GRADUATE MANAGEMENT ADMISSION COUNCIL

Validity Study of Non-MBA Programs

Eileen Talento-Miller

GMAC® Research Reports • RR-09-12 • November 24, 2009

Although the validity of Graduate Management Admission Test® (GMAT®) scores for predicting performance in Master of Business Administration (MBA) programs is well established (Kuncel, Crede, & Thomas, 2007; Oh, Schmidt, Shaffer, & Le, 2008; Talento-Miller & Rudner, 2008), less information is known about the validity of scores for other graduate business programs, such as Master of Accountancy (MAcc) programs or Master of Science (MS) programs with various business concentrations. To ensure that inferences made based on test scores are appropriate, validity should be examined for all groups and situations for which test scores are used (AERA, APA, & NCME, 1999). It is therefore important to gather information about non-MBA programs to determine how they may differ from MBA programs regarding interpretations of GMAT scores and other typical admission criteria.

Previous research has suggested that different program types or disciplines may observe different levels of validity. The study by Talento-Miller and Rudner (2008) showed that differences in predicting performance were observed in executive MBA programs compared to full- or part-time MBA programs. Since these are all MBA programs, one might expect the courses of study within the programs to be the same. Entering students may differ drastically, however, since executive programs generally require extensive work experience. Studies of doctoral business programs yielded predictive validity estimates that were lower on average than those found in MBA programs (Siegert, 2007; Zwick, 1993), providing further evidence that program differences may moderate the level of predictive validity observed for GMAT scores. Kuncel, Hezlett, and Ones (2001) suggested that validity of test scores might differ by discipline. Incoming students would have different backgrounds and would be subjected to different requirements within their programs.

These same differences might be expected of entrants into non-MBA graduate business programs compared to MBA programs. Masters programs of specific disciplines would undoubtedly include different business courses than typical MBA programs, such as more advanced subject-specific curricula. Admission requirements may be likely to call for specific courses on transcripts or particular undergraduate majors. In addition, the MBA generally is regarded as a post-experience program that requires nearly all incoming students to have a minimum amount of work experience. Masters programs in business, on the other hand, may be more likely to accept students directly out of their undergraduate programs.

Differences in the backgrounds of entering student populations and differences in the requirements within the programs indicate that the predictive validity of admission factors for MBA programs may not generalize to other types of graduate business programs. The purpose of this study is to determine the level of predictive validity from admission factors used for non-MBA graduate business programs and to compare the results against what is known for MBA programs. Differences for these masters programs compared to MBA programs may have implications for best practices for numerous admissions staff in the business school, particularly those who evaluate applicants for multiple programs.

Methodology

Participants

To identify potential participants for a validity study, GMAT score-sending patterns were examined by program code. Institutions often have multiple program codes to which scores are sent, which they

use to differentiate among applicants to different programs, such as part-time MBA, distance MBA, Master of Accountancy, and so on. Program codes for non-MBA programs were ranked based on the number of GMAT score reports sent. The programs with the highest score-reporting volume were invited to participate in a special study of the Validity Study Service (VSS) of the Graduate Management Admission Council® (GMAC®) during the summers of 2008 and 2009. Invitations were extended through phone calls, emails, and postal letters.

To be included as part of the non-MBA study, programs needed to provide data for students enrolled in the program. Required data included GMAT scaled scores, undergraduate grade point average (UGPA), and mid-program grade point average (MPGPA). Mid-program grade point average could represent grades in the first year of a two-year program, halfway through a longer or shorter program, or just grades in a program's core courses (exclusive of electives). For the individual program studies, the validity information was customized to include either additional predictors and/or criteria, or was calculated separately by specified groups. Examples of additional data likely to be included would be gender, undergraduate major, and final grades of each student.

Data Analysis

Predictive validity was calculated separately for each program using combined data. Validity coefficients were obtained for individual predictors as well as combinations of predictors using ordinary least squares linear regression. Summaries of program-level results include mean and median validity values for each of the standard combinations.

