GMAC®

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

When Are Grades No Longer Valid? A Look at the Effect of Time on the Usefulness of Previous Grades

Eileen Talento-Miller and Fanmin Guo

GMAC® Research Reports • RR-09-05 • February 1, 2009

A great deal of controversy exists regarding the use of standardized tests for high stakes purposes, including for admissions to higher education (Kuncel & Hezlett, 2007; Zwick, 2007). One argument against testing for this purpose is that information about previous experience, such as grades and transcripts, provides good quality data about candidates that is relevant to program performance. A question about grades, however, is whether their relevance fades with time. In other words, do grades earned in classes taken just a few months or years ago show the same amount of predictive validity as grades earned in classes taken several years or even decades before starting a given program? As more and more adults return to higher education after several years out of the classroom, answering this question becomes increasingly important. In particular, programs such as those for Master of Business Administration (MBA) degrees that often require minimum amounts of work experience (Dreher & Ryan, 2000), generally receive older transcripts for applicants compared with other programs. For these applicants, how well do grades predict their performance in upcoming academic pursuits?

The validity of grades is well established for predicting performance in both undergraduate and postgraduate programs. For undergraduate programs in the United States, predictive validity studies looking at both previous grades and admission test scores show consistently higher correlations with grades compared with test scores (Burton & Ramist, 2001; Geiser & Santelices, 2007). This relationship reverses when one looks at postgraduate programs; however with less contribution to prediction from previous grades and higher validity from admission test scores (Kuncel & Hezlett, 2007, Talento-Miller & Rudner, 2008). It is possible, for MBA programs at least, that the "age" of previous grades has some moderating effect on their contribution to prediction.

A few studies have examined the effect of time variables such as age, amount of work experience, or time since undergraduate degree in the prediction of success in postgraduate management programs. The findings are mixed, with some studies suggesting that older, more experienced students perform better in these programs, controlling for other variables (Ahmadi, Raiszadeh, & Helms, 1997; McClure, Wells, & Bowerman, 1986), and other studies showing inconclusive results (Arnold, Chakravarty, & Balakrishnan, 1996; Dreher & Ryan, 2000; Graham, 1991; Paolillo, 1982; Talento-Miller, 2004). In the Ahmadi et al. (1997) study, where age was a significant predictor, the average age of the students was less than 30 years old. There is some concern among programs that cater to an older group that the typical admission factors might not be as useful in predicting performance.

Executive MBA (EMBA) programs, which are designed for individuals with extensive business experience, have expressed concern about using the same admissions factors as MBA programs designed for a less experienced group. In the study by Arnold et al. (1996), which examined an EMBA program, the authors suggested that admissions staff do not place as much emphasis on grades when the previous degree was earned more than 10 years prior to starting the program. Although 10 years between the undergraduate and postgraduate program might seem to be an extreme case, it actually could be quite common for EMBA programs. Gropper (2007) stated that five to eight years of professional experience was

the minimum requirement for typical EMBA programs, and in his study, the average amount of work experience of his participants exceeded 13 years. For this group, undergraduate grade point average (UGPA) was not a significant predictor of any of the program outcomes examined in the study (Gropper, 2007). Similarly, a meta-analysis of multiple studies suggested that the validity of UGPA for predicting program performance tends to be lower for EMBA programs relative to full-time programs, even though the validity from admission test scores tends to be higher (Talento-Miller & Rudner, 2008). Based on the accumulated research, it appears that there is some effect of time on the utility of grades in predicting performance.

Interactions have been observed between age and performance, as well as the validity of UGPA. A study by Gayle and Jones (1973) suggested that the effect of age on prediction was not linear, and in fact, reversed direction at age 30. For students up to 30 years old, age was positively related to performance, but after 30 the effect of age resulted in lower grades in the postgraduate program. In a similar fashion, the prediction found from previous grades differed between the two age groups, with UGPA being positively related to performance for the younger group but having a negative effect on prediction for students older than 30 (Gayle & Jones, 1973). Similarly, Palmer and Wright (1999) showed that while UGPA was a significant positive predictor of performance for students 32 years old and younger, the effect of UGPA was not significant for students over 32. In a large-scale study across several programs, Hecht, Manning, Swinton, and Braun (1989) found that grades were less useful for predicting performance of students 34 or older than for those 24 or younger.

