



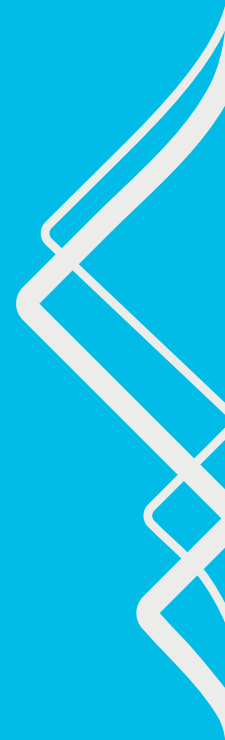
Cambridge Assessment
Admissions Testing

BMAT
BIOMEDICAL ADMISSIONS TEST

Applying the socio-cognitive framework to the BioMedical Admissions Test

Executive summary

August 2017

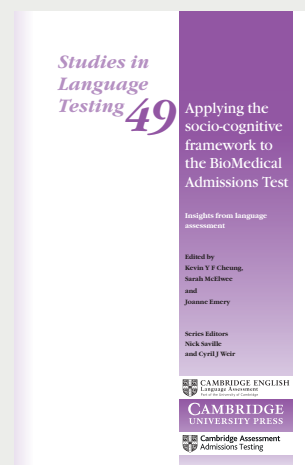


This summary gives an overview of BMAT research, and is intended for:

- higher education institutions that are considering the use of admissions tests and wish to validate assessments in a systematic way
- policy makers in medicine/healthcare professions
- researchers working in educational assessment or medical education.



The full research volume provides a comprehensive evaluation of BMAT's validity, including best practice on incorporating admissions testing into selection processes and common issues in admissions testing research.





In summary

Chapter 1 provides an introduction to:

- Applying the socio-cognitive framework to the BioMedical Admissions Test (the BMAT research volume) and how it was developed
- the BioMedical Admissions Test (BMAT), including how it is used in selection processes, the test format and the Cambridge Approach to test development
- the comprehensive socio-cognitive framework used to underpin the test's development and validation
- the chapters of the BMAT research volume: each focusing on an aspect of the socio-cognitive model.

The BMAT research volume

The BMAT research volume presents work on the test, conducted over more than a decade of research and validation. This is the first time this work has appeared together in one collection, made available for everyone.

This research evidence has been built over time, beginning at the test design and development stages and continuing since the test has been operational.

Why BMAT is used

BMAT is used by institutions worldwide, to help them select students with the right skills to succeed on medical, biomedical, dental and related undergraduate and graduate courses.

BMAT is developed, administered and marked by Cambridge Assessment Admissions Testing, part of a not-for-profit department of the University of Cambridge.

“ This book collects together, in large part for the first time, a body of evidence regarding admissions testing. The framework within which this has been done is robust and **few can be anything but impressed by the development processes, validation, contextual study and continual review and evaluation that Cambridge Assessment uses** in the BioMedical Admissions Test. ”

Martyn Partridge | *National Heart and Lung Institute, Imperial College London*

Summary of BMAT sections¹

Section 1: Aptitude and Skills	Tests the ability to apply key skills used in undergraduate study: <ul style="list-style-type: none"> • Problem-solving • Understanding argument • Data analysis and inference.
Section 2: Scientific Knowledge and Applications	Tests the ability to apply scientific knowledge typically covered in non-specialist school science and mathematics courses.
Section 3: Writing Task	Tests the ability to communicate effectively in writing, organising ideas and presenting them clearly and concisely.

How BMAT is used as part of the selection process

BMAT has been used by world-leading institutions since 2003. It is used to:

Select applicants who will succeed on the course

Our research shows a positive relationship between BMAT scores and on-course success. These research findings are presented in Chapter 6.

Support inclusion

BMAT measures a range of skills needed for success in higher education, giving applicants the opportunity to show their potential in addition to their school academic record.

Support internationalisation

BMAT provides a common measure for comparing applicants from a variety of educational backgrounds and with a wide range of qualifications.

¹ Download the full BMAT specification at: admissionstesting.org/bmat-preparation

“ One of the reasons we use BMAT is because we have a lot of applicants who have studied overseas with qualifications that we may or may not be familiar with. We have applicants with other degrees and applicants who have been out of education for a while. **All sorts of applicant categories, which are very hard to compare** to the standard GCSE/A Level applicant. ”

Pat Harkin | Deputy Director of Admissions and Associate Director of Student Support, University of Leeds

The chapters of the BMAT research volume

The BMAT research volume presents evidence for the validity of BMAT. Each chapter focuses on one aspect of validity. These are all component parts of overall validity, and each aspect of validity contributes cumulatively to the confidence associated with use of test scores.

