GMAC®

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

Predicting Graduate Program Success for Undergraduate and Intended Graduate Accounting and Business-Related Concentrations

Kara M. Owens and Eileen Talento-Miller

GMAC® Research Reports • RR-06-15 • August 7, 2006

Abstract

The validity of inferences made from test scores is an especially important topic for many audiences. Test-takers, admissions professionals, and program faculty want to ensure that admission test scores provide accurate information about an applicant's potential for success and are not biased for different subgroups. This study examined the validity of interpretations made about first year or mid-program graduate-level performance for different business-related undergraduate major and planned graduate concentration groups. Specifically, 20,437 examinees were grouped based on data collected from administrations of the Graduate Management Admission Test® (GMAT®) and validity studies conducted during 1999–2004. Results indicated that the GMAT® exam allows for valid interpretations to be made and that a single prediction equation for first year or mid-program success is not biased for any of the business-related groupings examined in this study.

Purpose

This study investigated the differential validity of the Graduate Management Admission Test® (GMAT®) and undergraduate grades as predictors of graduate-level grade performance for examinees who either completed an undergraduate business major (UGM) discipline or planned to enroll in a specific graduate-level business program. Specifically, the undergraduate and graduate business fields such as accounting, economics, finance, management, and others were investigated.

Although the particular courses taken to complete a certain UGM would undoubtedly affect the undergraduate grade point average (UGPA) of a graduate business school applicant, no studies were found to address possible differences in the predictive validity of grades for specific business majors. While the overall validity of the interpretations made from the GMAT® examination has been evaluated by many (e.g., Hecht & Schrader, 1986; Kuncel, Credé, & Thomas, 2004; Olsen, 1957) and estimates have been above values considered good (Talento-Miller & Rudner, 2005), no studies have investigated potential differences in the predictive validity of the GMAT® examination for examinees completing specific business-related UGM categories or those planning to attend graduate programs with concentrations in a business-related field. Predictive validity studies such as these are especially important, as differences would affect the inferences made from the admission factors and, consequently, the graduate admission decision.

The present study used data collected from 1997 to 2004. Data was based on student-level information submitted to the Graduate Management Admission Council® (GMAC®) by business graduate programs that participated in the GMAC® validity study service (VSS).

Validity differences were examined in terms of the amount of validity and prediction results calculated for

each business-oriented UGM and planned graduate concentration (PGC) group. As such, specific research questions were developed for this study.

Research Question 1: Does the predictive validity of the GMAT® exam and UGPA differ depending on UGM?

Research Question 2: Does the predictive validity of the GMAT® exam and UGPA differ depending on PGC?

Research Question 3: How does the use of a single prediction equation impact individuals from different UGM groups?

Research Question 4: How does the use of a single prediction equation impact individuals from different PGC groups?

Theoretical Framework

It is standard practice for test organizations to provide evidence of the validity and reliability of interpretations made about performance on the examinations they create and/or administer. This research becomes especially important when there are consequences or stakes tied to an examinee's performance on a test. "The higher the stakes associated with a given test use, the more important it is that test-based inferences are supported with strong evidence of technical quality" (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 1999, p. 139). Because admissions tests are used to make decisions about preparedness for and selection into graduate-level education, there are important stakes associated with examinee performance. For instance, the GMAT® exam is often used in combination with other admission criteria to predict first-year performance in graduate business programs. As such, it is important to evaluate the effectiveness of the GMAT® exam for predicting performance in business-related graduate programs.

In addition, it is necessary to validate the inferences made from the GMAT® exam and other admission tests for various subgroups to whom the test is administered (AERA, APA, & NCME, 1999). For instance, the tests should be validated separately for gender, citizenship, ethnicity, and other important subgroups. This research will determine if constructs irrelevant to the skills intended to be measured by the admission test are differentially influencing performance based on examinee subgroup membership. It would be unfair to administer a test that is consistently biased for or against specific subgroups of the population.

Previous research has compared predictive validity for different citizenship and gender groups for the GMAT® exam (Crooks & Heuvelmans, 1999; Dobson, Krapljan-Barr, & Vielba, 1999; Koys, 2005; Talento-Miller, 2005; Wilson, 1985). However, limited research has investigated the differences in prediction based on UGM or PGC subgroups. A 1957 study by Olsen examined UGM groups for performance differences on the GMAT® exam. Olsen found that examinees with a non-business UGM performed better on the exam than those that majored in a business-related field, but the examinees with a business background had higher UGPAs. When these predictors were controlled, however, graduate-level performance did not differ between the two groups.

