GMAC®

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

Which Programs Have the Highest Validity: Identifying Characteristics that Affect Prediction of Success[1]

Eileen Talento-Miller & Lawrence M. Rudner

GMAC® Research Reports • RR-05-03 • August 23, 2005

Hundreds of studies are conducted to determine the predictive validity of admissions tests. Combined studies on the validity of the Graduate Management Admission Test® (GMAT®) show that scores can be used to predict performance in a variety of programs. However, at the individual program level there can be considerable differences in the amount of validity observed. Though variation would be expected from different measurements of the same construct, not all of the variation among observed validity values across studies can be attributed to statistical artifacts. This suggests actual program differences are contributing to the effectiveness of GMAT® scores in prediction, which is consistent with previous research suggesting there may be characteristics of programs studied that explain observed differences. Data on predictive validity was collected from 163 studies with complete data conducted through the Validity Study Service (VSS) of the test sponsor, the Graduate Management Admission Council® (GMAC®). For each program with available validity data, information was gathered that described characteristics of the program and its students. Discriminant analyses were used to determine if the program characteristics could correctly classify cases into groups based on the amount of validity observed. The resulting validity coefficients were close to normally distributed with a mean of 0.506 and a standard deviation of 0.132. The best discriminators were related to program structure, admissions, and, to some extent, content. This information can be used to determine situations in which levels of predictive validity will generalize.

Introduction

Concurrent with the need to verify the validity of an instrument is the need to evaluate its efficacy in different situations. It is for this purpose that local validity studies are conducted. Admission tests, for example, are often validated for each individual program using the test. However, when it is not feasible to conduct a study, program staff rely on being able to generalize the results from other studies to their own situation. To determine whether or not any given program can generalize results of previous studies to their situation, it must first be understood to what extent program differences can affect observed validity. This study examines whether program factors can be used to identify studies that have exceptionally high validity.

The Graduate Management Admission Test® (GMAT®) has been in use since 1954 to aid in admission decisions for graduate management programs such as MBA programs. From the first instances of its use, validity evidence has been collected to evaluate the ability of the test to predict performance in the core business courses generally found in the first half of the program (Hecht & Schrader, 1986; Olsen, 1957). Over the hundreds of individual validity studies conducted in the past 50 years there has been considerable variation in the observed validity coefficients for the predictors and combinations of predictors, including GMAT® scores and undergraduate grade point average (GPA). A recent meta-analysis by Kuncel, Crede, and Thomas (2004) summarized available validity evidence for the GMAT® exam and undergraduate GPA while correcting for statistical artifacts such as sampling error, range restriction and criterion unreliability. While they actually reported high validity, they also showed the statistical artifacts were predictive of graduate school performance. The variance of the corrected validity coefficients, however, showed there were differences among the validity for the individual studies that could not be accounted for by the

[1] An earlier version of this paper was presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Canada, April 12–14, 2005.

statistical artifacts alone. The Kuncel study did not address any reason for the remaining variation in observed results across schools.

Previous research suggests that program differences, such as program type, can have an effect on the validity observed in a study. Analyses suggest that predictive validity of GMAT® scores and undergraduate GPA is higher for Executive MBA (EMBA) programs compared to other program types and lower for part-time programs compared to other program types (GMAC®, 2003; Talento-Miller, 2004). Program type is just one of the variables that can be identified that may explain some of the differences in observed validity for individual studies.

This paper seeks to determine if exogenous factors can explain differences in validity observed for individual schools or programs. The exploratory meta-analysis will examine the characteristics of programs and their effects on levels of validity coefficients resulting from predictive validity studies of the GMAT® exam.

Methods

Data Collection

The first set of data came from summaries of the 264 validity studies conducted from 1997 to 2003 through the Validity Study Service (VSS) at GMAC®. The VSS, which has been in operation since 1977, accepts data from individual schools and programs to provide them with validity evidence of GMAT® scores and undergraduate GPA for performance in their program. Because of the requirements of the VSS and the analyses conducted, the variables collected from the many studies have uniform characteristics (Wightman & Leary, 1985; Zhao, et al., 2000).

Information on each program was gathered from two published guides to business schools: Miller, 1999 and *Peterson's guide to MBA programs*, 1998. Schools that were not included in these guides were not included in the study to insure the consistency of data collected. The information collected from these guides was coded as program variables and information about enrolled students. Data related to the program were business school enrollment, average class size, number of full-time faculty, full-time to part-time faculty ratio, amount of tuition (out-of-state, if amounts differed), and selection ratio, defined as number of applicants to number enrolled. The variables about students in the programs were percent of female students enrolled, percent of foreign students, mean student age, and percent of students with the maximum criterion score (a 4.00 GPA).