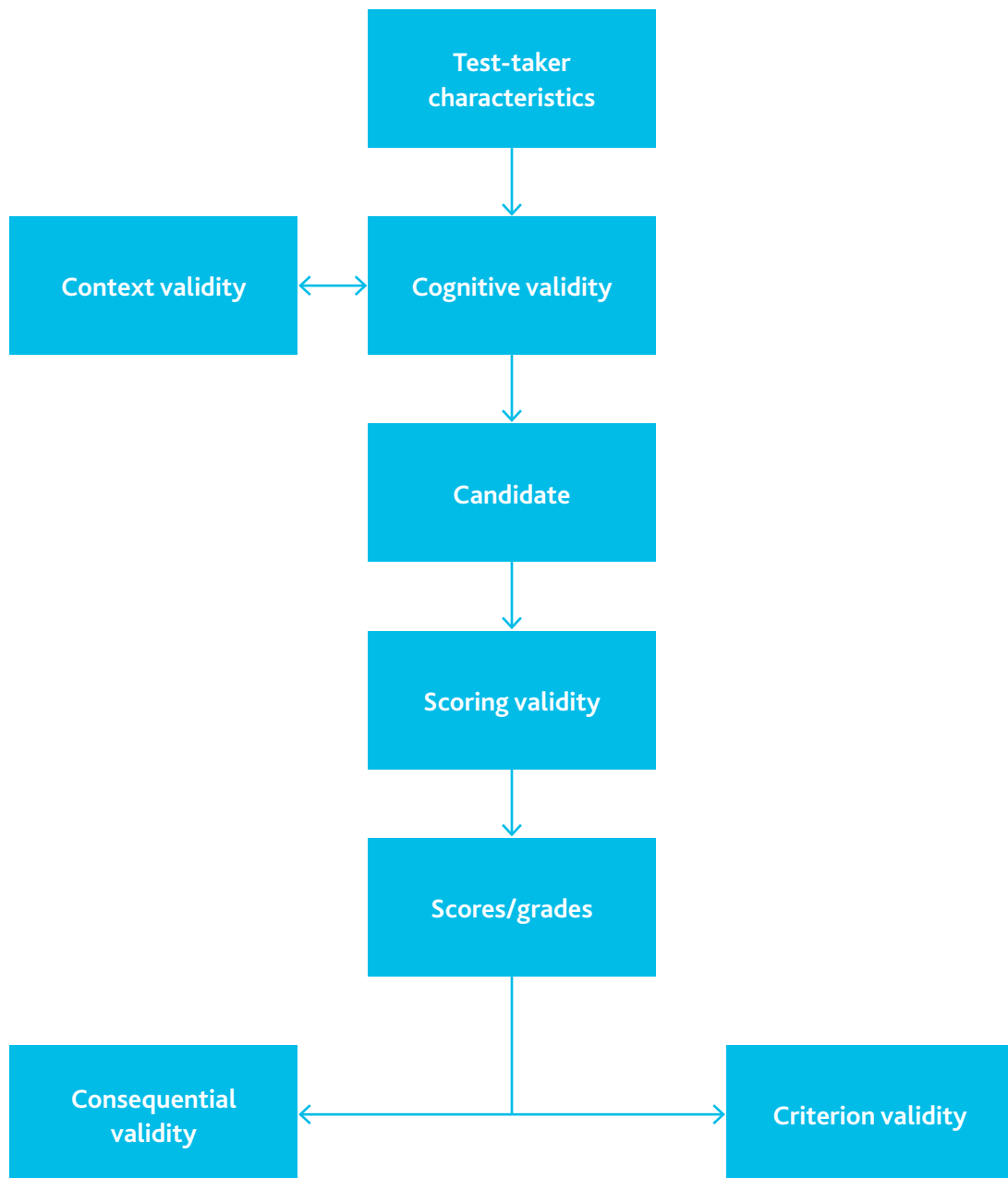
This multi-aspect model of test validity is known as the *socio-cognitive approach*.

Table 1: The socio-cognitive framework for test development and validation

The aspect of validity covered in each chapter	Explanation of each aspect of validity	Questions to be considered in building validity evidence
Chapter 2: Test-taker characteristics	Profiling intended test-takers to ensure tests are fit for purpose, fair and not biased in favour of any group of test-takers.	<ul style="list-style-type: none"> • What are the characteristics of the test-takers? (e.g. age, gender) • Does the test make suitable accommodations for candidates with special needs? • Are candidates sufficiently familiar with what they have to do in the test? • Are candidates put at ease so that they are enabled to achieve their best?
Chapter 3: Cognitive validity	Identifying the skills that are relevant to successful study.	<ul style="list-style-type: none"> • What are the skills/cognitive processes elicited by the test tasks?
Chapter 4: Context validity	Determining whether test tasks and administration conditions accurately measure what they are intending to measure and are fair and appropriate for all candidates.	<ul style="list-style-type: none"> • Is there any evidence that the response format is likely to affect performance? • Are the marking criteria explicit for the candidates and the markers? • Is the timing of each part appropriate? • Is the content knowledge suitable and unbiased? • Are the administration conditions satisfactorily consistent and secure?
Chapter 5: Scoring validity	Ensuring that tests are scored accurately and results are reliable (stable, consistent and free from errors of measurement).	<ul style="list-style-type: none"> • Are items of appropriate difficulty and do they discriminate between candidates? • Is there a sufficient level of test reliability? • Is there any evidence of item bias? • Are the candidates' responses their own? • Are there clearly defined marking criteria that cover the construct (the concept or characteristic that a test is designed to measure)? • Are markers trained, standardised, checked and moderated? • Is marking reliable and consistent?
Chapter 6: Criterion validity	Identifying the relationship between BMAT scores and other measures of performance.	<ul style="list-style-type: none"> • Do test scores relate to future outcomes? (predictive) • Do test scores relate to other tests or measurements? (concurrent)
Chapter 7: Consequential validity	Analysing the impact of the admissions test on individuals, institutions and society.	<ul style="list-style-type: none"> • How is the test perceived by stakeholders? • Are actions based on test scores appropriate? • Is there any evidence that test scores predict similarly across different candidate groups?