Pitcher, Deemer, and Smith (1968) also explored differences in the prediction of first-year performance for different UGM groups. The results revealed that performance in full-time graduate business programs for humanities and social science UGM groups was overpredicted, indicating that these students were predicted to perform better than they actually did. On the other hand, performance for the engineering UGM group was underpredicted when GMAT® scores were combined with UGPA as predictors. Additionally, multiple correlations with first-year grades using these predictors varied depending on UGM. The multiple correlations were again highest for humanities and lowest for engineering UGM groups. These studies indicated that UGM may be related to prediction of first year graduate performance.

While additional research has examined differences among UGM groups for various other admissions tests (Baird, 1975; Boldt, 1986; Hale, Rock, & Longford, 1991), further research investigating UGM differences in terms of the validity of interpretations made from GMAT® exam scores is very limited. Likewise, research investigating differences in predictive validity estimates for GMAT® exam scores based on PGC has not been conducted or made publicly available. The current study compared predictive validity for various business-related UGM and PGC groups. This study examined first-year and mid-program performance and GMAT® exam scores

and UGPA to determine if relationships between these variables changed based on a student's UGM or PGC.

Methods

Data

Data were collected from two sources: a database with background and test score information on GMAT® examinees and a database of validity studies conducted through GMAC®. The latter database contained information from 273 validity studies conducted between 1999 and 2004, for a combined sample representing more than 40,000 students. From these validity studies, information was gathered on GMAT® scores, UGPA, and graduate school performance for each student. Background data was collected at the time students took the GMAT® exam. Upon taking the test, examinees had the option of answering background information questions about topics such as gender, UGM, and PGC. It is important to note that PGC does not reflect actual graduate study, only the concentration planned at the time of taking the test. No data was available to determine the focus of students once they entered their graduate program.

To add UGM and PGC information to the performance data, students from the validity study database were matched to the GMAT® examinee database. If data could not be matched uniquely or if information was not provided regarding UGM or PGC, the individual was excluded from the study. The final matched database included 20,437 students from 168 validity studies. For PGC, only those categories with more than 100 complete cases were retained. The remaining business-related concentrations were collapsed into the "Business-Other" category. Table 1 shows the summary information for this sample. The separate major analyses only included the 5,402 cases that selected a business UGM and the 11,246 cases that indicated a business PGC, though all cases with valid data were used for the single prediction equation. Individual prediction equations for the different business-related majors and concentrations were compared to the single equation to determine if differences exist between a sample of the entire test-taker population and examinee groups with specific backgrounds and interests.

Table 1. Frequencies for Undergraduate Majors and Planned Graduate Concentrations

	N	% of Total	% of Business
Undergraduate			
Accounting	1,941	9.5%	35.9%
Economics	1,687	8.3%	31.2%
Finance	643	3.1%	11.9%
Management	1,131	5.5%	20.9%
Non-business	15,035	73.6%	—
Graduate			
Accounting	585	2.9%	5.2%
Finance	3,045	14.9%	27.1%
Information Systems	443	2.2%	3.9%
International Business	459	2.2%	4.1%
Marketing	729	3.6%	6.5%
Operations	184	0.9%	1.6%
General Management	2,163	10.6%	19.2%
Business-Other	3,638	17.8%	32.3%
Non-business	9,191	45.0%	—
All	20,437	100%	100%

Analyses

Correlations were used to determine the overall predictive validity for all examinees and for each group. The three GMAT® section scores—verbal, quantitative, and analytical writing (AWA)—and the UGPA for students were correlated with their standardized first-year or mid-program graduate GPA (Z-GPA). In addition, the predictive validity for the combination of these four variables was assessed. Each of the five analyses (one for each of the individual predictors and one for the combination of predictors) was conducted for the overall group and then repeated for each of the UGM and PGC groups of interest.

Because of grading and other differences across programs, graduate GPA values were standardized within school and multiple regression equations were built with dummy codes representing each validity study. This methodology, described in Talento-Miller, Rudner, and Owens (2006) allows data to be combined across studies without introducing excessive extraneous variance from program differences. No attempts were made to correct the validity coefficients for errors caused by statistical artifacts such as restriction of range and sampling effects. As a result, it is expected that the values observed in the current study would be lower than the actual expected validity across programs.