Due to missing data or inconsistency in the dataset, some cases had to be removed. One criterion for removal was lack of available data on school characteristics. There were a few cases removed whose GPA criterion scale was not consistent with the other cases. To maintain independence, multiple datasets from the same program were removed. Only the most recent study was included for any particular program within a school. The final data set included information from studies for 163 programs within 154 schools and included summary information from each of the studies: the means, standard deviations, and correlations of GMAT® scores, undergraduate and graduate grade point average. These variables, together with the program characteristics, represented the independent variables, or predictors, for the present study.

The multiple correlation of graduate GPA with GMAT® verbal and quantitative scores and undergraduate GPA was used as the measure of predictive validity for each study. Each of the simple correlations was corrected for restriction of range in order to calculate a corrected multiple correlation. The restriction of range correction was based on the 2001–2002 GMAT® score and GPA information from all students who sent their scores to the program studied. Stolzenberg and Relles (1985) argued that since sending GMAT® scores to a program represents a step in the admissions process, then individuals sending their scores to a program represent the full population of potential applicants. Thus, this correction yields the correlation coefficients the school programs could expect to see if they admitted all their applicants. The corrected validity coefficient for each study was classified into three relatively equal-sized groups based on validity levels. Validity around 0.3 to 0.4 is considered good for admissions tests (Kaplan & Sacuzzo, 1997); if validity is adjusted for restriction of range, values above 0.4 would be considered good (Talento-Miller & Rudner, 2005). The three groups in this study could be described as "moderate," "high," and "exceptionally high" validity. For clarity, though, these groups will be referred to as low, middle, and high validity levels.

Data Analysis

Discriminant analysis was used to identify differences in characteristics for programs with differing levels of validity. To insure clear differences, the middle validity category was removed and the analyses classified cases into low or high validity. Principal component analysis was used to group variables to determine if there was an underlying structure in the predictors that could be helpful in describing validity results. Discriminant analyses were performed with principal components and also with the individual predictor variables predicting low and high classifications.

Results

Validity coefficients were computed by regressing mid-program grades on GMAT® Verbal Scaled Scores (GMAT-V), GMAT® Quantitative Scaled Scores (GMAT-Q) and Undergraduate Grade Point Averages (UGPA) for each of the 163 programs in this study. The regression was computed from the correlation matrix with mid-program grades corrected for range restriction. The validity coefficients were close to normally distributed with a mean of 0.506 and a standard deviation of 0.132. There was a slight negative skew (–0.106) and a slight positive Kurtosis (0.127). The programs were then divided into three equal-sized groups based on their within program validity. Table 1 provides descriptive statistics for the three validity-level groups and the total distribution.

Table 1. Descriptive Statistics for the Validity Coefficients by Validity-Level Group

Validity Level	N	Mean	Std. Deviation	Median	Interquartile Range
≤0.444	55	0.368	0.0747	0.391	0.338–0.425
0.445–0.563	54	0.506	0.0347	0.499	0.475–0.542
≥0.564	54	0.649	0.0656	0.630	0.598–0.687
Total	163	0.506	0.1302	0.498	0.434–0.599

The validity levels for all three groups are quite respectable. The lowest group has a median validity of 0.391; the top group a median validity of 0.630. The interquartile ranges are fairly small. A one-way analysis of variance confirms that the groups have significantly different validity levels ($F = 290.3$, $p < 0.05$). A post hoc analysis using Scheffé shows that all pairs of validity-level groups are significantly different at the $p < 0.05$ level.

A principal component analysis was performed on the program and study characteristics. Three variables were extracted accounting for 51.3% of the variance. The varimax rotated structure, shown in Table 2, provides three well-defined components. The first component provides for characteristics of the student body. This includes demographic characteristics, class size, and GMAT® scores. The second component provides for program characteristics, including tuition, selection ratio, and percent of non-US students. The third component provides for grades and includes both undergraduate and mid-program grade point averages.

Table 2. Rotated Factor Structure			
Predictor Variable	Component		
	I	2	3
Business School Enrollment	0.205	**0.565**	0.061
Percent Female	**–0.635**	0.083	0.312
Percent Non-US Students	0.175	0.453	–0.060
Mean Student Age	–0.259	–0.150	0.198
Number of Full-time Faculty	**0.547**	0.378	0.248
Mean Class Size	**0.712**	0.230	0.237
Out-of-State Tuition	0.099	**0.683**	–0.047
Percent with 4.00 GPA	–0.278	**–0.570**	0.059
Selection Ratio	**0.667**	0.450	–0.119
FT to PT Faculty Ratio	0.122	**–0.560**	–0.029
UGPA Mean	0.249	0.043	**0.858**
UGPA Standard Deviation	–0.074	0.041	**0.893**
GMAT-V Mean	**0.869**	0.205	–0.020
GMAT-V Standard Deviation	**–0.752**	0.185	0.023
GMAT-Q Mean	**0.807**	0.361	0.012
GMAT-Q Standard Deviation	**–0.830**	–0.148	0.006
Mid-Program GPA Mean	–0.227	–0.350	**0.509**
Mid-Program GPA Standard Deviation	–0.019	–0.277	0.020

Note:
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization. Rotation converged in four iterations.

