GMAC®

Graduate Management Admission Council®
Creating Access to Graduate Business Education℠

GMAT® Validity for Six European MBA Programs

Eileen Talento-Miller

GMAC® Research Reports • RR-05-07 • October 5, 2005

Introduction

Several members of the European Union recently signed the Bologna Accord, which heralds significant changes to the higher education systems in these countries (Loades, 2005). As a result of some of the changes, it is expected that there will be increased interest in graduate education, including graduate business education. In anticipation of these educational changes in Europe, a study was designed to evaluate the effectiveness of the GMAT® examination for admission to graduate management programs in Western Europe.

This study explores the validity of GMAT® scores for predicting performance in six graduate management programs in Western Europe. The observed validity values are compared to the population of existing validity study results to determine if prediction for Western European programs is meaningfully different from results for US programs. Because the investigated programs are considerably diverse, in addition to studying validity for the Western European programs individually and in combination, this investigation also explores the validity of the GMAT® in Western European programs for students with differing backgrounds, including native language and citizenship. Analyses are performed separately by gender, previous educational experience, language group, and citizenship group—analyses similar to those conducted in an earlier European validity study by Crooks and Heuvelmans (1999).

Related Literature

Previous research has found that GMAT® scores are good predictors of performance in graduate management programs (e.g. Hecht & Schrader, 1986; Kuncel, Crede, & Thomas, 2004; Olsen, 1957; Sireci & Talento-Miller, in press), though few studies have looked specifically at predictive validity for programs outside the United States. A study by Dobson, Krapljan-Barr, and Vielba (1999) analyzed the data for one school in the U.K. and found that GMAT® verbal scores were good predictors of course exam performance, but GMAT® quantitative scores were not. This study also found differences based on whether students were native English speakers. The surprising result was that non-native English speakers performed worse on their course exams than would have been predicted based on their GMAT® scores (Dobson et al., 1999), which suggests that using the admission test does not disadvantage this group. Koys (2005) recently conducted a validity study for programs in three different countries, finding good predictive validity for GMAT® scores of students in these programs. All three programs were run by a U.S. business school, however, and were not limited to European countries. A study by Crooks and Heuvelmans (1999) specifically looked at the validity of multiple schools in Western Europe. Their findings were that validity of GMAT® scores for these programs was similar to the average of the validity studies conducted for U.S. and other programs. The study also examined specific groups within the sample and concluded that there was no bias in prediction by nationality, age, mother tongue, gender, or previous educational study. Because of the date the study was conducted, the Crooks and Heuvelmans (1999) validity study of European programs does not include any information about the Analytical Writing Assessment (AWA) portion of the GMAT® exam, nor does it address scores received since the GMAT® exam was converted to a computer-adaptive test.

To assess validity of GMAT® scores for different programs around the world, the Graduate Management Admission Council® (GMAC®) conducts validity studies free for any school that provides data. This Validity Study Service (VSS) has been available for decades and has collected data on hundreds of programs. Based on the data from recent studies, the validity of GMAT® scores can be compared across US programs and programs outside of the United States. Table 1 shows the results for several comparisons. Simple and multiple correlations were

corrected for restriction of range based on the ratio of the variance in the school sample to the variance in the population of applicants.[1] Results from these previous studies suggest that predictive validity from GMAT® scores is higher on average for non-US programs. Interestingly, GMAT® AWA scores appear to have higher average validity and GMAT® Quantitative scores appear to have lower average validity for the non-US programs. The small number of non-US programs studied and the appearance of moderate differences by program location underscores the need for additional information in this area.

Table 1. Comparison of Validity for US and Non-US Program VSS Studies

	N	GMAT® Total	GMAT® Verbal	GMAT® Quant	GMAT® AWA	VQA*
US: Mean (SD)	253	0.458 (0.14)	0.320 (0.14)	0.338 (0.16)	0.179 (0.13)	0.470 (0.13)
Non-US: Mean (SD)	14	0.484 (0.18)	0.337 (0.16)	0.288 (0.19)	0.257 (0.16)	0.516 (0.13)
US-NonUS: Effect Size**	267	-0.176	-0.119	0.305	-0.566	-0.344

* VQA is the multiple correlation of GMAT® Verbal, GMAT® Quantitative and GMAT® AWA.

