The case against the BMAT: Not Withering but Withered?

John F Bell
Research Division
Cambridge Assessment
1 Hills Road
Cambridge
CB1 2EU
bell.j@cambridgeassessment.org.uk

Competing Interests: JFB is employed by Cambridge Assessment, which developed and administers the BMAT.

Summary Points

The BMAT does predict performance on university examinations.

The BMAT can only be classified as an intellectual aptitude test if every academic test is so classified.

McManus *et al.* selected evidence that supported their case while failing to provide a full account of the studies they cited.

The thinking skills that are assessed by the BMAT are used by doctors in both practice and research.

In the September 2005 issue of the BMJ there is a paper by McManus *et al.*[1] that was described in *The Guardian* as placing aptitude tests for medical admissions under withering academic fire. The argument of McManus *et al.* is summarised as follows:

"Intellectual aptitude tests are seen as a way to stretch the range, continuing to differentiate when A level grades are at their ceiling. The problem is that although these tests undoubtedly provided variance (as indeed would random numbers), it is not useful variance since it does not seem to predict performance at university."

Does this argument apply to the BioMedical Admissions Test (BMAT)[2], a subject-specific admissions test (see Table 1)? The test results are intended to be a significant component of the selection decision, to be used in conjunction with past examination performance, evidence from the UCAS form, and performance in interview. The BMAT's purpose is solely to predict potential in an academically demanding undergraduate biomedical degree, and not fitness to practice medicine or veterinary medicine. It is designed to discriminate among able applicants, including those who have achieved the highest possible A-level grades.

Table 1: Structure of the BMAT

Section 1	Generic skills often used in undergraduate study, including Problem Solving, Understanding Arguments, Data Analysis, and Inference Abilities.
Section 2	Material normally encountered in non-specialist school Science and Mathematics courses (i.e. up to and including National Curriculum Key Stage 4, Double Science and Higher Mathematics).
Section 3	Choice of three short-stimulus essay questions of which one must be answered.

To understand the claims of McManus *et al.* about admissions tests in general and the BMAT in particular, it is useful to evaluate them using the sort of skills assessed in the first section of the BMAT. The argument can be expressed as a syllogism:

Aptitude tests do not predict degree performance (major premise).
The BMAT is an aptitude test (minor premise).
Therefore the BMAT does not predict degree performance (conclusion).

Their conclusion requires that both of the premises are true. If one or both are false, then the conclusion cannot be drawn (although it could still be true). However, if the conclusion is false, then one or both of the premises must be false. In this paper, it will be demonstrated that both of the premises and the conclusion are false.

Is it true that aptitude tests do not predict degree performance?

The summary points of McManus *et al.* state, "Since the 1970s, university achievement has been shown to be predicted by A-levels but not by intelligence tests", and three examples are given. Firstly, McManus *et al.* quote the conclusion of Choppin *et al.*'s report on the ISPIUA project[3] (a trial of a test similar to the US SAT) that the "TAA appears to add little predictive information to that already provided by GCE and school assessment." It is reasonable to conclude that this test does not add predictive information beyond the A-level. However, when taken alone it does provide valuable predictive information. In fact, Choppin *et al.* do not entirely reject the use of their test but state that that it could be valuable in certain fields of study. In their supplementary material, McManus *et al.* present a table derived from Table 12 of Choppin *et al.*, which gives the simple correlations between the first year performance (based on tutors' assessments) as 0.10 for the mathematics score and 0.15 for the verbal score. This choice of tables is interesting, given that the more important final degree results are presented in Table 24. The correlations are 0.29 for mathematical score (identical to the correlation for A-level grade) and -0.12 for verbal score; i.e., university achievement is predicted by some components of intellectual aptitude tests but not by others. Indeed, Choppin and Orr[4] do not reject the use of aptitude tests, concluding that, "Firstly, it seems likely that universities, and perhaps other institutions of higher learning, could make use of the test in special cases where candidates for admission could not offer the more usual academic credentials."