Figure 1 shows the direction of the relationships between different aspects of the socio-cognitive framework. While all aspects of validity need to be considered at test development stages, some types of validity evidence cannot be collected until after the test event – particularly those aspects which relate to the effects and consequences of using the test results.

Figure 1: The socio-cognitive framework used to underpin BMAT's development and validation



Chapter 2

Medical school, biomedical sciences and dentistry applicants: Considering the test-taker in test development and research



In summary

This chapter discusses:

- test-taker characteristics and how they are monitored to inform the design of the test
- ensuring fairness for different applicant groups through the monitoring of test performance data.

Test-taker characteristics

We analysed 14 years of candidate data, from 2003 to 2016:



Gender

Each year, BMAT has been taken by a higher proportion of female candidates (56–59%). This reflects the distribution of gender amongst both those applying and successful applicants.



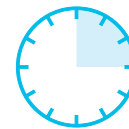
School type (UK centres only)

The proportion of candidates from independent (fee-paying) schools has decreased from 40% to 29% over the 14-year period.



Location

The proportion of BMAT candidates from non-UK centres has increased from 10% to 48% over the 14-year period.



Special needs access arrangements

Between 1.2% and 2.9% of candidates require extra time² for special needs. The proportion does not appear to be increasing over time.

² This type of provision typically accounted for the majority of access arrangements made.

“ This chapter will reassure institutions that they are not disadvantaging any group because BMAT predicts the on-course performance of all test-taker groups equitably. ”

Chapter 2 co-author, Amy Devine | Senior Research Manager, Cambridge Assessment Admissions Testing

Ensuring fairness for different applicant groups

a Performance of candidates with disabilities

Candidates with disabilities do not appear to be disadvantaged by the test. The marks of candidates who requested access arrangements were not uniformly lower than those of other candidates, nor was there an imbalance in their success rate compared to other candidates, in being offered a place of study.

b Performance of mature candidates

Mature applicants scored lower than non-mature applicants on:

- BMAT Section 1 in most test years (small effect sizes)
- BMAT Section 2 in all test years (medium effect sizes)
- BMAT Section 3 in four test years only (small effect sizes).

However, recent analyses comparing BMAT performance of *graduate* and *under-21* applicants has not replicated this difference in Section 2 scores, suggesting that time out of education may be the causal factor.



What does this mean?

BMAT scores predicted on-course examination results, on average, for all test-takers regardless of group. This provides crucial evidence that BMAT scores mean the same for different test-taker groups.

Case study: UCL Medical School

The first year that UCL Medical School applicants took BMAT, the admissions office decided to select using pre-existing processes and only look at BMAT scores *after* admissions decisions had been made.

The selected cohort of students had a reasonably broad spectrum of BMAT scores, and these scores were then validated against how well students performed at the end of the first and second year on the course.

The UCL Medical School admissions tutor said: *"We were reassured that students were performing well in relation to BMAT and also that we weren't discriminating against any particular groups by ethnicity, gender or school type."*

“ Even when slightly unexpected things come up, such as certain gender differences, we're able to see very convincing data that this is not in any way an unfairness or an inherent bias in a test; it reflects genuine differences between candidates. ”

Martyn Partridge | *National Heart and Lung Institute, Imperial College London*

Chapter 3

Cognitive validity – what skills are we assessing?



In summary

Chapter 3 in the BMAT research volume discusses:

- the skills included in BMAT and the rationale for selecting these
- the original test development
- examples of research studies that evaluate how well BMAT tasks assess these skills.

The skills included in BMAT

BMAT focuses on the skills and application of knowledge that are needed for biomedical study, rather than clinical practice. This has important implications when considering the cognitive validity of the test. Clearly there are other factors that contribute to on-course success. Therefore, admissions tests such as BMAT should be seen as one tool in the selection process, alongside other methods.

The original test development

The design of BMAT was influenced by the requirements of institutions and informed by two pre-existing tests³ that had been shown to make a positive contribution to student selection.

One significant decision when developing BMAT Section 2 (Scientific Knowledge and Applications) was that it should assess not only that a candidate has certain core scientific knowledge, but that they can apply it in a way that demonstrates an understanding of the scientific principles that underpin their knowledge. This is designed to distinguish a range of ability within a group of candidates with near-perfect grades in school examinations.

Various sources continue to agree on the core skills relevant for biomedical study: problem-solving (Quality Assurance Agency 2015), scientific reasoning (General Medical Council 2009) and writing abilities (Goodman and Edwards 2014, McCurry and Chiavaroli 2013). These studies confirm that the skills assessed by BMAT remain relevant to the contexts in which the test is used. Example test items and responses are analysed in the full BMAT research volume.

³ The Oxford Medical Admissions Test (OMAT) and the Medical and Veterinary Admissions Test (MVAT).

⁴ Verbal protocols can be an account of how a person is *currently* solving a problem, *did* solve a problem or *would* solve a problem.

“ We believe BMAT allows students to shine ... Problem solving is fundamental to medicine, which in many respects is a detective game. It's about looking for clues, asking questions about what's in front of you and thinking it through ... We also want people who can recognise there are many sides to an argument, listen to other people's point of view and write coherently. ”

Darren Beane | Head of Admissions, Recruitment and Widening Participation, Brighton and Sussex Medical School

Examples of research studies

Cognitive processes are difficult to investigate because they cannot be directly observed and the data collected in experimental settings is rarely available in formal testing contexts.