To provide the most useful information about the predictive validity, this study employed statistical methods designed to reduce artifactual error. For example, although inferences are made based on admissions information for the entire applicant population, only admitted students can supply information on the criterion variable of program grades. The data used for analysis, then, arguably represent the selection of only the most able candidates. This selection effect is commonly referred to as the *restriction of range problem*. Statistical adjustments to the simple and multiple correlations were applied using the formulas provided in Hunter and Schmidt (1990), which is based on the standard deviation of the applicant population compared to the observed data. The applicant population is defined operationally for this study as any scores sent to the program code within the most recent period, since sending of scores to a program can arguably be considered the beginning of an application process (Stolzenberg & Relles, 1985). The data used to calculate the corrections include the official GMAT scores of all senders, as well as self-reported information on UGPA. Self-reported information was considered sufficient for this purpose since a study by Talento-Miller and Peyton (2006) showed that the mean and standard deviations of self-reported and true UGPA of graduate business students were nearly identical, suggesting that errors in self-report for this group tend to cancel each other out when reported in aggregate.

In order to compensate for differences among programs when combining data, this study applied the methodology described in Talento-Miller, Rudner, Owens, and Guo (2006). To account for possible grading differences, the outcome variable of MPGPA was standardized within programs so that data from each program had a mean of 0 and a standard deviation of 1. Subsequent regression analyses using combined data included dummy variables representing program membership to account for the additional error resulting from program differences. The Talento-Miller et al. study (2006) showed that applying this methodology results in validity values from combined data that more closely approximate results achieved by summarizing from individual studies. The benefit of using the combined data is the ability to make group comparisons (such as gender or discipline) with more robust sample sizes than are possible within any individual study. For the data combined across programs, the hypothesized applicants include the population of GMAT examinees in the most recent three-year period who sent scores to non-MBA masters programs, which was the basis for the parameters used in the restriction-of-range adjustments. The purpose in selecting this population was to derive values that are appropriate for

generalizing across non-MBA business masters programs.

This study employed Pratt Index calculations to determine the contribution of each of the variables in the multiple regression (Johnson & LeBreton, 2004; Thomas, Hughes, & Zumbo, 1998). The product of the simple correlation of the predictor variable and the beta weight in the multiple regression was divided by the coefficient of determination for the combination to calculate the percentage contribution of each variable toward the validity. The relative contributions represent the importance of each variable toward prediction. Pratt Index values were not calculated for the dummy variables that represented the different programs, since program differences would be noise in the data, or sources of error, whose values would not sum to 100 percent. The values were rescaled to determine the contribution of just the relevant variables, so that the sum of the values would then equal 100 percent.

Results

Twelve different business schools submitted a total of 15 datasets for non-MBA programs with a total of 1,627 usable cases. The majority of programs were masters programs in accounting and all were located in the United States. Data were available within multiple programs on gender, citizenship (domestic versus nondomestic), race/ethnicity (majority versus minority), and undergraduate major (business or accounting versus other), so comparisons were made using these groupings in the combined dataset.

Table 1 summarizes the results of the validity studies from the 15 programs. One program did not supply enough valid information on AWA scores, so the summary information including AWA as a predictor is based on only 14 programs. For the individual predictors, the validity values were similar for GMAT Total scores and UGPA. The combinations of GMAT scores and UGPA showed about a standard deviation increase over the individual predictors, an arguably large effect size, indicating that the two types of information (test scores and grades) contribute important information to predicting academic success for these programs.

The combined dataset showed similar results, as shown in Table 2. Using multiple sources of admission information results in a substantial increase of variance explained. Squaring the validity coefficient for either UGPA or Total score alone shows the variance explained to be 15 percent to 16 percent for these individual variables. Combining the two predictors, however, doubles the amount of variance explained to 32 percent. The values from the combined dataset were close to the average validity values across programs, although the correspondence was closer for the combined predictors as opposed to the individual predictors, which was consistent with expectations from the methodology used (Talento-Miller et al., 2006). The Pratt Index values showed that in the multiple regression the contribution from UGPA accounted for about half of the prediction, but each variable contributed uniquely and meaningfully to the overall validity.