The two remaining studies[5][6] used the AH5 test of general intelligence, designed to assess ability in the upper part of the range. The content is very different from that of the BMAT. In Part I of the AH5,

items are based on four principles: directions (some of which are primarily verbal, some primarily numerical), verbal analogies, numerical series and 'similar relationships'. Part II consists of items exemplifying four types of principle: analogies, series, 'directions' (requiring mirror imagery and shape construction) and 'features in common'. In both studies, the components of a reasoning test do not perform uniformly, and there is a relationship between at least one of these components and some of the outcomes.

McManus *et al.* also state that, "A recent study of dental school admissions evaluated an aptitude selection test, carefully founded in the psychology of cognitive abilities and skill acquisition. The scores were of no predictive value for clinical achievement at the end of the course"[7]. This is indeed the conclusion given in the abstract on Pubmed (http://www.ncbi.n1m.nih.gov/entrez/query.fcgi?DB=pubmed). However, the conclusion in the main text of the paper also states, "The value of using the AA and PAT (the cognitive tests) as predictors of performance in the preclinical years is that admissions committees can selectively exclude those applicants who will most likely be unsuccessful, repeat a year, or spend a disproportionate amount of time in the laboratories."

Finally, McManus *et al.* also refer to experiments involving situational tests in Flemish medical schools[8] and state that the situational selection tests "…were found to predict final year outcomes better than tests of cognitive abilities." This is again a paraphrase of the abstract. However, it is difficult to justify this statement from the evidence in the paper. The intercorrelation between final first year score and the cognitive test score is 0.24 and the intercorrelation between final first year score and the situational test is 0.20. The authors do state that "…situational tests … significantly predict over traditional cognitive ability measures." However, further reading reveals this refers to a small increase in the explained variance after adding the situational variables. Another interesting feature is that the cognitive test score was obtained by combining the scores on four tests, a reasoning test (correlation with final first year score = 0.33), visual information processing (correlation = -0.03), memory association (correlation = 0.13) and pattern recognition (correlation = 0.11). Lievens and Coetsier argue that the situational test "may be a useful complement to traditional selection procedures". However, it is important to recognise that the reasoning test was their best predictor.

There is another relevant study: Mellanby *et al.*[9] investigated the relationship between the AH6 Group Test of High Intelligence[10] and degree performance at Oxford. They formed a scale by using subcategories derived from the marks on the final degree examinations and found that verbal scores correlated 0.341 with examination marks. The average A-level points score was 29.5 (i.e. this means at least 75% must have obtained AAA).

In a critical thinking question, candidates might be expected to conclude from this evidence:

- Cognitive abilities tests measure different traits, and correlations between these traits and university outcomes differ.
- The evidence suggests that carefully constructed or selected tests could have predictive validity for candidates without A-levels or high attainers when everybody has the same A-level score.

Is the BMAT an intellectual aptitude test?

At one point, McManus *et al.* suggest that the BMAT "will correlate highly with conventional IQ tests." Since all academic examinations correlate quite highly with each other, arguing that BMAT and conventional IQ tests are the same would be like claiming that GCSE English and GCSE mathematics are the same because they are highly correlated (0.78)[11]. It is necessary to consider the content validity of the test by showing how well the content of the test is drawn from the class of relevant situations or the domain of subject matter, and that the situations and the subject matter are relevant to the use to which the test is put. This is a value judgement but one that can be based on professional judgement. Unlike a general IQ test, the BMAT is designed specifically for medical admissions and the content has to satisfy the university staff directly involved in making admissions decisions. Anyone with experience of intelligence tests should have difficulty in concluding that the totality of the content of the BMAT is anything like that of an IQ test. In particular, Section 2 of the BMAT contains questions on biology, chemistry, physics, and mathematics at the level of Key Stage 4. IQ tests try to avoid questions that rely on specific subject knowledge.