The purpose of the current study is to determine whether the age of grades will affect their ability to predict future performance. For admissions, this information is critical in determining how other criteria fit into the decision process. Although this is particularly important for postgraduate business programs that have a work experience requirement, the findings would apply to any type of program that accepts applicants with a wide range of backgrounds and uses previous grades as an important admission component.

Study

Methodology

This study uses data collected through the piloting of new assessments to measure aspiring postgraduate business school applicants' knowledge in specific content areas. The pilot was conducted to assess the quality of items written to create assessments in the subjects of accounting, finance, and statistics. The items were organized into nine forms—three for each of the three content areas—containing approximately 40 items each (ranging from 39 to 43). Subsequent analyses of the forms indicated good reliability for both internal consistency (0.73 to 0.87) and alternate forms (0.68 to 0.87). The means and standard deviations of the item difficulty indices were almost identical for the nine forms; therefore, no equating of the scores from different forms was performed, although common items were included across the nine forms for equating, if necessary. In addition, content validity was assured through the use of content experts in each of the three subject areas. Because of this, it was determined the scores from these forms were adequate measures of current performance in each of these content areas and could be compared with previous grades to address the effect of time.

The new content assessments were geared toward those intending to enroll in postgraduate business programs. As such, the pilot participants were recruited from individuals registered to take the Graduate Management Admission Test® (GMAT®) exam, a common admission requirement for postgraduate business schools around the world. A random sample of recent registrants was extracted from the list of available candidates, and invitations to participate in the study were sent via email. Participants could take any or all of the nine available forms and were offered incentives for completion that included a condition that they exhibited a good faith effort to answer the questions to the best of their ability. In addition to completing the test forms, participants were asked to complete an optional background questionnaire that included a few demographic questions in addition to questions about educational experience. Specifically, the survey asked

participants for the grades they received in accounting, finance, and statistics courses and the date they took their most recent course in each subject.

A total of 727 participants completed 2,506 test forms. A review of the demographics of the participants showed they were similar to the population of GMAT examinees, suggesting limited response bias in respondents versus nonrespondents. Most respondents completed the background survey questions, although not all had taken courses in the subjects listed. The number of participants with course grades and dates for each of the subject areas ranged from 241 to 295. Tables 1 and 2 contain information about the pilot participants.

Table 1. Demographic and Educational Characteristics of Participants Compared with GMAT Examinees in 2005–2006

		Frequency	Percent	GMAT Examinees
All Respondents		***727***	***100.0%***	***204,509***
Gender				
	Male	443	60.9%	60.5%
	Female	273	37.6%	39.0%
Native Language				
	English	437	60.1%	53.6%
	Non-English	273	37.6%	41.5%
Highest Degree				
	Bachelor's or lower	528	72.6%	68.4%
	Some graduate study or a graduate degree	181	24.9%	19.3%
Undergraduate Major				
	Business	286	39.3%	44.9%
	Humanities	44	6.1%	4.4%
	Science	147	20.2%	21.3%
	Social Science	69	9.5%	13.5%
	Other	159	21.9%	3.8%
Intended Program Type				
	Full-Time	531	73.0%	52.8%
	Part-Time	108	14.9%	23.6%
	Executive MBA	36	5.0%	4.7%
	Undecided	28	3.9%	7.0%
Intended Concentration				
	Accounting	77	10.6%	9.3%
	Finance	195	26.8%	15.7%
	General Management	121	16.6%	9.0%
	Other	232	31.9%	42.6%
	Undecided	77	10.6%	9.7%
Previous Course Experience				
	Accounting	278	38.2%	Not Available
	Finance	244	33.6%	Not Available
	Statistics	196	40.7%	Not Available

Table 2. Summary Descriptive Statistics for Pilot Participants			
	N	Mean	Standard Deviation
Age	708	26.6	5.1
Undergraduate GPA[1]	674	3.58	0.42
Accounting grade	273	3.44	0.60
Finance grade	241	3.53	0.55
Statistics grade	296	3.43	0.62
Accounting score	418	0.659	0.153
Finance score	333	0.673	0.142
Statistics score	365	0.612	0.140
[1] All grades and grade-point averages are on a 4.0 scale.			