Table 1: Summary Predictive Validity Results for 15 Masters Programs

	Mean	SD	25th Percentile	Median	75th Percentile
Verbal (V)	0.331	0.149	0.248	0.361	0.408
Quantitative (Q)	0.204	0.229	0.046	0.312	0.352
AWA (A)	0.216	0.114	0.148	0.248	0.290
Total (T)	0.397	0.148	0.301	0.431	0.485
UGPA (U)	0.431	0.157	0.314	0.411	0.490
V + Q + U	0.574	0.132	0.466	0.532	0.642
T + U	0.547	0.149	0.430	0.560	0.618
V + Q + A + U	0.596	0.130	0.506	0.585	0.657
T + U + A	0.577	0.139	0.488	0.574	0.635

Table 2: Validity Results From Combined Data and Contribution of Predictors							
	N	**R**	**V**	**Q**	**A**	**T**	**U**
Verbal (V)	1,624	0.291	100%				
Quantitative (Q)	1,624	0.269		100%			
AWA (A)	1,333	0.238			100%		
Total (T)	1,686	0.401				100%	
UGPA (U)	1,627	0.393					100%
V + Q + U	1,624	0.537	26.1%	21.9%			52.0%
T + U	1,627	0.568				51.2%	48.8%
V + Q + A + U	1,333	0.580	22.5%	17.8%	8.8%		50.1%
T + U + A	1,333	0.598			8.9%	44.5%	46.5%

Results by Group

Four dichotomous groupings were evaluated in the combined dataset, representing comparisons of gender, citizenship, race/ethnicity, and undergraduate major, as defined by the individual programs. Within the GMAC VSS, citizenship groupings are defined as domestic versus nondomestic, to accommodate the comparison of programs around the world. Since all the programs included in this study are located in the United States, however, the citizenship comparison could also be classified as US citizens versus non-US citizens. Each of the programs included in the race/ethnicity comparison defined the groupings as majority versus minority. Without further elaboration on the grouping, it is not possible to determine which race/ethnicity designations would qualify under the minority category, or even whether the categories included only US subgroups. For the undergraduate major category, some programs categorized majors into accounting versus other, while others categorized majors into business versus other. The programs that examined these undergraduate major groupings showed similar patterns in validity, whether the grouping was based on accounting major or business major. The similarity of the accounting and business major groups justified combining these two in comparison with the other major grouping.

Table 3 gives the results of the group analyses, showing the total sample size for each group, the combined validity for the combination of V + Q + A + U, and the Pratt Index relative contribution for each variable in the combination. In addition, the last column of the table gives the average standardized residual for the group. The standardized residual is based on a single prediction equation built from the V + Q + A + U analysis for the full sample. Since the residual is defined by predicted minus actual MPGPA, negative average values indicate the group received higher grades than predicted, while positive average values suggest the group received lower grades than predicted based on their admission information.

Table 3: Validity Results and Variable Importance by Group							
	N	**R**	**V**	**Q**	**A**	**U**	**ASR**
All	**1,333**	**0.580**	**22.5%**	**17.8%**	**8.8%**	**50.1%**	**0.00**
Gender							
Male	704	0.550	23.0%	8.4%	14.5%	54.1%	0.038
Female	629	0.591	20.7%	30.3%	6.0%	43.0%	–0.042
Citizenship							
US	1,058	0.592	13.7%	18.5%	10.3%	57.5%	–0.024
Non-US	189	0.620	42.0%	30.5%	10.5%	17.1%	0.140
Race/Ethnicity							
Majority	450	0.623	5.2%	32.9%	11.8%	50.0%	–0.010
Minority	202	0.638	0.0%	58.8%	9.0%	32.4%	0.022
Major							
Acct/Bus	220	0.607	1.3%	28.8%	3.0%	66.9%	0.030
Other	125	0.501	66.2%	7.6%	0.0%	26.4%	-0.060

In contrast with the overall results, there were some notable findings by group. Compared to the validity for all students and the other groups, Verbal scores contributed relatively more to prediction for the non-US citizens and other undergraduate majors, but less to prediction for both race/ethnicity groups and the accounting/business undergraduate major groups. For Quantitative scores, the major differences appeared with higher contributions to prediction for the minority grouping, and relatively low contributions for the male group and other undergraduate major group. Writing scores contributed little to nothing to prediction for both undergraduate major groups. The highest contributions for UGPA were in the accounting/business undergraduate major group and the US citizenship group; the lowest contributions were observed in the non-US citizenship group and other undergraduate major group.