Early in their paper, McManus *et al.* give three broad reasons why A-levels may predict outcomes in medicine: cognitive ability, substantive content and motivation. Later, however, they propose that "A-

levels predict university achievement mainly because they measure the knowledge and ideas that provide the conceptual scaffolding necessary for building the more advanced study of medicine. McManus *et al.* then go on to make the following analogy: "Intelligence in a builder, although highly desirable, is no substitute for a professional knowledge of the craft of house building." Nobody is arguing that professional knowledge is not important but are thinking skills so unimportant for a doctor? When patients visit their doctors, surely they expect them to use thinking skills to select relevant information, recognise analogous cases, and determine and apply appropriate procedures. In addition, they would also expect doctors to be well-informed and so when faced with, say a BMJ article, they would use their skills to understand arguments by identifying reasons and assumptions, detect flaws, and draw conclusions. All of these skills are equally important for doctors pursuing a career in research. These are just the sort of thinking skills assessed by Section 1 of the BMAT. The fact that the scores for the sections of the BMAT are reported separately means that admissions tutors are able to consider these skills separately from the science knowledge. This is not possible with A-levels. At the moment a very high proportion of applicants to prestigious medical schools go on to obtain three grade As at A-level. This means that A-levels do not predict performance effectively in the current grading system. Nevertheless, universities must make selection decisions. It seems reasonable to conclude that:

- The totality of the BMAT could not be described as an intellectual aptitude test unless the definition of an intellectual aptitude test is so general as to include almost any academic test.

Is it true that the BMAT does not predict degree performance?

Since both the major and the minor premise of the syllogism are not definitely true, it is not possible to deduce that the BMAT does not predict degree performance. However, it is also true that falsifying the premises does not prove that the BMAT does predict degree performance. However, it is not necessary to rely on deduction; it is possible to use direct evidence.

Although the BMAT is new and the first student entry selected using it have only just completed their first year examinations, it is very similar to its predecessor, the Medical and Veterinary Admissions Test (MVAT), which was used with applicants to the University of Cambridge. (Many of the items for the BMAT match the specifications for those used in the MVAT.) Evidence from the MVAT led to the adoption of the BMAT by other institutions. In the pilot year of the MVAT (at Cambridge), the scores were available but not all College admissions tutors used them (interestingly, no tutor, including those sceptical of the value of the test, accepted a candidate with a very low MVAT score). The results of the admission process are presented in Figure 1. Clearly, candidates with very low scores were not accepted, although candidates with high scores were not necessarily accepted (the test is meant to be only one source of information used in the decisions so a high score would not be expected to guarantee success). This provides evidence that the test could be used to *deselect* applicants (Cambridge usually interviews all UK applicants). A similar pattern of results to that shown in Figure 1 is observable for the 2004 Cambridge applicants and the BMAT.
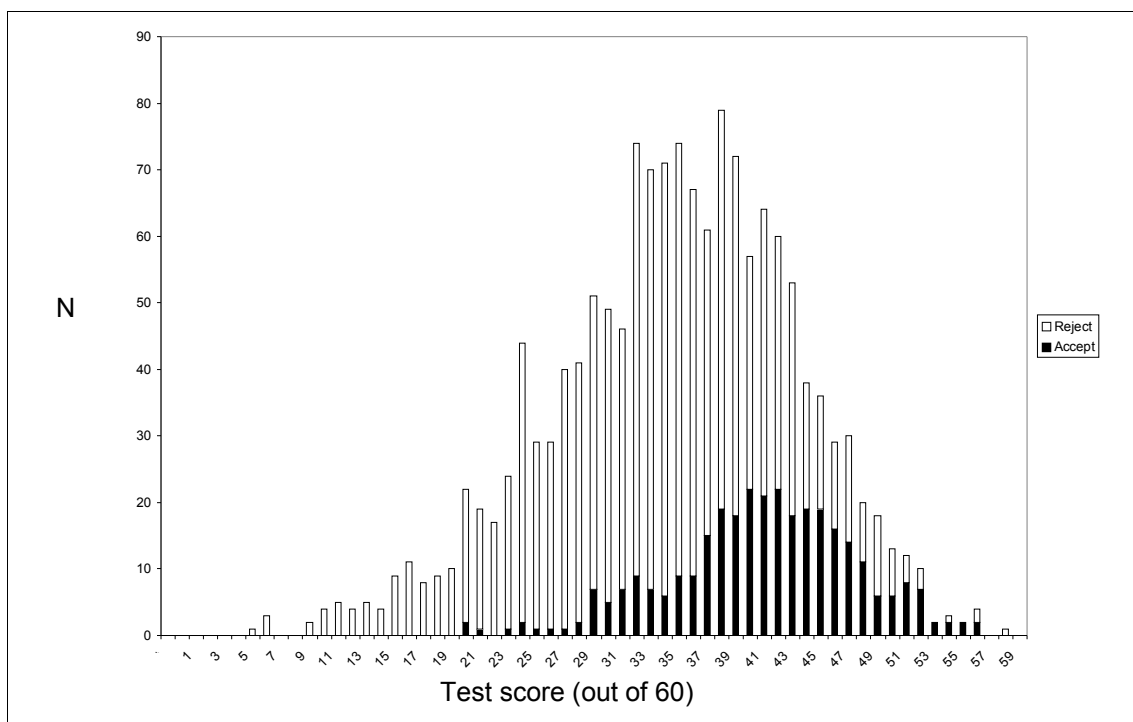
Figure 1: Distribution of MVAT pilot test scores for accepted and rejected applicants