Even though allowances were made to account for differences in graduate programs in assessing the predictive validity, it should be noted that other systematic differences may impact the ability estimates of predictive validity by UGM and PGC. For instance, certain undergraduate and graduate business concentrations may have more stringent grading practices than other concentrations. Because predictive validity is influenced by variability in the predictor and criterion variables, any differences would impact the estimates of predictive validity. Table 2 lists the descriptive statistics for each of the groups and illustrates the differences in variability across groups which may impact the validity coefficients.

The regression equation for the combined predictors using all valid cases was used to predict graduate GPA. The average standardized residuals from this prediction were calculated for each of the UGM and PGC groups to determine the impact of treating these groups the same in admissions. Because graduate GPA was standardized, the residuals represent the effect size of the differences in predicted versus observed grades for each of the different UGM and PGC groups. This effect size is expressed as standard deviation units and can be interpreted to determine if practical differences in predicted and observed performance exist among groups.

Results

The results for each of the four UGM and eight PGC groups along with the group overall are presented in Tables 3 and 4 and graphically depicted in Figures 1 and 2. Based on the results displayed in Table 3 and Figure 1, it appears that predictive validity is similar for accounting, economics, and management undergraduate majors across all individual predictors as well as the combination of predictors. The pattern suggests that validity is higher for finance majors than it is for the other groups, but lower for the group equation representing all students. It is important to note that, for most comparisons, the magnitude of the differences is quite small, particularly for the combined predictors. Additionally, differences in validity estimates for some of the individual predictors are quite small, indicating that for some of the UGM groups, there are only slight differences in the amount of variance accounted for by different individual predictors.

Table 2. Descriptive Statistics for UGM and PGC Groups										
Concentrations	AWA		Verbal		Quantitative		UGPA		Z-GPA	
	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
Undergraduate										
Accounting	4.18 (0.80)	1512	32.0 (7.0)	1929	35.4 (7.8)	1929	3.280 (0.420)	1849	0.108 (0.941)	1941
Economics	4.28 (0.83)	1159	34.1 (7.1)	1678	38.2 (7.4)	1678	3.233 (0.417)	1592	-0.057 (0.966)	1687
Finance	4.20 (0.72)	453	33.5 (6.8)	637	36.9 (7.6)	637	3.262 (0.421)	614	-0.007 (0.982)	643
Management	4.12 (0.72)	884	31.0 (6.8)	1123	33.1 (8.4)	1123	3.207 (0.427)	1087	-0.133 (1.024)	1131
Non-business	4.34 (0.87)	11458	33.5 (7.3)	14905	38.3 (8.3)	14909	3.195 (0.435)	13655	-0.045 (1.013)	15035
Graduate										
Accounting	4.06 (0.74)	484	30.6 (6.5)	584	34.1 (8.0)	584	3.293 (0.433)	565	0.016 (1.004)	585
Finance	4.31 (0.82)	2071	34.1 (7.0)	3023	38.9 (7.9)	3023	3.251 (0.426)	2810	-0.008 (1.008)	3045
Information Systems	4.23 (0.91)	277	32.2 (7.5)	440	38.1 (8.2)	440	3.183 (0.450)	405	0.007 (0.941)	443
International Business	4.34 (0.90)	332	32.5 (7.9)	454	37.9 (8.2)	454	3.190 (0.460)	410	-0.062 (1.015)	459
Marketing	4.46 (0.83)	510	32.9 (7.1)	724	36.3 (8.4)	724	3.194 (0.426)	658	-0.117 (0.977)	658
Operations	4.25 (0.86)	126	33.3 (7.7)	183	38.5 (8.3)	183	3.189 (0.435)	164	0.089 (0.961)	184
General Management	4.44 (0.86)	1711	34.2 (7.1)	2151	38.0 (8.4)	2151	3.214 (0.441)	1954	0.065 (1.017)	2163
Business-Other	4.23 (0.76)	2686	33.5 (7.3)	3595	36.7 (8.2)	3599	3.191 (0.426)	3430	-0.073 (0.996)	3638
Non-business	4.30 (0.90)	7269	33.0 (7.3)	9118	37.8 (8.2)	9118	3.201 (0.431)	8401	-0.052 (1.005)	9191

Table 3. Predictive Validity across Groups by Undergraduate Major					
	AWA	Verbal	Quantitative	UGPA	A+V+Q+U
All	0.208	0.293	0.306	0.303	0.403
Accounting	0.365	0.366	0.344	0.418	0.483
Economics	0.401	0.411	0.409	0.453	0.501
Finance	0.506	0.561	0.561	0.581	0.584
Management	0.401	0.437	0.440	0.455	0.496
Non-business	0.224	0.305	0.328	0.301	0.423