Data Analysis

Correlation analyses were used to determine to what extent grades in subject courses would predict performance on that subject's test forms. Course grades were self-reported on a four-point scale. Reported values ranged from 1.0 to 4.0. Test performance was based simply on proportion correct on each form and ranged from 0.12 to 0.95. Because participants could take multiple forms, they could have up to three test scores in a single subject area. Each of these test scores was included as a separate outcome, so some individuals were represented multiple times within the data for the particular content areas. The alternate form reliability analyses that were conducted previously did not suggest there would be radical differences in scores across forms; however, the use of the same grades to predict different test scores may have affected the correlations due to the lessened variability. Presumably, this effect would be consistent across the time variable being examined for the study and should not affect the outcome of the central question.

The amount of time since grades were given was calculated as the number of months from the reported date of the course to the date of the pilot. Two participants reported their last course was after the date of the pilot, which generated negative values for months since their grade was received. Once these participants were removed from the analyses, the range of time since grades were received ranged from 0 to 243 months with median values ranging from 27 to 46 months among the content areas. All three of the content area distributions of time since grades were received were positively skewed.

Because the time distributions were skewed, it was unlikely the relationship with test performance would be linear. Correlations of time with scores, although in the expected direction of lower performance from older courses, were nearly zero, with values ranging from -0.042 to -0.081. It was therefore decided to create grouping variables for time. Groups were created within each content area for recent (0 to 24 months), moderate (25 to 60 months), and older (61+ months) grades. These particular groupings allowed for relatively robust group sizes with sample sizes ranging from 82 to 192. While the groupings could be considered arbitrary, they can be described in terms of the time needed to satisfy typical work experience requirements for MBA programs, with the recent group only qualifying for postgraduate business programs requiring little or no experience, and the older group meeting the five-year minimum requirement of most EMBA programs.

Correlations between test performance and previous grades were calculated within time groups for each content area. The correlations were then compared among groups as well as compared with correlations using all data. Values were evaluated for both practical and statistical significance.

Results

Table 3 presents overall correlations and those for each of the time groups. The highest correlations were observed for the recent time group, with higher values

than overall data as well as the other time groups. Even with smaller sample sizes, the values for the recent group were statistically significant at the 0.005 level. Marked differences existed in the correlations for the older grades group compared with the other groups; correlations were relatively small and not statistically significant with $p > 0.05$, for all content areas. For the moderate group—those receiving grades around two to five years before taking the tests—values were statistically significant for accounting and finance, but not for the statistics group. Across all three content areas, correlations decreased across the time groups, with older grades showing lower correlations with only one exception. For the statistics content area, both the moderate and older time groups showed negative non-significant correlations, though the values did not decrease across time. Because of this, and because negative correlations between grades and test scores are counterintuitive, the findings for the statistics content area should be further explored in future research.

Table 3. Correlations of Grades with Test Scores Overall and for Time Groups (Sample Size)

	Overall	Recent	Moderate	Older
Accounting	0.242** (n = 418)	0.314** (n = 192)	0.303** (n = 113)	0.064 (n = 113)
Finance	0.324** (n = 331)	0.488** (n = 151)	0.366** (n = 98)	0.191 (n = 82)
Statistics	0.064 (n = 367)	0.255** (n = 122)	-0.117 (n = 100)	-0.017 (n = 142)

**Correlation is significant at alpha = 0.005.

Discussion

The results of the study suggest that older grades are not as useful in prediction as those earned more recently. In particular, grades that are more than five years old had arguably no relationship to test performance for accounting and statistics content areas, and a relatively low relationship for finance. If these findings generalize to an admission setting, then the implication would be that programs requiring five or more years of work experience or those with a large proportion of applicants out of school for a long period of time would need to rely on other factors, such as admission test scores, to obtain valid predictions of future program performance. Many MBA programs, for instance, require post-baccalaureate work experience. In EMBA programs, it is not unusual to see the average student spend 5 to 10 years in business careers before enrolling in postgraduate classes. It would be vitally important for such programs to evaluate carefully the effect of their admission requirements on performance to determine the extent to which older grades are able to predict performance and whether additional predictors might need to be considered more heavily.