It is now possible to consider the relationship between performance on the BMAT and first year medical examinations at Cambridge. Figure 2 shows the results of logistic regression modelling of achievement at (a) Class I, and (b) at least Class II, together with (c) a histogram of the distribution of the BMAT scores (Sections 1 and 2 total). Of the candidates entering Part 1A (first year) examinations in 2005, 227 students took the BMAT in 2003. Clearly, even with the very select entry at Cambridge, the BMAT is very effective at predicting the probability of obtaining Class I. The patterns observed in these graphs are very similar to those obtained with the BMAT's predecessor, the MVAT.
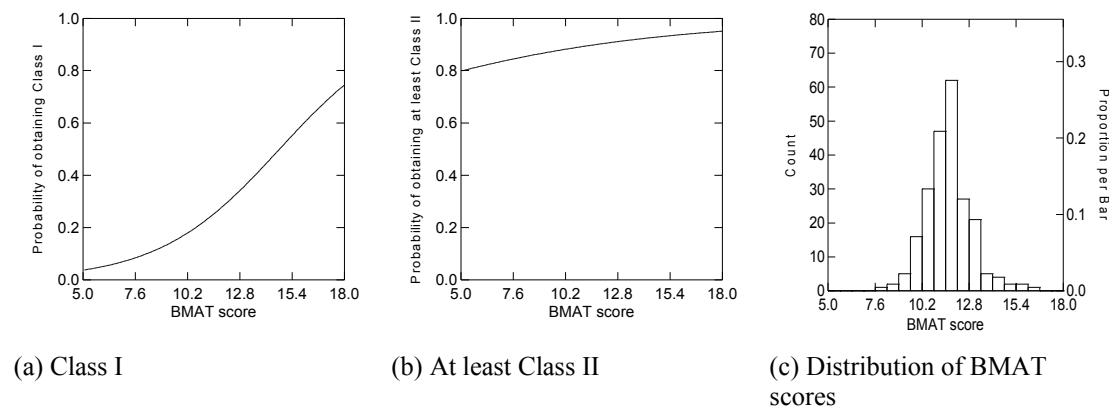