Figure 1. Comparison of Predictive Validity across Groups by Undergraduate Major

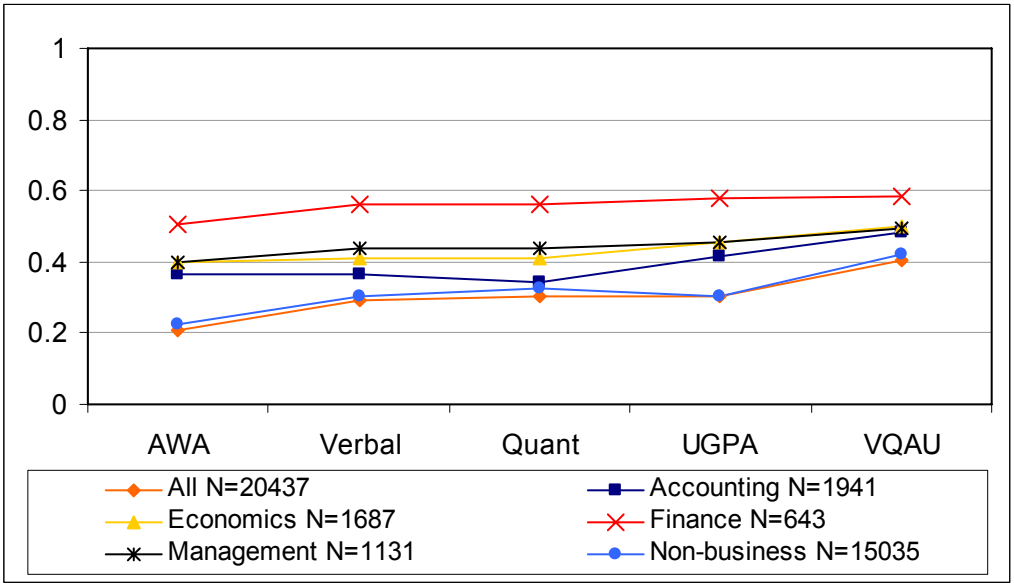
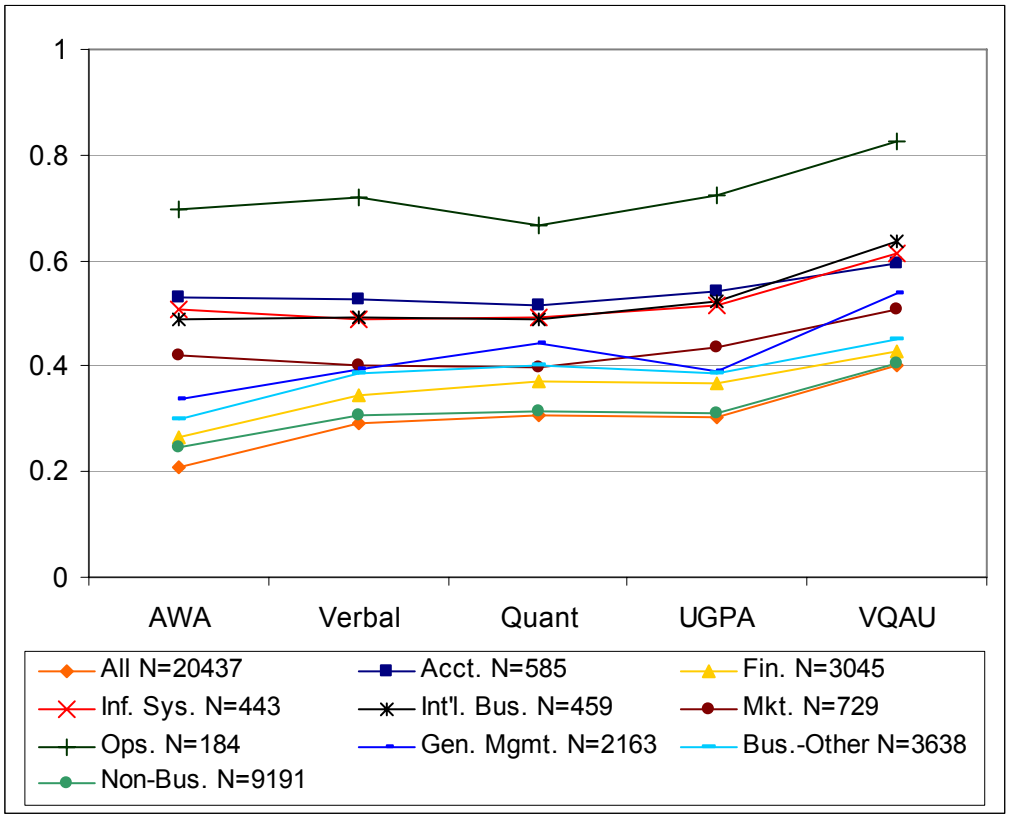