It is currently most common to collect data by conducting think-aloud studies, which are interpreted using verbal protocol analysis.⁴ Statistical techniques, known as factor analysis, are also used to examine performances on items and the relationships between them.

Key research studies: The verbal protocol analysis, which informed the development of BMAT

Green (1992) focused on questions similar to problem-solving items used in BMAT today. The analysis indicated that most items functioned well and did not merely require the application of routine procedural methods.

Thomson and Fisher (1992) focused on questions similar to the 'understanding argument' items in BMAT. Analysis confirmed that candidates use targeted cognitive processes when answering the majority of items. Complex wording and design of options can result in items that assess reading comprehension rather than critical thinking, but minor changes and edits reduce ambiguity and improve how well items elicit targeted skills.

Key research study: Using factor analysis to investigate cognitive validity

Factor analysis has shown that it is valid to interpret BMAT Section 1 (Aptitude and Skills) as measuring a unified construct of thinking skills, and valid to interpret BMAT Section 2 (Scientific Knowledge and Applications) as measuring a unified construct of scientific reasoning. There is also some evidence that an aggregate score for Sections 1 and 2 could be appropriate.



What does this mean?

Cognitive validity ties into all the other aspects of validity. Having established what the test needs to measure, other research questions can then be answered. It informs how questions are commissioned, reviewed and used in BMAT papers today, to ensure the best measurement performance of the questions.

An understanding of cognitive validity also helps institutions to explain the test to colleagues and applicants – 'This is what we're testing and this is how it fits with our course.'

“ We use BMAT because we think that the three test sections are well aligned with the way we deliver medical education at Lancaster and map really closely with our course. We believe that preparing for BMAT helps students to develop skills that will be useful at medical school. ”

Karen Grant | *Director of Admissions and Deputy Director of Medical Studies, Lancaster University*



In summary

This chapter discusses:

- task design considerations
- how task design and administration decisions are maintained in practice.

Task design considerations

Designing tasks for an admissions test such as BMAT requires careful consideration of various factors, because the response format, test timing and task content can all influence the skills assessed by a test.

Multiple-choice questions and tasks requiring constructed responses have specific advantages and disadvantages, so BMAT uses a combination of these task types across its sections.

Key research study: Are the number of questions and time constraints on BMAT appropriate?

One of the ways in which time constraints can be investigated is to review whether candidates demonstrate a high omission of test items, particularly towards the end of a test paper.

There was a slight increase in omit rates towards the ends of Sections 1 and 2, which are the multiple-choice sections. However, these omit rates were low, with less than 5% of candidates failing to respond at the ends of the test sections. Statistical analysis indicated that candidates were not guessing disproportionately at the end of BMAT sections.

The statistical evidence in this study indicates that time pressure in BMAT is not excessive for most candidates.

“ This chapter helps institutions to understand how tests are constructed and how test material is developed. It demonstrates the very rigorous process we go through to check the quality of the material and to ensure that the administration of the test is conducted fairly and securely. ”

Case study: 2014 revision of the BMAT Section 2 specification

Central aspects of context validity include: ensuring that test content is of a suitable level of difficulty and targeted to the intended test-takers, and that the assumed knowledge for a test is explicitly defined.

A key consideration in the development of BMAT was that the test would require candidates to do minimal new learning or preparation. It was therefore decided that the test specification should cover topics candidates would have been expected to study up to age 16.

The 2014 specification revision had a number of stages:

- 1 It began with a review of major science and mathematics syllabuses to identify common areas.
- 2 A round-table discussion with senior academics involved in student selection refined these areas.
- 3 Education professionals from around the world reviewed the specification, to confirm that content was targeted at the level expected of candidates in their final year of school study in their respective countries.
- 4 The specification was then used to develop suitable candidate preparation materials.

How task design and administration decisions are maintained

Various steps are used to ensure that, in practice, BMAT assesses the correct skills as intended. Our approach to authoring tasks uses detailed test specifications, review by subject matter experts and standardised processes to safeguard against threats to context validity.

It is critical that the test is administered securely and in a standardised way. Our approach to administration involves strict test regulations and centre approval processes to monitor these issues. The advantages and disadvantages of different administration methods are continuously reviewed.



What does this mean?

This chapter maps out very clearly the stages of developing BMAT test material. Our processes are designed to ensure that the test material and administration is:

- appropriate: it tests the right skills and qualities needed for success on undergraduate courses, at the right level of difficulty
- fair: it gives free and equal access to all applicants, regardless of background.

“ BMAT is proven, ethically aligned and, most importantly, it is fair. It gives our applicants the best possible opportunity to demonstrate their knowledge, excellence and potential, whatever their social backgrounds or educational chronology. We are confident that BMAT will help us to get the right undergraduates from amongst the brightest candidates. ”

Gail Nicholls | Director of Admissions, University of Leeds School of Medicine



In summary

Chapter 5 in the BMAT research volume investigates:

- the scoring validity of multiple-choice questions in BMAT Sections 1 and 2
- the scoring validity of written tasks in BMAT Section 3.

Scoring validity of multiple-choice questions in BMAT Sections 1 and 2

The Rasch model, an Item Response Theory (IRT) psychometric model, is used to calculate and report scores. Scores are capped upwards at 9.0 and downwards at 1.0, with an approximate mean of 5.0 for the reference group. This ensures that test results are comparable within a cohort.