(a) Class I   (b) At least Class II   (c) Distribution of BMAT scores

Figure 2: Relationship between Part 1A degree performance and BMAT score

There are, of course, limitations to what an academic selection test can do. One of these can be summarised as: 'a selection test can only predict those who can, not those who will.' This is illustrated in Figure 3, which presents the relationship between the MVAT Section 2 score and the Molecules In Medical Science (MIMS) first year examination at Cambridge. (Note that this is the examination with the strongest relationship to the MVAT but is also the one on which students tend to perform worst.) It is possible to discern that the scatter of points is approximately triangular. Very few candidates with low MVAT scores obtained high marks on the MIMS examination (e.g. in category C for the MVAT

and a Class I on the MIMS). However, those with high MVAT scores are not necessarily successful. This is because there are many non-academic reasons for poor progress at medical school (e.g. illness, decline in motivation, discovery of the college bar). Some of these reasons could be detected by other selection methods but there are some that could never be predicted by any test. In addition, those teaching in the preclinical courses have reported that the introduction of the test has greatly reduced the number of students who, at the start of the course, find themselves struggling with basic science concepts.
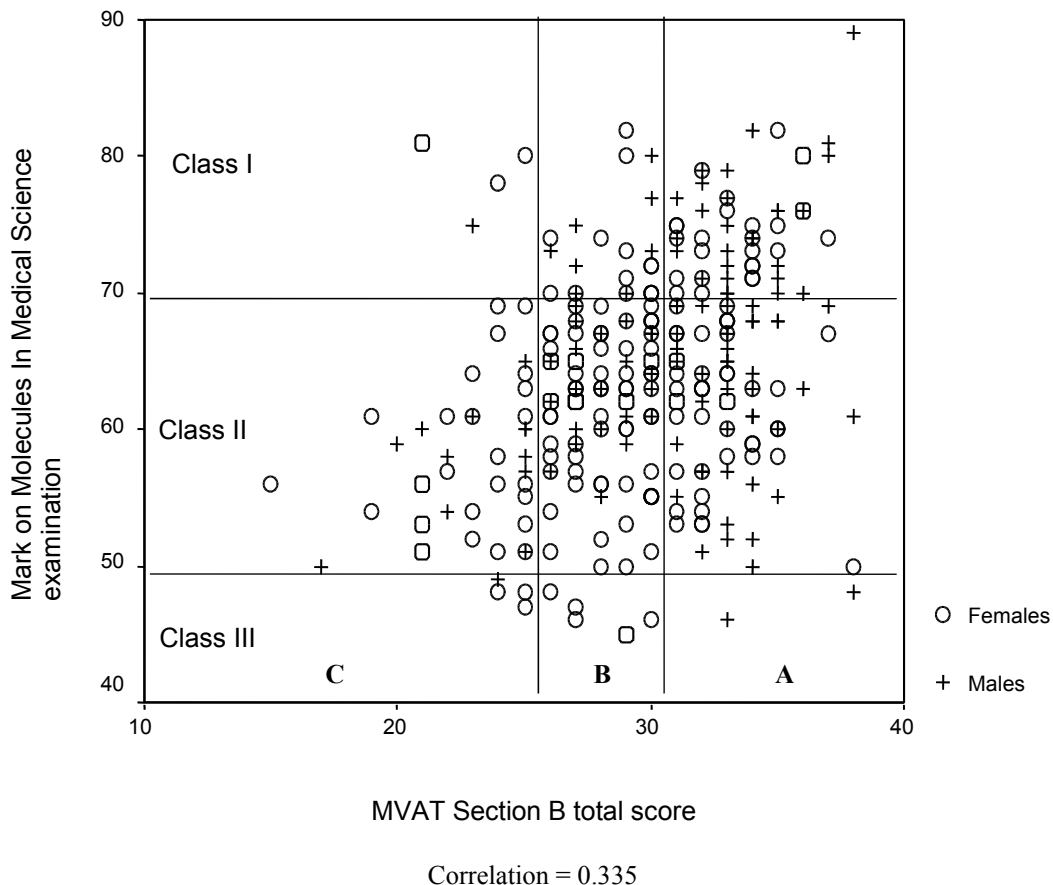

MVAT Section B total score

Correlation = 0.335

Figure 3: Molecules In Medical Science against MVAT Section B total

From the predictive validity evidence, the use of BMAT as a hurdle to decide on which applications proceed to interview is justified because applicants with low scores are rarely admitted (it is accepted that an individual applicant's circumstances might provide evidence that the BMAT scores underestimate their potential – if the BMAT team is informed of any problems then this information is passed with scores to the relevant medical schools). However, a high score should not guarantee a place in medical school (although all other things being equal the applicant with the high score is more likely to be successful).