The remaining analysis compares the lowest validity-level group with the highest validity-level group. There are statistically significant mean differences between these two groups ($p < 0.05$) on several variables. The highest validity group had a lower percentage of female students ($F = 7.59$), a lower percentage of students with a 4.0 ($F = 8.39$), a higher number of full-time faculty ($F = 8.23$), larger class sizes ($F = 7.74$), a higher selection ratio ($F = 6.63$), and a higher mean GMAT-V score ($F = 4.02$). The data were analyzed using listwise deletion. Using only programs with complete data did not meaningfully alter the group validity statistics. The resultant group sizes were 42 and 46 for the low and high validity groups, respectively.

A discriminant function analysis predicting high versus low validity group membership based on the three component factor scores, however, was not significant (Canonical correlation = 0.288, Wilk's Lambda = 0.917, $\chi^2_3 = 7.333$, $p > 0.05$). Thus, composite factor scores representing program, student, and grading characteristics were not able to differentiate high versus low validity-level programs.

A discriminant function analysis using all of the variables was able to adequately predict high versus low program validity-levels (Canonical correlation = 0.641, Wilk's Lambda = 0.590, $\chi^2_{18} = 40.683$, $p < 0.05$). With equal priors, the function correctly classified 80.7% of the programs. The standardized coefficients, shown in Table 3, indicate that the best discriminators, controlling for the presence of the other variables, are mean UGPA, mean GMAT-Q, number of full-time faculty, mean class size, and percent of students with a GPA of 4.0. Negative coefficients are associated with higher validity-level group programs. Thus, in the presence of each of these variables, the student body characteristics associated with higher validity programs are more full-time faculty, larger mean class size, and lower mean GMAT-Q scores. The program characteristic most associated with higher validity levels is fewer students with a GPA of 4.0. The grading characteristic most associated with higher validity levels is low mean UGPA.

Table 3. Standardized Discriminant Function Coefficients

Variable	Coefficient
Business School Enrollment	0.347
Percent Female	0.503
Percent Non-US Students	0.146
Mean Student Age	–0.178
Number of Full-time Faculty	**–0.719**
Mean Class Size	**–0.704**
Out of State Tuition	0.133
Percent with 4.00 GPA	**0.679**
Selection Ratio	–0.320
FT to PT Faculty Ratio	0.201
UGPA Mean	**0.860**
UGPA Standard Deviation	–0.339
GMAT-V Mean	–0.158
GMAT-V Standard Deviation	–0.549
GMAT-Q Mean	**0.769**
GMAT-Q Standard Deviation	–0.040
Mid-Program GPA Mean	–0.213
Mid-Program GPA Standard Deviation	0.189

Discussion

In seeking to identify characteristics that will lead to levels of validity for different studies, several possibilities were examined. Although characteristics could be grouped through the principal components analyses, these components were not meaningful in terms of clearly identifying differences among studies with low and high validity. Based on the variables examined, there was no identifiable general structure related to the prediction of high versus low validity. The variables that were strongest in the subsequent discriminant analysis represented aspects of each of the previously identified components.

Because the best discriminators did not fall easily into the descriptions of the principal components, the variables themselves need to be examined to determine how they might contribute to classifying the cases. In terms of admission requirements, lower average UGPA and GMAT-Q values were characteristic of programs with high validity. It is interesting to note this relationship in the presence of GMAT-V mean, for which higher values are associated with high validity. If the lower values on the two admission factors indicated a less selective school, then one would expect that the selection ratio variable would reinforce that notion. However, since selection ratio is defined as the number of applicants to the number of enrolled students, and since for the analysis higher selection values were associated with high validity, then the lower values for the admission factors is not related to selectivity. The implication is that validity values are higher for business school programs that are popular but do not have the strenuous requirements of some of the most visible business programs.

That the more selective programs, represented by higher selection ratios, have high validity may seem counterintuitive. One would expect that a highly selective group would have remarkably similar characteristics, or low variance. However, since variables were corrected for restriction of range, this statistical artifact was effectively removed. The fact that selection ratio was not one of the best discriminators suggests that the correction did not unduly influence the analysis.

Though the admission requirements may not be as strenuous, the required courses are not as easy for the programs with high validity as measured by the percent of students receiving the maximum 4.00 GPA. One would expect large numbers at the ceiling to attenuate the validity of a study, so it is not surprising that high validity is associated with a lower percentage of ceiling scores. Although not as many students received the maximum GPA, grades were high as evidenced by higher mean mid-program GPA related to high validity. Taken together with information about admissions and the fact that GPA is based on a combination of classes, it may be that while students performed well in most classes, a few more difficult classes prevented them from achieving a perfect GPA.

