plots between GRE-Q and GGPA-tot Separated by Degree Level

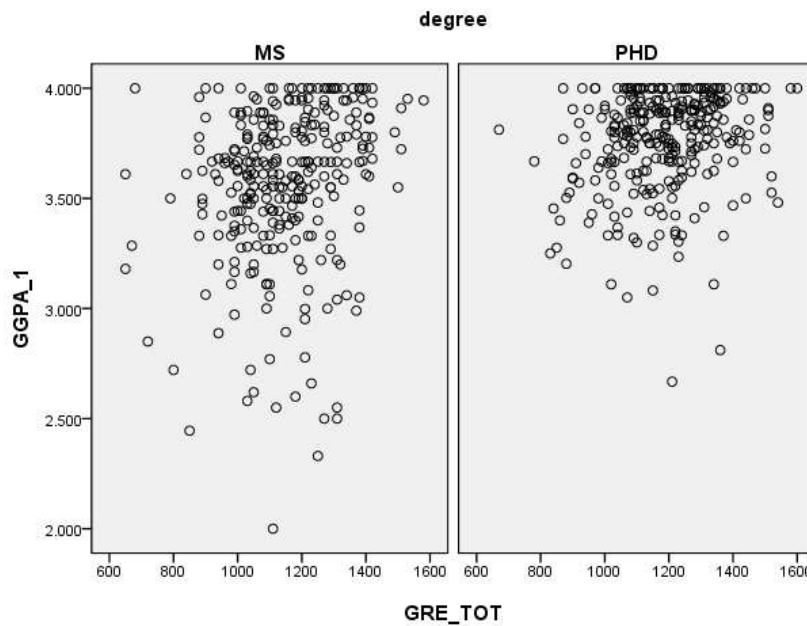


Figure 16. Scatter Plots between GRE-tot and GGPA-1 Separated by Degree Level

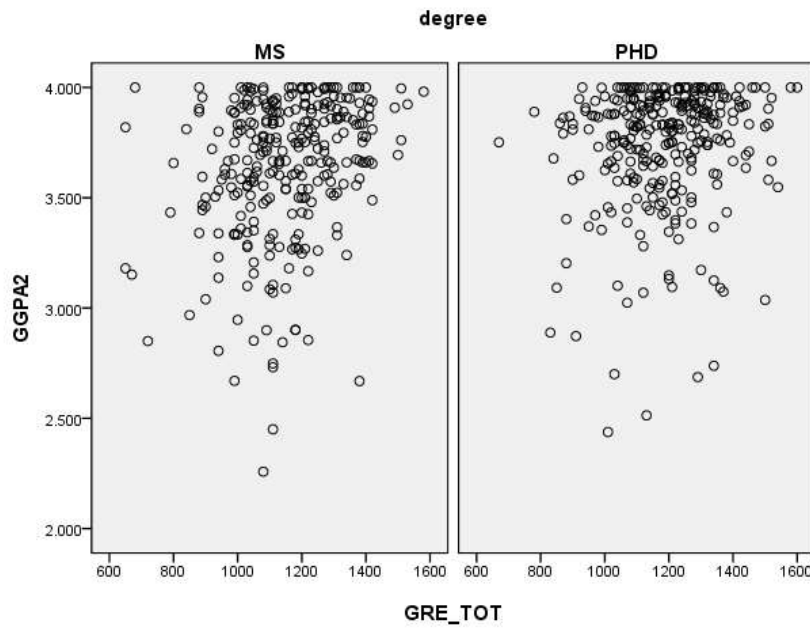


Figure 17. Scatter Plots between GRE-tot and GGPA-2 Separated by Degree Level

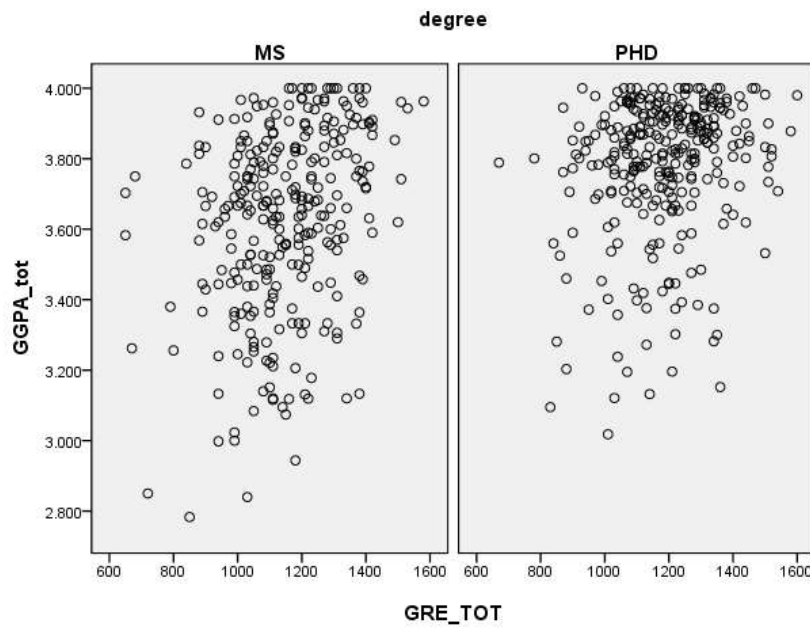


Figure 18. Scatter Plots between GRE-tot and GGPA-tot Separated by Degree Level