** Effect size is calculated as the mean difference divided by the pooled standard deviation.

Methods

Individual programs were invited to participate in a special study examining the predictive validity of the GMAT® exam for European programs. Programs were identified by examining the number of GMAT® scores sent to each school. The programs were asked to provide the relevant data for three years or three classes of students. In the early phase of data collection, six schools submitted data for a total of 1,241 students, which formed the basis for this preliminary investigation.

Predictive validity is calculated through the relationship between the test scores that are used to predict performance and the performance outcome of grades in the program. These relationships are captured through simple and multiple correlations that are corrected for restriction of range. Because of the admission process, data is only available for a select group of applicants. Correcting for restriction of range yields results that are applicable to the entire applicant population, not just the pool of selected candidates. The corrections for the individual programs are based on the comparison of test scores for the students in the program versus the scores of all applicants to that program. The predictive validity values for overall GPA obtained for each of the six programs was compared to the average of results calculated across the 253 validity studies conducted for US programs through the GMAC® VSS between 1997 and 2005. Effect size in terms of standard deviation differences were used to compare results because this statistic is not affected by sample size. These analyses represent potential differences at the school level. A different question would be addressed by examining results at the individual student level by aggregating data from all schools. This student-level analysis also allows for comparisons of student subgroups.

The standard data for all schools was combined to form the basis for subsequent analyses. Since different grading scales existed at each of the schools, all GPA values were standardized within school. To account for school differences, school membership was entered as dummy variables for all multiple regression analyses. Correlations were corrected for restriction of range using a comparison

[1] Potential applicants were defined as any GMAT® examinee who sent scores to the program studied. Restriction of range corrections were based on the Hunter and Schmidt formula:

$$r_{ij}^{*} = \frac{U\, r_{ij}}{\sqrt{(U^2 - 1)\, r_{ij}^{\,2} + 1}}$$

where r_{ij}^{*} is the adjusted bivariate correlation of variables i and j,

r_{ij} is the observed bivariate correlation, and

$$U = \frac{\sigma_{pop}}{\sigma_{obs}}.$$

of the standard deviation of GMAT® scores for the sample of 1,241 students versus the population standard deviation. The corrected simple correlations were used to calculate the multiple regression. Predictive validity was calculated for overall GPA using the sample of all students. Analyses were then conducted for subgroups of the population to determine if differences existed.

Results

Table 2 summarizes the results of the school-level analysis for each of the six programs. Predictive validity in the 0.3-0.4 range is generally considered good for admission tests (Kaplan & Sacuzzo, 1997). When validity is adjusted for restriction of range, good values are likely to be above 0.4 (Talento-Miller & Rudner, 2005). It is clear that GMAT® scores are a good tool for admission to these programs. Table 3 shows how the average validity of these programs compares to the average validity calculated across 253 recent VSS studies for US programs. This is the same data that is compared in Table 1. Because of differences in programs and differences in students, variation is expected in the amount of validity observed. To insure that there were no differences in the amount of variation between the European and US studies, Hartley's test for homogeneity of variance was conducted. The results showed no significant differences in variation (all $p > 0.05$).

Compared to the average validity for US programs, these programs show higher validity for most of the GMAT® variables. Effect size was calculated as the difference between the mean European program validity and the mean VSS study validity divided by the VSS standard deviation. Guidelines for interpreting effect size suggest that values around 0.2 would be considered a small effect, those around 0.5 would be considered medium, and those around 0.8 would be large (Cohen, 1969). The overall effects using GMAT® Total scores or the combination of Verbal, Quantitative, and AWA scores are small, but positive for Western European programs compared to all programs. Looking at the separate scores, however, provides much more information about differences. These six European programs benefit from better prediction from both GMAT® Verbal and AWA scores compared to all programs, but have quite a bit less of an effect from GMAT® Quantitative scores. In all but one school, both the GMAT® Verbal and AWA scores had higher predictive validity than GMAT® Quantitative scores in predicting overall GPA. However, for the one school that was the exception, the prediction from Quantitative scores was substantial.