Although the question of the utility of grades is very important for admissions purposes, the current study did not examine predictive validity in an admission context. This study used grades in specific content areas, as opposed to grade point averages, and the outcome variable was a test score, again as opposed to grade point averages, which differs from the typical dependent variable for admissions validity studies. Since most validity studies do not examine data indicating how old previous grades are, future research should be designed to take this information into account when examining the influence of various admission factors on performance. The current study design operated under the assumption that the relationship of time would not be strictly linear, which is something to consider for future research. The time groups assigned could be considered arbitrary—studies may want to examine if an "ideal" time period can be identified in which grades are most useful.

There were other important limitations to this study. For one, grades were self-reported. A study by Talento-Miller and Peyton (2006) showed that although self-reported grades were generally accurate, a small relationship existed between accuracy of self-report and time since graduation. Since this is central

to the question being addressed in this study, it would be prudent to obtain more reliable grades in future studies. The current study included only potential postgraduate business school students. The issue at hand may apply, however, with varying degrees of relevance, to any program—undergraduate, postgraduate, or professional—that may have applicants who have been out of school for a long time. Additional research can address how these findings generalize to different settings.

The findings of the study are not really surprising: Over time, people forget. Academic content, at least in these subjects, is not "like riding a bike—once you learn, you don't forget how." The key for future research would be to learn how best to utilize this information to get accurate and fair pictures of how students with older grades will perform in future academic pursuits.

Study 2

One of the limitations of the previous study was the circumspect connection between the data used and the admission process. To remedy that, an additional study was designed specifically to address the use of UGPA in admissions and the ramifications in the prediction of later performance. The use of combined grades (UGPA), as opposed to grades for a single subject, and the outcome more relevant to validity studies of admission data (program grade point average) allows for extension of previous findings. In addition, this study examined the effect of time on the validity of GMAT scores to determine if results would be similar to what was observed for grades.

Methods

This study uses data described in Talento-Miller and Peyton (2006) that combines data from validity studies conducted through the Graduate Management Admission Council® matched to a GMAT examinee dataset containing responses to background information questions answered when the examinee either registered for or finished taking the exam. The validity study data included information on admission factors, such as GMAT scores, UGPA, and the first-year grade point average (FYA) earned within the program. Year of graduation and year the GMAT exam was taken were gathered from the GMAT examinee database. Because the validity data were based on more than 100 different studies from individual graduate business programs around the world, differences in grading scales could affect the measurement of relationships, as FYA was standardized within programs, after programs containing fewer than 20 cases in the matched dataset were eliminated. Because of the large number of undergraduate programs represented, it was not possible to standardize UGPA; therefore, values that were greater than the typical United States maximum of 4.0 were also eliminated from the database. This left 7,264 cases from 90 different programs with enough information on the relevant variables to include in the analyses.

Age of UGPA and GMAT scores was estimated based on the available information. The validity study database included the year the study was conducted. The validity studies could include students who were admitted across multiple years or could have included data older than that present year's class. Nevertheless, this variable was used as the best proxy to determine approximate admission date. Because the outcome variable was first-year average, the admission date was theorized to be exactly one year earlier than the study year. Study years ranged from 1996 to 2003 and GMAT score dates ranged from 1997 to 2002. Graduation dates were listed from the 1970s and earlier, but any values beyond 2002 were not included, with the assumption that those respondents misunderstood the question. Age of UGPA upon admission was then calculated as estimated admission year minus graduation year. Age of GMAT scores was calculated similarly by subtracting exam year from estimated admission year. Age groups were then defined similar to groups in Study 1, although groups had to be defined relative to years rather than months. Recent values were determined to be up to two years old, moderate values ranged from three to five years, and older values were six or more years old. Because of the dates included in this study, it was not possible to analyze GMAT scores six or more years old at admission. Table 4 shows the sample sizes in each group for UGPA and GMAT scores.

