CAMBRIDGE ASSESSMENT

A report on the predictive validity of the BMAT (2004) for 1st year examination performance on the Veterinary Medicine course at the University of Cambridge

Joanne L. Emery

26 February 2007

UNIVERSITY *of* CAMBRIDGE
Local Examinations Syndicate

Background and Methods

The BMAT 2004 was taken by 4529 applicants to medicine and veterinary medicine courses at the Royal Veterinary College, the University of Bristol, the University of Cambridge, the University of Oxford and University College London. Of these, 371 applied to course D100 (Veterinary Medicine) at the University of Cambridge.

First year examination data from 2006 were supplied by the University of Cambridge and matched by name and college to students' BMAT results. Matches were made for 63 out of 68 students and data were anonymised after matching. The examinations consist of three separate components whose marks are combined into an overall mark and classification (Class I, II, III or fail). The components are 'Homeostasis' (HOM), 'Molecules in Medical Science' (MIMS) and 'Veterinary Anatomy and Physiology' (VAP). Spearman correlations were calculated between BMAT scores (Section 1: Aptitude and Skills, Section 2: Scientific Knowledge and Applications) and examination marks. Descriptive statistics and logistic regression analyses were carried out on the overall exam classifications. Hypotheses were one-tailed as positive correlations were expected.

There are problems with the correlation coefficient as a measure of predictive validity. This is discussed fully in Issue 3 of Research Matters (Bell, 2007). Correlation is the usual measure of predictive validity but tends to produce weak coefficients for various reasons where selection is concerned. The scores of accepted applicants often cover only a restricted range at the upper end of the measures and we cannot know how those rejected would have performed. Unreliability in outcome measures (such as subjectively-marked assessments) further weakens coefficients. Another caveat is that student selection is usually compensatory in nature, where a high score on one criterion can make up for a poor score on another, and this can cause prediction to appear absent or even negative. For such reasons, uncorrected coefficients of over 0.35 are considered very beneficial in predictive validity studies (see Table A).

Table A: Guidelines for interpreting correlation coefficients in predictive validity studies

Validity Coefficient	Interpretation
Above 0.35	very beneficial
0.21 to 0.35	likely to be useful
0.11 to 0.20	depends on circumstances
Below 0.11	unlikely to be useful

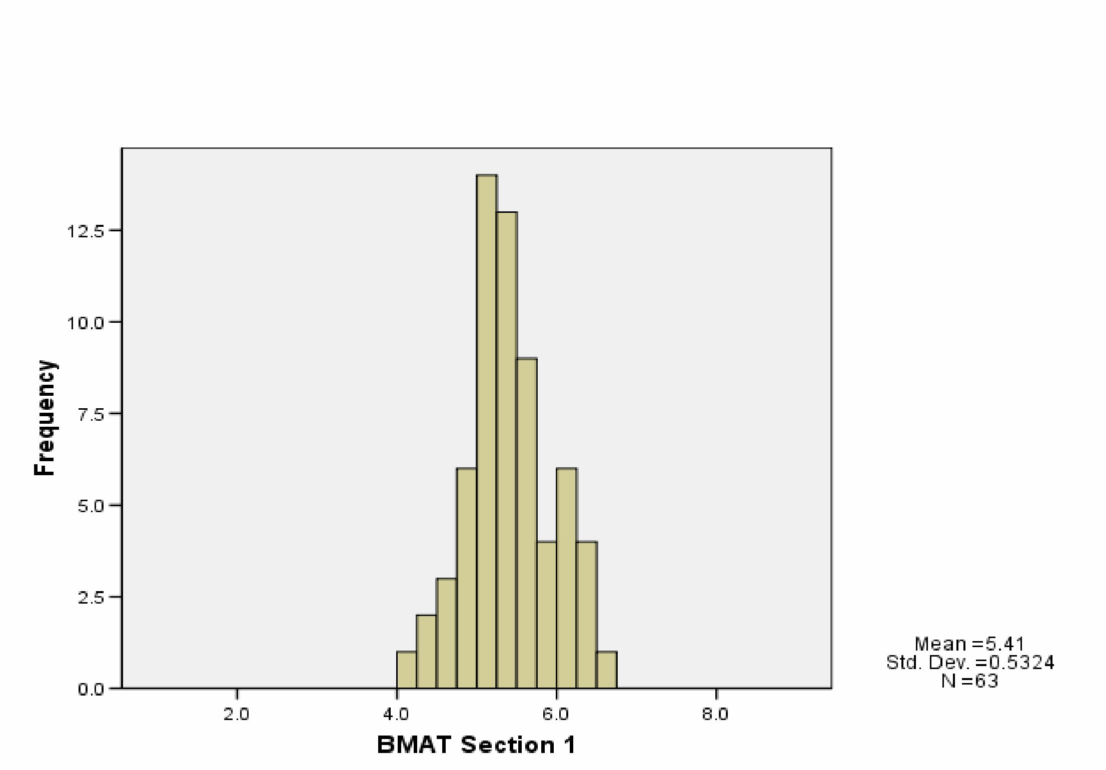
Source: US Department of Labor, Employment Training and Administration, 1999

There are corrective formulae for range-restriction and unreliability (Sackett and Yang, 2000) but these are only appropriate in particular circumstances and should not be routinely applied. The complexity of the selection process here (a compensatory mix of qualitative and quantitative information) makes them inappropriate. Thus simple, uncorrected coefficients are presented throughout this report. These may be underestimates of the true strength of the predictive relationships.