Table 2. Predictive Validity for Overall GPA by Program

	N	GMAT® Total	GMAT® Verbal	GMAT® Quant	GMAT® AWA	VQA
A	380	0.436	0.385	0.048	0.353	0.476
B	100	0.555	0.513	0.077	0.389	0.570
C	269	0.352	0.243	0.095	0.111	0.300
D	265	0.456	0.509	-0.170	0.156	0.529
E	100	0.403	0.477	-0.122	0.467	0.548
F	127	0.681	0.237	0.582	0.021	0.680

Table 3. Comparison of Validity of European Programs to VSS Studies of US Programs						
	N	GMAT® Total	GMAT® Verbal	GMAT® Quant	GMAT® AWA	VQA
European: Mean (SD)	6	0.481 (0.12)	0.394 (0.13)	0.085 (0.27)	0.249 (0.18)	0.517 (0.13)
US VSS: Mean (SD)	253	0.458 (0.14)	0.320 (0.14)	0.338 (0.16)	0.179 (0.13)	0.470 (0.13)
EU-US: Effect Size	259	0.165	0.529	-1.554	0.534	0.362

When using the student as the unit of analysis, the combination of GMAT® Verbal, Quantitative, and AWA scores results in good predictive validity for overall GPA, with the results suggesting that these variables can explain over one-fifth of the variance in grades for these European programs. Table 4 shows the simple correlations for each of the GMAT® variables with overall GPA and the relative contributions of the three scores when used in combination. GMAT® Verbal scores contributed most to prediction. Although the simple correlation of GMAT® AWA with grades might have suggested more of an effect, its impact in the combination with the other GMAT® scores was lessened due to their interrelationship, particularly the strong association with GMAT® Verbal scores. When the student data was analyzed, GMAT® Quantitative scores showed more predictive power than observed in the summary across individual programs, though the effect was still small. This result shows that each of the three GMAT® section scores can contribute uniquely to predicting successful performance in these programs.

Table 4. Predictive Validity for Overall GPA for Students		
	Correlation with GPA	Contribution to Prediction*
GMAT® Total	0.468	-
VQA	0.467	-
Verbal	0.364	63%
Quant	0.120	13%
AWA	0.266	24%
	Total	100%

* The contributions of variables in the prediction were calculated using the Pratt Index (Johnson & LeBreton, 2004) with

$$PI_i = r_i^* \beta_i / R^2$$

where r_{ij}^* is the adjusted bivariate correlation of predictor variable i with criterion variable j;
β_i is the standardized beta weight for variable I; and
R^2 is the squared multiple correlation of the set of variables with j.

In general, the results for the student dataset were similar to the summary of results across the six programs. Comparisons of the correlations for these analyses to the population of VSS studies results yielded similar findings, with medium positive effect sizes for GMAT® Verbal and AWA scores and a large negative effect size for GMAT® Quantitative scores. Because the results were similar, it appeared that error due to combining the data was minimal, so examining groups within the dataset was feasible.

Results by Group

Analyses were conducted for student groups by gender, undergraduate major, mother tongue, and nationality. Examples of categories within each group are listed in the appendix. The results are summarized in Table 5. Each of the groups had a large number of cases with all sample sizes greater than 100. To determine if there were differences, the predictive validity was examined for differences in magnitude and in the contributions of variables. For the most part, the predictive validity of the GMAT® scores for groups was similar to or better than the predictive validity for all students. A notable difference was the group of 'American nationalities', which included students from North, South, and Latin American countries. Upon examination, it became apparent that this group had the least variation in overall grades, which most likely contributed to the lower observed validity.