Items of appropriate difficulty that discriminate between candidates

A wide range of statistical methods is used to ensure scoring validity. One example provided in the chapter is the analysis of item facilities, which finds that BMAT has a mean item facility around the desired .5, with very few items outside of acceptable ranges. This indicates that BMAT items are consistently set at appropriate difficulty levels for the cohort.

Reliability of a test

This is estimated from the consistency of scores within a test (internal consistency). The majority of internal consistency estimates for BMAT are in the lower range of coefficients considered acceptable. However, this is not unexpected given that Section 1 assesses three different thinking skills and Section 2 assesses four different subject disciplines.

Estimates of internal consistency could be improved by making BMAT sections more one-dimensional, but this would be detrimental to the quality of inferences that could be made on the test. Internal consistency could also be improved by increasing the number of test items. However, this would increase testing time or reduce the time available for each item, with knock-on effects for the context validity of BMAT.

Ensuring that responses come from the candidate

The use of statistics can also support detection of malpractice within test cohorts. For example, analysing response strings to identify cases with strong patterns of common wrong answers. Any cases which are flagged are referred to a malpractice panel for scrutiny.

Investigating item bias

Differential Item Functioning (DIF) analysis can be used to evaluate any items that are biased towards candidates of different groups, controlling for variance in overall test ability.

“ Scoring validity can be considered to include what might be referred to as ‘reliability’, but covers all decisions which impact the meaningfulness of the scores and how they are used. The scoring validity of BMAT multiple-choice questions is established in different ways to BMAT written tasks, and this chapter covers both. ”

Key research study: Investigating item bias by matching candidate-level information to item-level data

In the 2010–12 test years, there was no evidence of DIF by gender, or by UK school sector.

In the 2013–16 test years, an IRT approach was used and found that three items out of a total of 186 displayed moderate DIF. These items were balanced between those slightly favouring males/females, and independent/state school candidates. Our Assessment Managers review all flagged items to confirm there are no task features that give an advantage.

Scoring validity of written tasks in BMAT Section 3

Clearly defined marking criteria

Responses are marked by two markers against the marking criteria, which are included in full in the chapter.

Training and moderation of markers

Rasch model analyses are provided to Assessment Managers, who can provide further training and additional supervision. Where there are concerns about a marker's performance, the marker can be dismissed and their scripts re-marked.

Reliable and consistent marking

As responses are scored with subjective judgement, each marker needs to be internally consistent⁵ and there also needs to be consistency of marking between markers.

A multi-faceted Rasch model⁶ shows that the quality of marking is generally good. The most severe and lenient markers deviated from the mean by less than a quarter of a score point, indicating that the vast majority of markers were marking consistently.



What does this mean?

This chapter outlines the statistical methods and processes used to ensure that BMAT is measuring what it is intending to measure and providing an accurate, reliable measure of candidates' abilities.

It also presents a critical evaluation of how statistical and technical aspects of test validity need to be considered alongside other aspects of test validity. The aim is to achieve a good balance – ensuring that BMAT is covering the range of different skills it needs to cover *and* that it is assessing these skills fairly and accurately, within practical constraints.

⁵ Given a particular quality of performance, a marker needs to award the same mark whenever this quality appears.

⁶ A model that can, in addition to person ability and item difficulty, account for additional facets seen in scoring that involves judgement.

“ We believe that this chapter will reassure stakeholders that we take the integrity of results extremely seriously and we do all that we can to ensure results are accurate and reliable. ”

Chapter 5 co-author, Mark Elliott | Senior Validation Manager, Cambridge Assessment Admissions Testing

Chapter 6

Criterion validity – the relationship between BMAT scores and other measures of performance (the predictive validity of BMAT)



In summary

This chapter discusses:

- how predictive validity is measured
- evidence that BMAT scores predict future course performance (predictive validity)
- the relationship between BMAT test scores and predicted A Level grades.

How predictive validity is measured

Establishing predictive validity is critical to the overall validity of an admissions test. Predictive validity is often taken as evidence that a test 'works' for making selection decisions. Predictive validity must be evaluated in 'natural' settings, in which the educational trajectories of candidates are assessed and correlated with test scores.

In order to ensure rigour, our approach advocates reporting uncorrected coefficients alongside known information about the selection procedures used, which can be achieved by conducting situated studies⁷ in collaboration with admissions tutors.

In conducting predictive validity studies, careful consideration must be given to:

- what outcome measures will be used, and when these outcomes should be assessed
- reducing confounding factors, which may lead to spurious results
- methodological challenges, such as range restriction.

⁷ Studying predictive relationships within a specific context. The strength of the observable predictive relationship varies between institutions, courses and cohorts. This may be explained by differences in how BMAT is used at different institutions and aspects of the educational context that vary between and within courses, or even across years.

“ Criterion validity deals with one of the issues that institutions are most interested in – the extent to which admissions test scores predict future course performance. This chapter also outlines our approach to predictive validity research. It is essential that we have a rigorous, transparent, careful and considered approach, so that institutions can be confident in the results being reported. ”

Chapter 6 co-author, Molly Fyfe | Senior Research Manager, Cambridge Assessment Admissions Testing

Evidence that BMAT scores predict future course performance



Predicting on-course success

BMAT shows predictive validity across a range of courses and contexts. Correlations are typically positive, with both Sections 1 and 2 significantly predicting early course examination performance.