Table 4 provides the predictive validity values for the various PGC groups. Though some of the concentrations yielded slightly different patterns of predictive validity, the combination including all of the predictors yielded the highest validity values for all of the groups. For the PGC comparisons, all groups except operations seemed to cluster together, as can be seen in Figure 2. Particularly, the validity values for the accounting, information systems, and international business groups were nearly identical. Values were very high for the operations group, but relatively low for the finance and business-other groups. However, results for the operations group should be cautiously interpreted due to the small sample size (N= 184) and low Z-GPA variance; this group may not be representative of all test-takers anticipating enrollment into an operations graduate program. Similar to the results for UGM, the group containing all examinees had the lowest predictive validity.

Table 4. Predictive Validity across Planned Graduate Concentration Group					
	AWA	Verbal	Quantitative	UGPA	A+V+Q+U
All	0.208	0.293	0.306	0.303	0.403
Accounting	0.531	0.526	0.515	0.540	0.594
Finance	0.267	0.346	0.371	0.366	0.428
Information Systems	0.507	0.488	0.493	0.516	0.612
International Business	0.487	0.494	0.490	0.523	0.638
Marketing	0.419	0.400	0.396	0.437	0.509
Operations	0.696	0.720	0.668	0.724	0.827
General Management	0.337	0.393	0.443	0.391	0.538
Business-Other	0.298	0.386	0.402	0.385	0.450
Non-business	0.247	0.306	0.314	0.310	0.406

Figure 2. Comparison of Predictive Validity across Planned Graduate Concentration Group



For both the UGM and PGC comparisons, the patterns of predictive validity from verbal scores, quantitative scores, and UGPA were similar, though predictive validity values fluctuated slightly depending on group membership. However, three groups showed somewhat different patterns of prediction depending on the variable(s) used. The accounting UGM group appeared to have better prediction than the other UGM groups, particularly with respect to UGPA. The operations PGC and the general management PGC groups also appeared to differ from the

trend. The former showed much higher predictive validity than any of the other PGC groups and the latter revealed greater predictive validity estimates for quantitative scores when compared to the other individual predictors for the general management group. For many of the groups, the validity values from the AWA scores were lower than for the other predictors. As expected, the combined prediction was an improvement over the individual predictors for all groups, but for some groups, the difference was small.

Using data from all students, a regression equation was built using the combination of predictors, and residuals were calculated. Because not all cases had complete data for all predictors, group and overall sample sizes differ for each of the comparisons, though each group still had more than 100 cases. The residual was calculated as observed graduate GPA minus the predicted value, so positive average residuals indicate the group performed better than predicted, and negative values indicate they did not perform as well as predicted.

The average residuals for each group are presented in Table 5. In general, most of the UGM and PGC groups performed better than expected, but the differences were small. Because the residuals are standardized, they can be considered a measure of effect size or practical significance. Values around 0.2 would be considered small effects according to Cohen (1969). Based on this cutoff, the effects of using a combined prediction equation across all groups were very small. The standardized residuals were re-scaled to determine the difference that would be observed for a four-point grading scale, using the average standard deviation of grades observed across the studies (SD = 0.338). The largest difference equated to little more than 0.05 on a 4-point scale, indicating that there is little practical difference in predictive validity depending on UGM or PGC group. Thus, a prediction equation based on the overall sample can be used for applicants from different UGM and PGC groups. Analyses of variance of the residuals performed separately for both the UGM and PGC groups indicated that the omnibus test was significant at a $p = .01$ level for both analyses. The significant omnibus tests were followed with Bonferroni adjusted post hoc tests. Though there were no significant differences in standardized residuals among comparisons of the various business-related undergraduate or graduate groups, there were significant differences when business-related groups were compared to non-business groups. Specifically, the non-business UGM group differed significantly from accounting and finance UGM groups, and the non-business PGC group differed significantly from the business-other PGC group in terms of standardized residuals. Essentially, error not accounted for by the combination of predictors or the standardized residuals differed in magnitude across these groups.