Average standardized residuals by group were low, suggesting that using a single prediction equation to predict MPGPA was effective for all groups studied. Because the residuals are standardized, the values were equivalent to a Cohen's *d* effect size, where amounts below 0.2 are considered small. The largest average residual of 0.140 was observed for the non-US citizens. Converted back to a 4.0 GPA scale, the results suggest the average earned MPGPA values for the non-US group would be 0.043 less than the average predicted GPA for the group. Tests of statistical significance verified that the difference among citizenship groups may warrant further consideration ($t_{1245} = -2.09$, $p = 0.037$), but all other group comparisons suggested that the differences could be attributed to chance variations (all $p > 0.10$).

Comparison of Program Types

The GMAT exam is frequently associated with MBA programs, so it is useful to determine how results for other program types compare to the results for MBA programs. Tables 4 and 5 compare the demographic characteristics and admission factors for all GMAT examinees—those intending to pursue MBA programs and those intending to pursue other masters programs for testing years 2007 to 2009. In general, it appears that applicants to other masters programs are relatively more likely to be female or come from a business undergraduate program than applicants to MBA programs. Masters program applicants are younger than those intending to enroll in MBA programs, with a small to moderate effect size (Cohen's $d = -0.33$). Average GMAT scores are lower for the other masters program applicants compared with MBA program applicants, and the average UGPA is higher; but all effects are very small (Cohen's $d < |0.20|$). These

differences among applicant populations reinforce the need to examine validity separately for program types.

Table 4: Descriptive Characteristics of GMAT Examinees by Intended Program

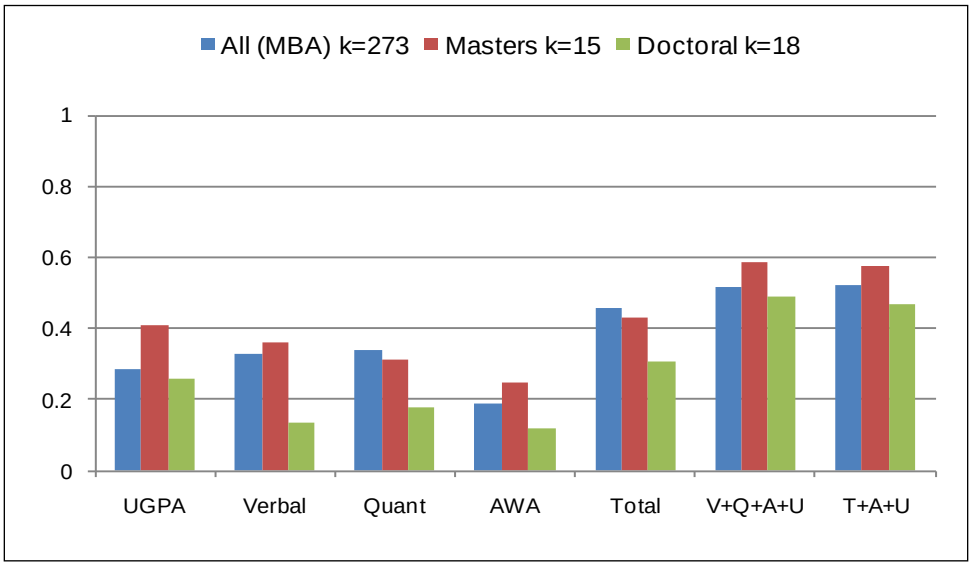
	All Examinees		MBA		Other Masters	
	N	%	N	%	N	%
Gender	731,647	100%	487,251	100%	88,282	100%
Female	288,738	39%	184,143	38%	46,272	52%
Male	442,909	61%	303,108	62%	42,010	48%
Citizenship	731,647	100%	487,251	100%	88,282	100%
Non-US	357,686	49%	219,850	45%	35,194	40%
US	373,961	51%	267,401	55%	53,088	60%
Ethnicity of US Citizens	373,961	100%	267,401	100%	53,088	100%
Native American	2,931	1%	2,169	1%	364	1%
Asian American	41,564	11%	29,990	11%	4,821	9%
African American	31,405	8%	23,421	9%	3,923	7%
White (Non-Hispanic)	257,448	69%	182,999	68%	38,778	73%
Mexican American	7,844	2%	5,735	2%	1,149	2%
Puerto Rican	2,555	1%	1,827	1%	356	1%
Other Hispanic American	10,779	3%	7,813	3%	1,548	3%
Other	11,848	3%	7,967	3%	1,243	2%
Multiracial	4,243	1%	3,073	1%	506	1%
Multiethnic	3,344	1%	2,407	1%	400	1%
Undergraduate Major Category	715,671	100%	484,561	100%	87,869	100%
Science	43,227	6%	30,182	6%	5,068	6%
Business	338,540	47%	231,626	48%	59,249	67%
Humanities	36,175	5%	26,822	6%	3,743	4%
Social Science	109,071	15%	79,490	16%	11,235	13%
Other Major	63,128	9%	23,268	5%	3,858	4%
Engineering	125,530	18%	93,173	19%	4,716	5%