Predictive validity evidence for Section 3 (Writing Task) is mixed. In some cases, Section 3 scores are the strongest predictors of performance, while in other cases there is no predictive relationship. Unsurprisingly, findings indicate that Section 3 scores are more likely to correlate with course modules assessed using written components.



Difference in predictive validity

Most studies find that BMAT Section 2 has the strongest predictive validity. However, the strength of predictive relationships varies between institutions, courses and cohorts.

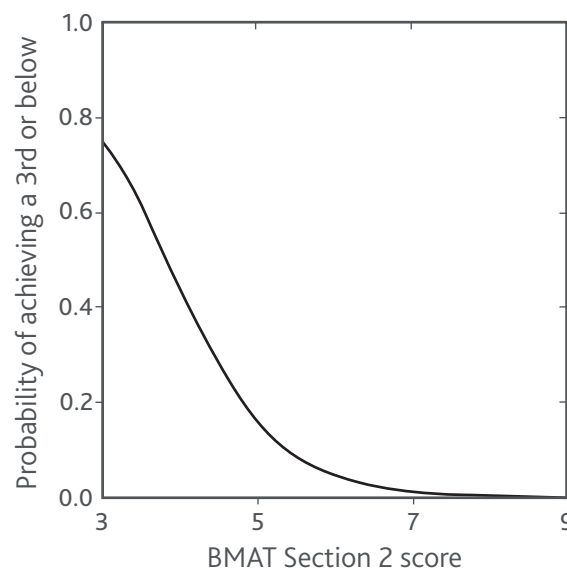
For example, for a *graduate entry* Medicine course, the BMAT Section 1 score and Section 3 'quality of content' score correlated positively with on-course performance. Scores from BMAT Section 2 did not show a significant correlation.



Predicting poor examination outcomes

In one cohort, there were sufficient numbers to permit this type of analysis. Candidates admitted with a low BMAT Section 2 score had a high probability of a poor examination outcome. Candidates with a BMAT Section 2 score of 5.0 or more had a very low probability of a poor outcome.⁸

Figure 2: Section 2 score and probability of achieving a 3rd or below



The relationship between BMAT Section 2 score and the probability of achieving a third or below in the first year of pre-clinical medical study. Examination data: Faculty of Biology, University of Cambridge.

⁸ A 3rd class result, failing the examinations or having left the course altogether.

Case study: The predictive validity of BMAT Sections 1 and 2 in the first four years of use at the University of Cambridge⁹

About the study: The study aimed to find out whether BMAT scores can significantly predict course performance, in candidates admitted with uniformly high A Level grades.

The selection process: A number of applicants with relatively low BMAT scores were offered a place each year, due to the compensatory nature of the selection process. However, correlations presented were likely to be underestimates of the true predictive validity of BMAT, due to the narrow score ranges overall, as we cannot know how rejected applicants with low BMAT scores would have gone on to perform.

Performance measures: The first and second year pre-clinical course examinations were the focus of this study, as BMAT focuses on academic readiness for demanding science-based study, not clinical skills.

Predicting course marks: Across all cohorts and course components, the correlation with course examination marks was consistently stronger for BMAT Section 2 than for BMAT Section 1. The majority of coefficients for Section 2 fell within the 'very beneficial' range (above 0.35) or the 'likely to be useful' range (above 0.21).

Correlation coefficients were slightly weaker for second year examinations. This was an expected outcome given that predictive relationships typically weaken with increasing time intervals.

Test cohort	Year 1 examination: total mark		Year 2 examination: total mark	
	Section 1	Section 2	Section 1	Section 2
MVAT ¹⁰ 2000	0.24***	0.44***	0.17**	0.26***
MVAT 2001	0.18**	0.26***	0.11	0.18**
MVAT 2002	0.19***	0.45***	0.11	0.23***
BMAT 2003	0.13*	0.26***	0.22***	0.25***

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Very beneficial range: above 0.35 | Likely to be useful range: above 0.21¹¹

Predicting 1st class attainment: Again, the relationship was stronger for Section 2 of the test. The lowest Section 2 scores were associated with a very low probability of achieving a 1st class outcome in both years of the course.

Conclusions: The findings from this early study are in line with many subsequent studies on the predictive validity of BMAT, which also show that Section 2 generally has stronger predictive validity.

The development of critical reasoning, problem-solving and writing skills is a common aim of medical education. However, there are challenges in establishing the predictive validity of thinking skills admissions tests, as these skills are rarely assessed discretely in higher education.

⁹ The University of Cambridge did not use BMAT Section 3 (Writing Task) scores in selection in these test years, instead considering candidates' responses as a qualitative piece of evidence.

¹⁰ BMAT was known as MVAT (Medical and Veterinary Admissions Test) prior to 2003, so the first three cohorts in this study sat MVAT rather than BMAT.