Table 4. Frequency of Variable Age Groups by Admission Factors				
	Overall	**Recent**	**Moderate**	**Older**
UGPA	7,264	1,981	2,305	3,018
GMAT-Total	7,253	6,570	683	—
GMAT-Verbal	7,242	6,563	679	—
GMAT-Quantitative	7,232	6,556	676	—
GMAT-AWA	5,806	5,284	522	—

Data analyses included simple correlations and multiple regression analyses. Correlations of each of the admission factors of UGPA and GMAT scores with standardized FYA (Z-FYA) were calculated overall and by variable age group. In addition, multiple regression analyses were used to determine how the age of either UGPA or GMAT scores affected the prediction. For UGPA, GMAT scores were entered first into the regression followed by UGPA, and the change in variance explained was compared across time groups. Similarly, for GMAT scores, UGPA was entered first and the change in variance explained was assessed across the GMAT age groups. For the regression equations, the GMAT predictors used were the scores on the verbal, quantitative, and analytical writing assessment (AWA) sections. Because GMAT total score represents performance on the verbal and quantitative sections, it was not included as part of the regression analyses.

Results

The patterns seen in the correlations in Table 5 suggest that, regardless of whether UGPA or GMAT scores are considered, as the admission factors age, their relationship with the outcome variable becomes weaker. Although GMAT scores older than five years could not be analyzed and scores older than two years were relatively rare, the trend of smaller relationships held across all variables. All correlations are statistically significant, in large part due to the very large sample sizes available for each of the groups. A relevant question would be whether the reductions across groups could be considered practically significant. For instance, in moving from the recent to the moderate age group for UGPA, there is a reduction of less than 2% of variance explained, but moving from moderate to older results there is a reduction of greater than 3% of variance explained. Moving from recent to moderate age groups for GMAT scores reduced variance explained by almost 3% for both the total and verbal scores. For AWA scores, however, the reduction was less than 2%, and for quantitative scores it was less than 1%. To determine how these variables interacted, further analyses were conducted using multiple regression techniques.

Table 5. Simple Correlations[1] with Z-FYA by UGPA and GMAT Age Groups				
	Overall	**Recent**	**Moderate**	**Older**
UGPA	0.230	0.297	0.269	0.203
GMAT-Total	0.234	0.245	0.179	—
GMAT-Verbal	0.188	0.200	0.112	—
GMAT-Quantitative	0.206	0.210	0.193	—
GMAT-AWA	0.179	0.185	0.121	—
[1] All correlations are statistically significant at $p < 0.01$.				

The results of the regression analyses can be found in Table 6. Again, the pattern clearly shows that the change in variance explained is less when the predictors in question are among the older groups. In one set of regressions, UGPA added either more than 7% or less than 2% to what was explained by GMAT

scores, depending on the age groups of the grades while, in the converse analyses, more recent GMAT scores added almost 5% and older scores added almost 2% to what was explained by UGPA.

Table 6. Stepwise Regression Results for UGPA and GMAT Age Groups

	All groups	Recent	Moderate	Older
R^2 for GMAT	*0.064*	0.069	0.045	0.068
Δ from UGPA	*0.030*	0.074	0.042	0.017
R^2 for UGPA	*0.052*	0.050	0.064	—
Δ from GMAT	*0.043*	0.047	0.017	—

Discussion

The results of Study 2 support the results found in Study 1 by demonstrating that the usefulness of grades in predicting future performance is reduced as those metrics move further away from the time they were awarded. The context is far more relevant to the admission situation and extends to include not only grade point averages, but also test scores.