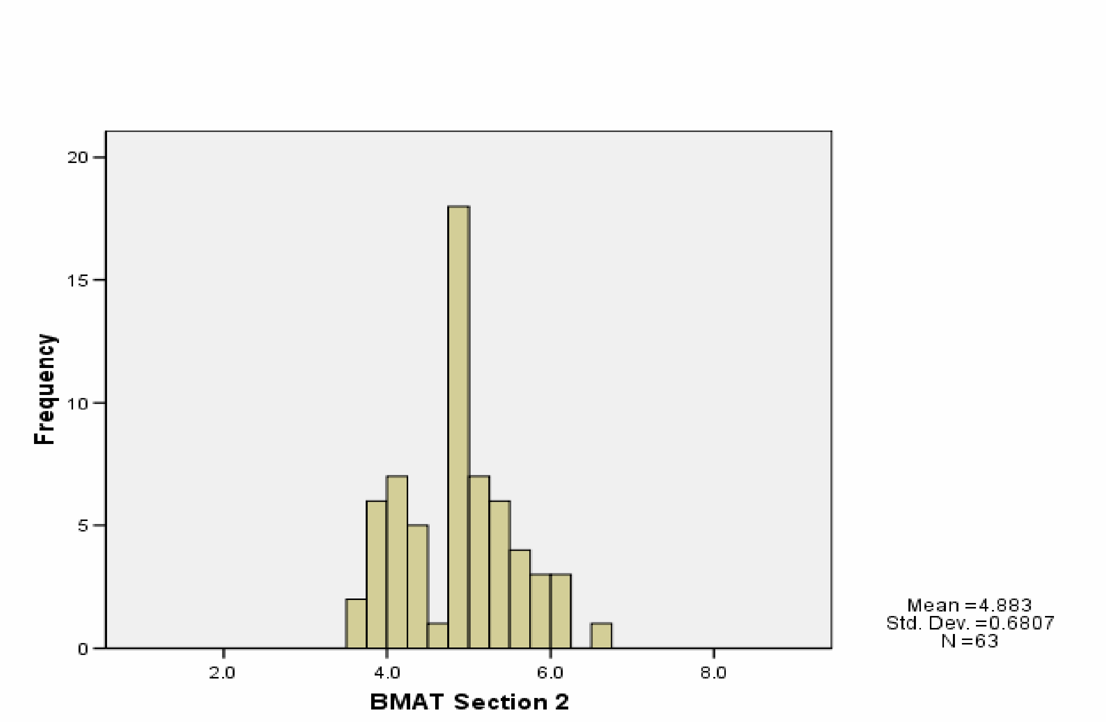
Results

Figure 1: Distributions of BMAT 2004 scores for students accepted onto course D100 (Veterinary Medicine) at the University of Cambridge

a) Section 1 – Aptitude and Skills



b) Section 2: Scientific Knowledge and Applications



The distributions of BMAT scores in Figure 1 show that performance was, on average, slightly better for Section 1 of the test than for Section 2 in this (the BMAT is designed to have a mean score of 5 for both sections). Ranges of scores are similar for both BMAT sections but a greater number of students scored above 6 on Section 1. Some candidates are admitted with relatively low scores on the BMAT due to the compensatory nature of the selection process.

Table 1: Spearman correlation coefficients between BMAT 2004 scores and 1st year examination performance

			BMAT Section 1	BMAT Section 2
Spearman's rho	HOM	Correlation Coefficient	.214*	.212*
		Sig. (1-tailed)	.046	.047
		N	63	63
	MIMS	Correlation Coefficient	.193	.153
		Sig. (1-tailed)	.065	.116
		N	63	63
	VAP	Correlation Coefficient	.209	.209
		Sig. (1-tailed)	.052	.051
		N	62	62
	total	Correlation Coefficient	.211*	.215*
		Sig. (1-tailed)	.048	.046
		N	63	63

*. Correlation is significant at the 0.05 level (1-tailed).

Table 1 shows that correlations for the HOM and VAP components, and for the overall total mark, are in the ‘likely to be useful’ range. Those for the VAP component, however, just miss statistical significance. Sections 1 and 2 appear to correlate equally well with examination performance and both correlate with the total mark at the 0.05 level of significance.

It should be noted that the strength of correlations typically varies between institutions and cohorts in predictive validity studies. For this reason, caution should always be exerted in citing a single number as a test’s coefficient. It should also be noted that correlation coefficients are weakened by range restriction, unreliability, compensatory selection and non-linear relationships. Thus weak coefficients do not necessarily imply the lack of a predictive relationship.