Table 5. Prediction of Overall GPA by Group

Group	N	Validity for VQA	Average residual
Gender			
Male	832	0.458	-0.055
Female	231	0.550	0.149
Major			
Business	358	0.425	0.030
Engineering	429	0.467	-0.140
Other	169	0.542	0.220
Language			
English	469	0.539	0.004
Western European	363	0.593	-0.214
Other	247	0.512	0.253
Nationality			
Western European	465	0.669	-0.248
Eastern European	121	0.504	0.107
Americas	253	0.157	0.080
Asia Pacific	242	0.679	0.285

To further examine group differences, the residuals from the overall model were examined for each group. The residuals are the difference between the predicted GPA and the actual observed GPA. Since the model is based on all students, the average residual can indicate whether predictions are equivalent across groups. Both the grades and the residuals are standardized, so the residuals are a measure of effect size. None of the effects for the groups were large or even medium effects. The largest effect was little more than a quarter of a standard deviation, which is not enough to make a practical difference between the predicted and the observed GPA. Because the residual is

predicted minus actual GPA, negative values mean that group, on average, performed better than expected from a model, while positive values mean that group did not perform as well as expected. With this in mind, it is apparent that both the 'Other language' group and the 'Asia Pacific nationality' group did not perform as well as expected based on their test scores, whereas the 'Western European nationality' group performed better than expected. When evaluated separately, the American nationality group had the lowest validity, but the small average residual for this group did not indicate that this group was disadvantaged by prediction using the test scores.

The amount of validity and the size and direction of the residuals provided some information indicating differences between the groups. More differences can be found when examining the relative importance of the variables for each group as calculated by the Pratt Index (Johnson & LeBreton, 2004). Results for school-level and student-level data showed the importance of GMAT® Verbal and AWA scores with less of an impact from GMAT® Quantitative scores. While that pattern was consistent for most groups, a few groups showed notable differences. The two groups with the largest contributions from GMAT® Quantitative scores were the related groups of Western European languages, with Quantitative scores contributing 58% to the prediction, and the Western European nationalities, where Quantitative scores contributed 71% to the prediction.

Due to the findings for the Western European language and nationality groups, these groups were examined further. The larger validity values could not be explained by greater variation in either the predictors or criterion, nor were there particularly notable differences in the means of the predictors. The overall GPA means for these two groups followed the outcome suggested by the residuals, which is that these two groups performed better than their comparison groups. All the information taken together suggest that admission factors for Western European students may not follow the same model as for other students seeking to enter Western European programs.

Summary

The GMAT® exam was shown to have good validity for predicting performance in graduate management programs taught at a select group of Western European schools. Student-level and school-level results suggested that GMAT® Verbal scores were particularly useful in predicting performance. While the GMAT® AWA section is a relatively new addition to the test, the usefulness of the AWA scores is readily apparent for these schools. The validity from GMAT® Quantitative scores is lower for five of the six investigated programs compared to results across all validity studies, and there is relatively little contribution from this variable when the variables are combined. This is consistent with previous results for non-US programs. Further study could determine how program differences or differences in admission contribute to these observed differences in validity.

Group analyses show only a few differences. Across the methods used to compare the groups, there were clear differences for the two related Western European groups. The findings suggest that admission factors do not predict the same for these groups as for others. The higher grades for students of Western European nationalities suggest there may be identifiable reasons affecting the performance for these students that also might inform good practices for other students.

Although this study analyzed a large number of students overall, which also allowed relatively robust group comparisons, there were only a few European schools that participated and the sampling of schools is not likely to be representative of all European schools with graduate management programs that use GMAT® scores in admission. Adding to this dataset would make the results more generalizable and could expand on the group findings, particularly for the Western European language and nationality groups. Further study may also look at the impact of additional variables such as age and TOEFL scores.

Contact Information

For questions or comments regarding study findings, methodology or data, please contact the GMAC Research and Development department at research@gmac.com.

Acknowledgements

Lawrence M. Rudner, GMAC®, for comments on earlier drafts of this report and Veronica Garcia, Research Writer/Editor, GMAC®, for writing and editorial services.

References

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Crooks, S. & Heuvelmans, A. (1999). The GMAT® as a predictor of academic performance on eight European MBA programs. Arnhem, The Netherlands: CITO.

Dobson, P., Krapljan-Barr, P., & Vielba, C. (1999). An evaluation of the validity and fairness of the Graduate Management Admission Test® (GMAT®) used for MBA selection in a UK business school. *International Journal of Selection and Assessment, 7,* 196-202.