Study 2 also suffered from a number of limitations, however. The data used for the validity studies were reported from schools and are likely far more reliable than self-reported data; but some inconsistencies, such as out-of-range UGPA values or non-US students, required a portion of the data to be removed. The graduation date taken from the examinee database was self-reported and might have suffered from misunderstanding, as evidenced by the fact that some examinees reported graduation dates that would have occurred after their school's study period. Study date was used as a proxy to calculate age values based on admission, and the amount of error introduced by using this imperfect measure could not be estimated. The time periods for Study 2 were chosen to be consistent with Study 1 and therefore may not represent ideal time grouping, particularly for GMAT scores, since the majority of scores were less than three years old.

It is difficult to say how the results from data compiled across multiple programs would apply to individual programs. A scan of data from a handful of schools showed that some programs had very few cases in the recent UGPA age group. These would presumably be the programs that had a strict work experience requirement. The prevalence of these types of programs in the later age groups might have affected the results. It may be expected that EMBA programs are overrepresented in the older UGPA age group. Because program type was not available in the database, it is not possible to tease out the effects of programs in each of the groups. The post hoc nature of the current analyses makes it difficult to answer questions directly that can be better addressed by collecting data specifically designed for this topic.

Future research should be conducted to address the limitations of the current studies. Validity studies can be conducted for individual programs, including information on age of grades and other admissions data such as test scores, or even ratings of the personal recommendations for candidates. Individual studies can then be combined to determine how different program types such as full-time MBA and EMBA programs are alike or different. More accurate measures of age might allow examination into whether there is a threshold that can be identified as the time when grades are most useful as predictors. It is recommended that GMAT scores only be used if they have been taken within five years of applying. The current research supports that assertion by showing a trend of reduced usefulness for older GMAT scores. Similarly, the results for undergraduate grades point to the need for continued research on this topic to better advise admissions policies for various programs.

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

Ahmadi, M., Raiszadeh, F., & Helms, M. (1997). An examination of admission criteria for the MBA programs: A case study. *Education, 117*, 540–546.

Arnold, L., Chakravarty, A., & Balakrishnan, N. (1996). Applicant evaluation in an executive MBA program. *Journal of Education for Business, 71*, 277–283.

Burton, N., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (Research Report 2001-2). New York: College Entrance Examination Board.

Dreher, G., & Ryan, K. (2000). Prior work experience and academic performance among first-year MBA students. *Research in Higher Education, 41*, 505–525.

Gayle, J. B. & Jones, T. (1973). Admissions standards for graduate study in management. *Decision Sciences, 4*, 421–425.

Geiser, S., & Santelices, M. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes.* (Research & Occasional Paper Series CSHE.6.07). Berkeley, CA: University of California, Center for Studies in Higher Education.

Graham, L. (1991). Predicting academic success of students in a master of business administration program. *Educational and Psychological Measurement, 51*, 721–727.

Gropper, D. (2007). Does the GMAT matter for executive MBA students? Some empirical evidence. *Academy of Management Learning and Education, 6,* 206–216.

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®.

Kuncel, N., & Hezlett, S. (2007). Standardized tests predict graduate students' success. *Science, 315*, 1080–1081.

McClure, R. H., Wells, C. E., & Bowerman, B. L. (1986). A model of MBA student performance. *Research in Higher Education, 25*, 182–192.

Palmer, J., & Wright, R. (1999). The influences of student age and gender on the predictive validity of GMAT scores and undergraduate GPAs. *Educational Research Quarterly, 23*, 27–33.

Paolillo, J. (1982). The predictive validity of selected admission variables relative to grade point average earned in a Master of Business Administration program. *Educational and Psychological Measurement, 42*, 1163–1167.

Talento-Miller, E. (2004). *Predicting success in business school and beyond: An examination of the factors impacting graduate school performance and occupational success for registrants of the Graduate Management Admission Test®*. Unpublished doctoral dissertation. [University?]

Talento-Miller, E., & Peyton, J. (2006, April). *Moderators of the accuracy of self-report GPA*. Paper presented at the annual meeting of the American Educational Research Association, April 7-11, 2006, 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.

Zwick, R. (2007). College admissions in twenty-first-century America: The role of grades, tests, and games of chance. *Harvard Educational Review, 77*, 419–429.