Table 5: Mean(SD) for Age and Admission Factors by Intended Program

	All Examinees	MBA	Other Masters
N	731,647	487,251	88,282
Age	26.96 (5.47)	27.03 (5.08)	25.30 (5.80)
Quantitative	35.99 (10.77)	35.53 (10.75)	34.74 (10.56)
Verbal	27.96 (9.09)	28.35 (8.93)	27.38 (8.66)
Total	538.86 (120.73)	538.49 (120.83)	525.54 (114.79)
AWA	4.55 (1.18)	4.63 (1.11)	4.49 (1.17)
UGPA	3.26 (0.48)	3.25 (0.47)	3.32 (0.44)

From more than 30 years worth of data from the GMAC VSS, of studies that identify the program, more than 90 percent were MBA or executive MBA programs. Although program was not specified in the Talento-Miller & Rudner (2008) meta-analysis, assuming the same proportion of MBA programs versus other programs in the study makes it reasonable to consider the average values as representative of MBA programs. Using that assumption and values from the study of doctoral business programs by Siegert (2007), Figure 1 compares the validity values across program types. The sample size represents the number of programs represented in the meta-analysis and the average validity values are based on the median.

Figure 1. Comparison of Validity Values Across Program Types



The evidence shows that UGPA and GMAT scores contribute to the prediction of mid-program GPA for different program types. The combined predictor comparisons indicate that overall median validity is similar across program types, exceeding 0.45 for all groups. The relatively higher average validity for the masters programs in the multiple regressions may be a result of higher validity for UGPA. Although the validity values for the individual variables such as UGPA and GMAT Total are high, it is clear that combining test score information with previous grades meaningfully improves prediction over considering one source of information alone, across all the program types.

Discussion

The predictive validity for GMAT scores is well established for MBA-type programs. Additional information regarding other types of graduate business programs supports the generalizability of the validity of GMAT scores for admission in various contexts. These other graduate business program types would be expected to differ in content, such as more advanced courses in a specific subject area, as opposed to a sampling of general courses across several areas. The MAcc program may presuppose knowledge of accounting, looking for specific courses to be included in the undergraduate education. There would also likely be a significant difference in the applicant population. Many MS programs may be considered pre-experience business courses, meaning no work experience is required. This differs from many MBA programs, which often require relevant work experience. The difference in the work experience requirement may result in quantitative (age) as well as qualitative (employment background) differences in the applicant populations. The differences in program content and applicant population suggest that evidence

is needed to support the generalizability of GMAT validity across these situations.