Table 5. Average Residuals by Group

Concentration	N	Residual	Scaled difference
Undergraduate			
Accounting	1,456	0.104	0.035
Economics	1,107	0.036	0.012
Finance	435	0.155	0.052
Management	859	–0.021	–0.007
Non-business	10,566	–0.023	–0.008
Graduate			
Accounting	475	0.017	0.006
Finance	1,921	0.045	0.015
Information Systems	268	0.013	0.004
International Business	313	–0.092	–0.031
Marketing	485	–0.039	–0.013
Operations	120	0.118	0.040

Table 5. Average Residuals by Group			
Concentration	N	Residual	Scaled difference
General Management	1,596	0.024	0.008
Business-Other	2,523	0.048	0.016
Non-business	6,722	–0.033	–0.011
All	14,423	0.000	0.000

Conclusion

In summary, GMAT® exam scores and UGPA allow for valid interpretations regarding first-year and mid-program performance in graduate education for students from various UGM and PGC groups. The high predictive validity values for different UGM and PGC areas indicated that UGPA and GMAT® exam scores were useful in selecting qualified applicants for admission regardless of undergraduate background and graduate program interest. Additionally, the standardized residuals for the different graduate and undergraduate areas indicated that there were no meaningful differences in the prediction of success depending on one's UGM or PGC for business-related areas. Thus, success can be adequately predicted for different business undergraduate majors and PGC groups using a single prediction equation based on GMAT® Verbal, Quantitative, and AWA scores combined with UGPA. Though slight differences in standardized residuals were found for some group comparisons (i.e., business-related versus non-business related groups), the small standardized residual values for these groups indicates that these differences are not practically meaningful. Future research can examine actual graduate concentrations rather than anticipated or planned graduate concentration enrollment and expand to other business-related fields to determine if differences in prediction exist. Additionally, research could examine second-year or end-of-program GPA to determine if there are long-term differences in predictive validity for different business areas.

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

An earlier version of this paper was presented August 2006 at the American Accounting Association Annual Conference in Washington, DC.

The authors wish to express their appreciation to Lawrence Rudner for his assistance in preparing the data for analysis.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, D. C.: APA.

Baird, L.L. (1975). Comparative prediction of first year graduate and professional school grades in six fields. Educational and Psychological Measurement, 35, 941-946.

Boldt, R. F. (1986, December). Generalization of GRE® general test validity across departments (Research Report No. 86-46). Princeton, NJ: Educational Testing Service.

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

Crooks, S., & Heuvelmans, A. (1999, October). The GMAT® as a predictor of academic performance on eight European MBA programs. Arnhem, The Netherlands: The National Institute for Educational Measurement.

Dobson, P., Krapljan-Barr, P., & Vielba, C. (1999). An evaluation of the validity and fairness of the graduate management admissions test (GMAT®) used for MBA selection in a UK business school. International Journal of Selection and Assessment, 7(4), 196-202.

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

Hale, G. A., Rock, D. A., & Longford, N. T. (1991). Effects of graduate coursework on the GRE® quantitative score for recent and nonrecent college graduate (Research Report No. 91-15). Princeton, NJ: Educational Testing Service.

Koys, D. (2005). The validity of the Graduate Management Admissions Test® for non-U.S. students. Journal of Education for Business, 80(4), 236-239.

Kuncel, N. R., Credé, M., & Thomas, L. T. (2004). A comprehensive meta-analysis of the predictive validity of the Graduate Management Admission Test (GMAT®) and undergraduate grade point average (UGPA). Manuscripts submitted for publication.

Olsen, M. (1957, January). 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®.

Talento-Miller, E. (2005, October). GMAT® validity for six European MBA programs (Research Report RR-05-07). McLean, VA: Graduate Management Admission Council®.

Talento-Miller, E., & Rudner, L. M. (2005). GMAT® validity summary report for 1997 to 2004 (Research Report No. RR-05-06). Mclean, VA: Graduate Management Admission Council.

Talento-Miller, E., Rudner, L., & Owens, K. (2006). Assessing validity by combining examinees across programs. Paper presented at the International Testing Commission Conference, July 6-8, 2006, Brussels, Belgium.

Wilson, K. M. (1985). Factors affecting GMAT® predictive validity for foreign MBA students: An exploratory study. Princeton, NJ: Educational Testing Service.

Young, J. W. (2001). Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis (Research Report No. 2001-6). New York: The College Board.