Hecht, L. W. & Schrader, W. B. (1986). Graduate Management Admission Test®: Technical report on test development and score interpretation for GMAT® user. McLean, VA: Graduate Management Admission Council.

Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage Publications.

Johnson, J. W. & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods, 7*, 238-257.

Kaplan, R. & Sacuzzo, D. (1997). *Psychological testing: Principles, applications, and issues* (4th ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Koys, D. (2005). The validity of the Graduate Management Admission Test® for non-US students. *Journal of Education for Business, 80*, 236-239.

Kuncel, N., Crede, M., & Thomas, L. (2004, April). *A comprehensive meta-analysis of the predictive validity of the Graduate Management Admission Test® (GMAT®) and undergraduate grade point average (UGPA).* Poster session presented at the annual meeting of the Society for Industrial and Organizational Psychology, Chicago, IL.

Loades, R. (2005). The future of graduate management education in the context of the Bologna Accord: Highlights [online]. Retrieved August 3, 2005, from http://www.gmac.com/NR/rdonlyres/BD4E25FC-FB9F-467E-9ECE-F0BF9FDEA7F9/0/GMACBolognaProjectHighlights.pdf.

Olsen, M. (1957). The Admission Test for Graduate Study in Business as a predictor of first year grades in business schools, 1954-1955. Princeton, NJ: Educational Testing Service®.

Sireci, S. & Talento-Miller, E. (in press) Evaluating the predictive validity of Graduate Management Admission Test® scores. *Educational and Psychological Measurement*.

Talento-Miller, E. & Rudner, L.M. (2005). *VSS Summary Report for 1997-2004* (RR-05-06). McLean, VA: Graduate Management Admission Council®.

Appendix

Undergraduate Major

Business

Accounting
Business Administration
Commerce
Economics
Finance
Hospitality Management
Information & System Management
International Trade
Marketing
Management Sciences
Operational Research
Production and Operations Management
Telecommunications

Engineering

Aeronautical Engineering Agriculture/Agricultural Engineering
Architectural Engineering
Biology
Chemical Engineering
Chemistry
Civil Engineering
Computer Science/Engineering
Electrical Engineering
Environmental Science
Geology
Industrial Engineering
Mathematics
Mechanical Engineering
Medicine
Nuclear Engineering
Physics
Statistics
Veterinary Medicine
Zoology

Other

Architecture
Arts
Classics
European Cultural Studies
Geography
History
Journalism and Communication
Languages/Literature
Law
Philology
Philosophy
Politics
Psychology
Social Sciences

Language

English

Western European

Catalan
Danish
Dutch
Finnish
Flemish
French
German
Greek
Italian
Nowegian
Portuguese
Spanish
Swedish
Turkish
Valencian

Other

Afrikaans
Arabic
Bulgarian
Chinese
Czech
Farsi
Hebrew
Hindi
Hungarian
Japanese
Korean
Latvian
Malayalam
Mandarin
Marathi
Polish
Punjabi
Romanian
Russian
Taiwanese
Tamil
Telugu
Thai
Ukrainian
Vietnamese

Nationality

Western European

Austrian
Belgian
British
Cypriot
Danish
Dutch
Finnish
French
German
Greek
Icelandic
Irish
Italian
Luxembourger
Maltese
Norwegian
Portuguese
Spanish
Scottish
Swedish
Swiss
Turkish

Americas

American
Argentinean
Bolivian
Brazilian
Canadian
Chilean
Colombian
Costa Rican
Ecuadorian
Guatemalan
Mexican
Panamanian
Peruvian
Salvadorian
Uruguayan
Venezuelan

Eastern European, Africa, Middle East

Albanian
Algerian
Bulgarian
Cameroonian
Croatian
Czech
Ghanaian
Hungarian
Iranian
Israeli
Jordanian
Kenyan
Latvian
Lebanese
Moldavian
Nigerian
Polish
Russian
Saudi Arabian
South African
Swazi
Tunisian
Ukrainian

Asia Pacific

Australian
Chinese
Filipino
Indian
Indonesian
Japanese
Malaysian
Nepalese
New Zealander
Singaporean
South Korean
Taiwanese
Thai
Vietnamese