Overall, the validity evidence based on the programs included in the current study support the generalizability of the usefulness of GMAT scores in selecting applicants for admission to non-MBA masters programs by showing the substantial improvement over other sources of information, specifically previous grades. Results showing the high relative contributions of UGPA to prediction are consistent with previous studies showing higher validity values from UGPA from younger students compared to older students (Hecht, Manning, Swinton, & Braun, 1989). This again suggests that the applicant population for the masters programs studied may be younger than the MBA-seeking population. One finding was the strong effect of UGPA for previous accounting and business undergraduate majors, with more than two-thirds of the prediction contributed by previous grades. This is not surprising if one logically surmises that the best predictor of grades in a graduate accounting program would be grades in an undergraduate accounting program. What is notable in this case is the opposite group. The other undergraduate majors, who were not designated as accounting or business, underscored the value of GMAT scores in admission, with more than 70 percent of the prediction contributed by the test scores. The standardized nature of the GMAT exam is designed to take advantage of this very situation—when you have different backgrounds of students, such as different majors, different universities, or even different countries. Clearly, the results of this study in general, and the results for the other undergraduate major group in particular, underscore this benefit.

There are several limitations to the current study. Obviously, the few programs included—although aggregated across more than a thousand students—is only a fraction of the total number of students and programs that could be studied, and there may be quite a bit more variation in the results than would be observed with a larger sample of studies. The current study included mostly accounting programs, so results may be different for programs that specialize in other disciplines such as finance or human resources. Once sufficient cases are available for study another program type that could be examined is the pre-experience MS program, which is gaining in popularity outside the United States. The standard predictors used in this study were GMAT scores and UGPA, but additional predictors may provide more insight into the admission process for these and other program types. The group comparisons involved only a limited subset of the studies based on who included the information in their dataset, so arguably the comparison back to the complete dataset may be viewed with some skepticism. Future studies could extend the available research by adding more information for groups and ensuring the generalizability to additional program types and different program locations.

Contact Information

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

References

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

Hecht, L., Manning, W., Swinton, S., & Braun, H. (1989). *Assessing older applicants for admission to graduate study in management: The role of GMAT scores and undergraduate grades*. McLean, VA: Graduate Management Admission Council.

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

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

Kuncel, N., Hezlett, S., & Ones, D. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*, 162–181.

Kuncel, N. R., 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 and Education, 6*, 51–68.

Oh, I., Schmidt, F., Shaffer, J., & Le, H. (2008). The Graduate Management Admission Test (GMAT) is even more valid than we thought: A new development in meta-analysis and its implications for the validity of the GMAT. *Academy of Management Learning and Education*.

Siegert, K. (2007). *Predicting success in graduate management doctoral programs* (GMAC Research Report RR-07-10). McLean, VA: Graduate Management Admission Council.

Stolzenberg, R., & Relles, D. (1985). Calculation and practical application of GMAT predictive validity measures. *GMAC Occasional Papers.* McLean, VA: Graduate Management Admission Council.

Talento-Miller, E., & Peyton, J. (2006, April). *Moderators of the accuracy of self-report grade point average*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Talento-Miller, E., & Rudner, L. (2008). The validity of Graduate Management Admission Test scores: A summary of studies conducted from 1997 to 2004. *Educational and Psychological Measurement, 68*, 129–138.

Talento-Miller, E., Rudner, L., Owens, K., Guo, F. (2006, July). *Assessing validity by combining examinees across programs*. Poster presentation at the Fifth Conference of the International Test Commission, Brussels, Belgium.

Thomas, D., Hughes, E., & Zumbo, B. (1998). On variable importance in linear regression. *Social Indicators Research*, *45,* 253–275.

Zwick, R. (1993). The validity of the GMAT for the prediction of grades in doctoral study in business and management: An empirical Bayes approach. *Journal of Educational Statistics, 18*, 91–107.