Discussion

To refute the claims made in the original BMJ paper all that is needed is the evidence of predictive validity. However, it is important that it should be made clear that while it is entirely appropriate to ask about the predictive validity of a test, it is unreasonable to attack the admissions procedures of established and prestigious medical schools on the basis of a flawed argument based on the unrepresentative and incomplete selection of the available evidence. The summary points of McManus *et al.* state that, "Schools are introducing tests of intellectual aptitude without evidence of appropriateness, accuracy, or added value, making them open to legal challenge." This was not the case. The development of the MVAT and its, only slightly modified, successor, the BMAT, have been

subject to careful scrutiny by the participating institutions. The MVAT had a pilot year and has been evaluated every subsequent year.

There are three ways of using a selection test. The first is as a hurdle. The applicants have to reach a set level of attainment to either be selected outright or continue in the selection process. No one would advocate the former for medical admissions and there is evidence that the BMAT could be used for the latter (although it would be sensible to allow some candidates to proceed to a next stage if there are special circumstances). This is because in the pilot year, low scoring MVAT candidates tended to be deselected using other information and those who were selected had a lower probability of success. The second way of using a selection test is as one part of evidence in a compensatory system. This involves considering all the evidence and allows a higher rating on one selection criteria to compensate for a lower rating on another. Finally, there is a hybrid way with both hurdles and compensation. In practice, this is usually what happens in a selection process. Some applicants are rejected because of particular failings but for the remainder the decision process involves balancing individual applicants' strengths and weaknesses.

The analysis of the BMAT data would suggest that a candidate with a high BMAT score would not have problems with the academic difficulties of a medical course but could be rejected for weakness on other characteristics. The BMAT (and the MVAT) have been considered as providing only some of the evidence used to make a selection decision. There are other non-cognitive characteristics that need to be considered to improve the overall assessment of candidates.

[1] McManus IC, Powis DA, Wakeford R, Ferguson E, James D, Richards P. Intellectual aptitude tests and A levels for selecting students for medical school. *BMJ* 2005, 331:555-559.

[2] Cambridge Assessment. How to prepare for the BMAT. http://www.bmat.org.uk. (accessed 3 Oct 2005).

[3] Choppin BHL, Orr L, Kurle SDM, Fara P, James G. *Prediction of academic success.* Slough: NFER Publishing, 1973.

[4] Choppin B, Orr L. *Aptitude testing at eighteen-plus.* Slough: NFER Publishing, 1976.

[5] McManus, IC, Smithers E, Partridge P, Keeling A, Fleming PR. A-levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ* 2003, 327:139-42.

[6] McManus IC, Keeling A, Paice E. Stress, burnout and doctors' attitudes to work are determined by personality and learning style: A twelve year longitudinal study of UK medical graduates. *BMC Medicine* 2004;2:29.

[7] Gray SA, Deem LP, Straja SR. Are traditional cognitive tests useful in predicting clinical success? *J Dent Educ* 2002;66,1241-5.

[8] Lievens F, Coetsier P. Situational tests in student selection: an examination of predictive validity, adverse impact and construct validity. *Int J Sel Assess* 2002; 10:245-57.

[9] Mellanby J, Martin M, O'Doherty J. The 'gender gap' in final examination results at Oxford University. *British Journal of Psychology* 2000;91:377-390.

[10] Heim AW, Watts KP, Simmonds V *AH6 Group tests of High Level Intelligence.* Windsor: NFER-Nelson.

[11] Calculated from Cambridge Assessment data by author.