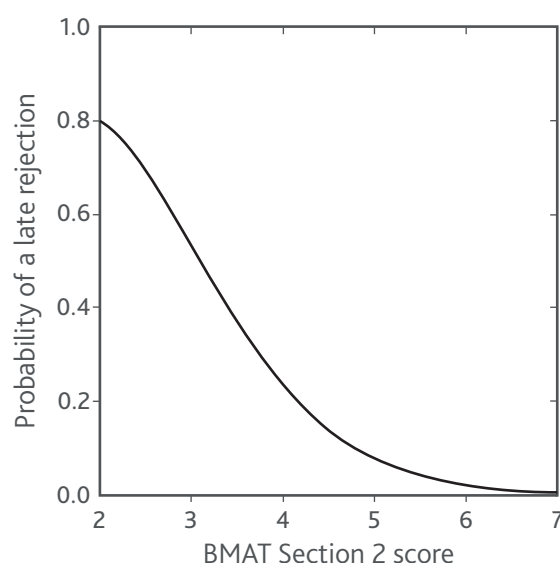
¹¹ These guidelines for interpreting uncorrected coefficients were published by the Department of Labor, Employment and Training Administration (US) (1999) *Testing and assessment: An employer's guide to good practices*, Washington, DC.

The relationship between BMAT test scores and predicted A Level grades

BMAT scores have been shown to correlate with A Level grades. All three BMAT sections show a strong positive relationship with the probability of achieving grades AAA.

Low BMAT Section 2 scores are also a particularly strong predictor of failing to achieve the grades needed for a conditional offer, which would result in candidates receiving a late rejection. This can be useful to institutions when making admissions decisions before A Level grades are known.

Figure 3: Section 2 score and probability of a late rejection



Logistic regression function showing the probability of a late rejection (failure to achieve A Level grades BBB) as a function of BMAT Section 2 scores (from Emery 2007c)



What does this mean?

Predictive validity research conducted in collaboration with institutions using BMAT shows that there is good evidence of BMAT's predictive validity across a range of courses and contexts. BMAT adds value to admissions processes by predicting on-course examination performance, and as an early indicator of A Level performance.

The Senior Admissions Tutor at Imperial College London commented on the use of BMAT for their courses: *"We realised that, with many, many hundreds of applications, we were basing our decisions to call for interview purely on what was considered relatively subjective. We started using BMAT to help us identify much more readily those who were likely to perform best on our course."*

Chapter 7

Consequential validity – the consequences of biomedical admissions testing on individuals, institutions and society



In summary

This chapter discusses:

- the wider social impact of the test, including how admissions tests impact on educational practices, individual applicants, the medical profession and society at large
- how admissions tests support efforts to widen participation and increase the diversity of medical, biomedical and dentistry students.

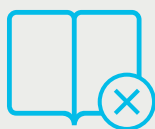
The wider social impact

An important area for research is to understand how preparing for tests like BMAT can impact upon student learning and test performance. This includes the possible impact of external help, through either commercial preparation courses or help received from school to prepare for BMAT.

BMAT is designed to complement candidates' existing school study and should not require candidates to invest large amounts of time or money. Our research into preparation practices has been used to provide extensive, free test preparation materials and guidance to all test candidates.

“ This chapter is about widening the way we think about admissions tests and considering the broader social impact of the test. In particular, it is reassuring to see research evidence showing that commercial preparation courses do not provide any advantage in terms of BMAT performance. ”

Key research study: BMAT preparation behaviours and the relationship with test scores¹²



Types of preparation that *do not* influence test scores

Candidates from UK independent schools were more likely to receive help from their school¹³ or from a commercial preparation course. However, no association was observed between better test performance and any of the following factors: school type, help received from school, or attending a course.



Feeling prepared

The majority of BMAT candidates surveyed felt 'very well' or 'quite well' prepared (66%). Candidates from UK state schools reported feeling better prepared than UK independent school and overseas candidates. The effects of *feelings* of preparedness on test performances are small, but significant.



Types of preparation that *do* influence test scores

Practising tests under timed conditions was associated with higher test scores on all sections of BMAT. The majority of candidates (83%) who looked at practice test papers reported practising under timed conditions. Attempting practice tests without time constraints and looking at papers without attempting to answer them were associated with poorer test performance.



Gender differences

More males (72%) felt 'very well' or 'quite well' prepared compared to females (62%). However, it is unclear whether differences in self-reported preparation are related to actual preparedness.



Time spent preparing

No association was found between the number of hours spent preparing and BMAT scores. The median amount of self-study time reported was approximately 30 hours.

¹² This is taken from a survey study conducted in 2015 by recruiting participants from the BMAT website. Candidate details were collected to enable matching to BMAT results data from the November 2015 session. Similar surveys were also administered in 2008 and 2007. Please note: There is no baseline measure of candidates' ability, so it is unknown whether candidates choosing different methods of BMAT preparation were of equal ability at the outset.

¹³ School help amounted to an average of 3 hours (similar for state and independent school candidates) and was most frequently in the form of advice on test content, rather than organised tuition/revision classes.

How admissions tests support efforts to widen participation

The social impact of BMAT extends to issues such as the diversity of the medical workforce. It is crucial that admissions tests such as BMAT do not pose an additional deterrent to potential applicants.

Key research study: How admissions tests are perceived by medical school applicants¹⁴

Admissions tests are not a barrier to applications: Admissions tests are seen as 'daunting' for similar reasons as interviews: they are 'one-off' chances. However, perceptions of admissions tests do not pose a barrier to applying to medical study. Candidates are committed to the idea of applying regardless.

Differences by test-taker characteristics: There are very few differences in how candidates from different social and educational backgrounds view the admissions process and admissions tests. More pronounced were gender differences. Females rated most criteria as more daunting than did males.



What does this mean?

Our findings indicate that applicants' perceptions of admissions tests do not pose a deterrent to applying to medical study. In addition, while candidates from different socio-economic backgrounds engage in different test preparation strategies, our research shows that this does not systematically give any group an advantage in terms of test performance.