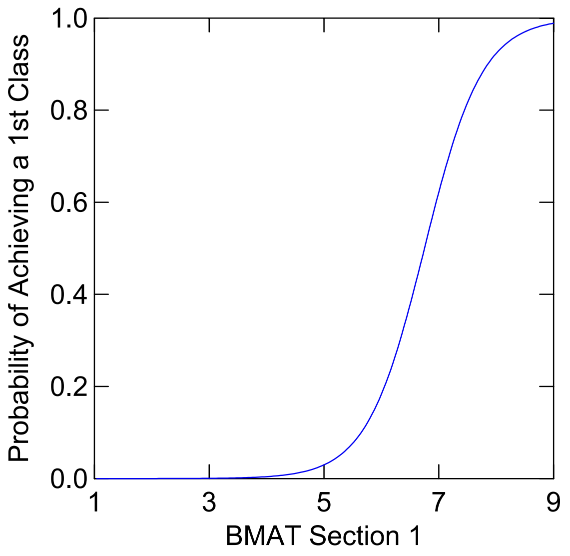
Table 2: Descriptive statistics of BMAT scores categorized by examination outcome

			Statistic	n
BMAT Section 1	I	Mean	5.87	6
		Std. Deviation	.44	
	II	Mean	5.37	43
		Std. Deviation	.51	
	III	Mean	5.33	13
		Std. Deviation	.61	
BMAT Section 2	I	Mean	5.60	6
		Std. Deviation	.79	
	II	Mean	4.80	43
		Std. Deviation	.63	
	III	Mean	4.81	13
		Std. Deviation	.68	

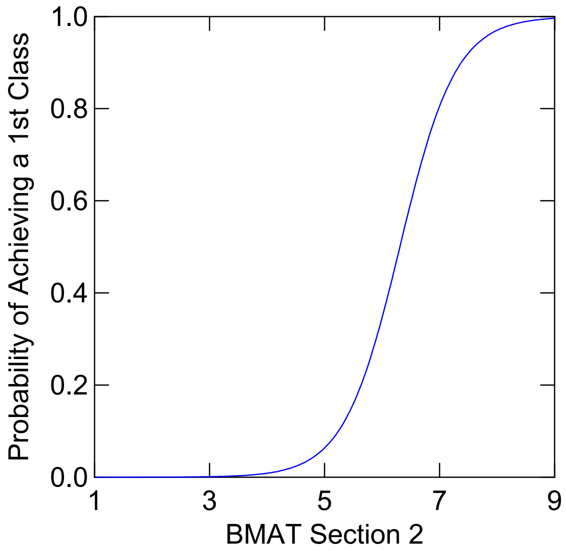
Table 2 suggests that higher BMAT scores were obtained, on average, by those who went on to perform better in their examinations (overall classification). Those who achieved a 'distinction' had the highest mean BMAT scores for both Sections 1 and 2 (numbers are low here, however, and should be interpreted with caution). Those who went on to fail their examinations had the lowest mean scores for Section 1 of the BMAT but not for Section 2. This is not unusual and reflects the possibility that many students who fail do so for reasons other than their academic ability.

Figure 2: Logistic regression plots showing the probability of achieving a Class I outcome as a function of BMAT scores

a) Section 1: Aptitude and Skills



b) Section 2: Scientific Knowledge and Applications



Logistic regression analysis predicts the probability of a discrete outcome (here the probability of achieving a Class I outcome in the exams) as a function of predictor variables. The steepness of the slope directly implies the strength of the relationship and a flat, horizontal function implies no predictive relationship. These functions should be interpreted with reference to Figure 1 which displays the actual range of BMAT scores achieved by the cohort (nobody scored above 7 on either test section).

Approximately 10% of students were awarded a Class I and the probability of this is strongly predicted by their BMAT performance. Both BMAT sections similarly predict the highest exam outcome but the slope is slightly steeper for Section 2 of the test when scores are very high (over 6). This means that increasing scores for Section 2 make the greater difference to the probability of achieving a Class I. Those with scores of around 5 on Section 2 have approximately a 0.05 probability of achieving a Class I whilst those scoring 6 have a probability of around 0.3.

Conclusions

Results suggest that the BMAT predicts examination performance in this particular cohort. Correlations, although statistically significant in most cases, do not appear very strong by traditional standards. This is a common finding in validity research and the reasons for this can be range restriction, compensation, unreliability in outcome measures (e.g. for subjectively-marked questions) or non-linear relationships. However, scores on the BMAT clearly predict the chances of obtaining an excellent examination outcome. An obvious restriction on the predictive validity of entrance tests is that the reasons for poor performance at university are many. An entrance test can only predict those who *can*: not necessarily those who *will*.

References

Bell, J F. (2007) Difficulties in evaluating the predictive validity of selection tests. *Research Matters: A Cambridge Assessment Publication*, **3**:5-10.

Sackett PR, Yang H. (2000) Correction for range restriction: an expanded typology. *Journal of Applied Psychology*, **85**:112-8.

US Department of Labor, Employment Training and Administration (1999) Employers' Guide to Good Practices. Accessed at www.doleta.gov