With regard to program characteristics, more full-time faculty and higher average class sizes were associated with high validity. These findings may appear contradictory, as it would seem if more faculty were available, then class sizes should be smaller. This finding would be understandable if the business school were large, necessitating a large number of faculty and large class sizes, however this information is in the presence of lower business school enrollment. One possible explanation that would account for greater numbers of full-time faculty would be specialization of faculty members, which would require more people to teach available classes in the different subject areas. The low full-time to part-time faculty ratio for high validity suggests large numbers of faculty altogether, not just full-time. In the sources used, the faculty are specified as graduate business faculty, but may be teaching in programs other than the MBA. Multiple programs would necessitate greater numbers of faculty while not affecting class size. A look at the mean class size variable revealed that while most values hovered around the median of 27 students, there were some extreme values at the upper end which may have influenced the prediction.

Taken all together, it appears that high validity values are associated with programs that are popular, do not have particularly strenuous admission requirements, have high grades within the program but not extremely so, and possibly have multiple program types which require large numbers of faculty. Taken another way, one might expect that programs with strenuous admission requirements, or programs with a large ceiling effect, high percentages of students with 4.00 averages in the program, to have lower validity, though one might not expect to observe both within the same program. Considering that levels of validity studied represented only the extremes of the group

after the middle level was removed, then it is apparent that if these effects stacked within a program studied, then the low (or high) validity would be almost assured.

This study found some factors that influence the level of observed validity. There were no clear patterns in the best discriminators based on the principal components. The original description of predictors that classified variables as student characteristics, program characteristics, or study characteristics found more effects from the latter two than from the former. In general, discriminating factors appeared to have more to do with the program structure and design than with any factors having to do with the student body or specifics of the validity studies. The UGPA and GMAT-Q mean values were related to admission to the program; number of full-time faculty and mean class size were related to the structure of the program; and percent receiving a 4.00 GPA can be related to program content as well as program structure. Some variables that might be expected to affect validity, such as standard deviations of admission characteristics or selection ratio, may have had a reduced effect when examined in the presence of the other variables.

There were limitations to this study. Because of the method chosen to gather information about programs, predictor variables were limited in scope and definition and some studies were excluded due to lack of available data. The number of validity studies was reduced, since including schools that had conducted multiple studies on the same program would have resulted in an overrepresentation of their program information. As such, the final sample was small relative to the number of predictor variables used. This study used existing validity studies of programs that chose to take part in the VSS conducted by GMAC®. Future studies may want to target programs that have particular characteristics.

If characteristics of school programs have an effect on the predictive validity of an admission test for the school, then each school who wants to include that admission test in its application process will have a better idea of the impact of the test based on the school's unique characteristics. The findings of this study add to the validity generalization literature, as it can define the situations to which a level of predictive validity will generalize.

Contact Information

For questions or comments regarding study findings, methodology or data, please contact the GMAC® Research and Development department at research@gmac.com. For additional information specific to the Validity Study Service at GMAC®, contact GMAC® staff at vss@gmac.com.

Acknowledgements

Veronica Garcia, Research Writer/Editor, GMAC®, writing and editorial services.

References

Graduate Management Admission Council®. (2003). *How effectively does the GMAT® predict the academic performance of Executive MBAs?* Unpublished manuscript.

Hecht, L. & Schrader, W. (1986). *Graduate Management Admission Test®: Technical Report on Test Development and Score Interpretation for GMAT® Users*. McLean, VA: Graduate Management Admission Council®.

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

Kuncel, N., Crede, M., & Thomas, L. (2004). A comprehensive meta-analysis of the predictive validity of the Graduate Management Admission Test® (GMAT®) and undergraduate grade point average (UGPA). Presented at the Annual Meeting of the Society for Industrial and Organizational Psychology, April 2-4, Chicago, IL.

Miller, E. (1999). *Barron's guide to graduate business schools* (11th edition). Hauppauge, New York: Barron's Educational Series, Inc.

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

Peterson's guide to MBA programs. (1998). Princeton, NJ: Peterson's.

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. (2004). *Predictive validity for part-time MBA programs*. Unpublished manuscript.

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

Wightman, L.E., & Leary, L.F. (1985). *GMAC® Validity Study Service: A three-year summary.* McLean, VA: Graduate Management Admission Council®.

Zhao, J.C., Patsula, L., Stone, E., Paez, R., Babula, J., Tye, E., & Cruz, D. (2000). *GMAT® Validity Study Service: A Three-year Summary (1996-97 through 1998-99).* McLean, VA: Graduate Management Admission Council®.