¹⁴ This was a survey study conducted with students interested in applying to medical school. The survey was distributed in 2014 at medical school open days for prospective applicants and on the BMAT website. Participation was voluntary.

“ It is part of our responsibility as a test provider to consider the wider impact of the test. These are issues that we care about and want to take on board when developing the validity of BMAT. ”

Chapter 7 co-author, Molly Fyfe | Senior Research Manager, Cambridge Assessment Admissions Testing

Chapter 8

Conclusions and recommendations



In summary

This chapter discusses:

- the key points covered throughout the BMAT research volume and within each chapter
- areas for future research that can support investigations of validity going forward.

The key points



The socio-cognitive framework

The BMAT research volume represents a comprehensive application of the socio-cognitive framework to admissions testing and demonstrates why this approach can be useful.



Future trends

The number of institutions in the UK and internationally that use BMAT as part of their admissions process is steadily growing. We will continue to review the test to ensure its fitness for purpose in a global context.



Supporting the wider selection process

The research volume shows that the admissions test itself is only one part of a much greater context that responsible test providers must consider. Focusing on the test in isolation could result in validity claims that are not defensible.

“ The whole area of admissions is one that is taken incredibly seriously by institutions as it has such a far-reaching impact on the individual. Institutions try to be as transparent and evidence-based as they can when making decisions, and this is something we will continue to mirror in how we design the test. ”

Andy Chamberlain | *Head of Cambridge Assessment Admissions Testing*

Areas for future research



An increasingly international BMAT candidature

With BMAT serving an increasingly global higher education arena, it is crucial to evaluate the performance of an international candidature from more diverse educational backgrounds and how international group performance interacts with traditional group differences. Linking our data to other sources, such as Universities and Colleges Admissions Service (UCAS) data, could also be useful to further develop understanding of group differences.



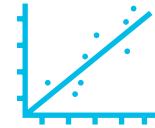
Scoring validity

BMAT's use in an increasing number of territories and contexts may require research into alternative scoring procedures that ensure comparability of scores across different sessions.



Cognitive validity

Further investigation of the balance between knowledge and novel problem solving could be beneficial for assessing scientific reasoning. This presents a possible avenue for further research that might be supported by technological advancements, which have been used to investigate scientific problem solving. These approaches can also investigate whether cognitive validity could be improved by lengthening a test section or including additional tasks, particularly for Section 3.



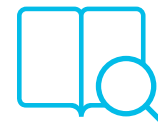
Criterion validity

New, large data initiatives present future research opportunities to investigate criterion validity. For example, we have been working with the General Medical Council (GMC) to provide data for the UK Medical Education Database (UKMED). However, this does not eliminate the need to collaborate closely with admissions tutors and understand issues specific to their contexts.



Context validity

BMAT is currently administered in paper-based format. This will continue to be reviewed. The introduction of computer-based testing is an area that will undoubtedly require further research: in particular, the equivalence of completing computer-based and paper-based tasks.



Consequential validity

Consequential validity is an area that has only been focused on relatively recently and there are many directions and areas of investigation. The interactions between consequential validity, other areas of validity and wider social theory represent areas to be explored with future research.

References

General Medical Council (2009) *Tomorrow's doctors: outcomes and standards for undergraduate medical education*, London: GMC. Available online: www.gmc-uk.org/Tomorrow_s_Doctors_1214.pdf_48905759.pdf

Goodman, N W and Edwards, M B (2014) *Medical Writing: A Prescription for Clarity*, Cambridge: Cambridge University Press.

Green, A (1992) *A Validation Study of Formal Reasoning Items*, Cambridge: UCLES internal report.

McCurry, D and Chiavaroli, N (2013) Reflections on the Role of a Writing Test for Medical School Admissions, *Academic Medicine* 88 (5), 568–571. DOI: 10.1097/ACM.0b013e31828b3dbb

Quality Assurance Agency (QAA) (2015) *Subject Benchmark Statement: Biomedical Sciences*, Gloucester: The Quality Assurance Agency for Higher Education. Available online: www.qaa.ac.uk/en/Publications/Documents/SBS-Biomedical-sciences-15.pdf

Thomson, A and Fisher, A (1992) *MENO: a Validation Study of Informal Reasoning Items*, University of East Anglia.

This is an executive summary of the full BMAT research volume.

The full research volume provides an in-depth summary of BMAT research to date, and promotes multidisciplinary collaboration between medical educators and other educational assessment experts, helping to improve selection for healthcare.

It was edited by:

Kevin Y F Cheung

Research and Thought Leadership Group,
Cambridge Assessment Admissions Testing

Sarah McElwee

Research and Thought Leadership Group,
Cambridge Assessment Admissions Testing

and

Joanne Emery

Consultant,
Cambridge Assessment Admissions Testing



Contact us to request your copy of the full BMAT research volume:
admissionstesting@cambridgeassessment.org.uk

We are Cambridge Assessment Admissions Testing, part of the University of Cambridge. Our research-based tests provide a fair measure of skills and aptitude to help you make informed decisions. As a trusted partner, we work closely with universities, governments and employers to enhance their selection processes.

Our assessments include:

- assessments in thinking skills
- admissions tests for medicine and healthcare
- behavioural styles assessment
- subject-specific admissions tests.

Visit our website to discover more: **admissionstesting.org/for-institutions**



Contact us

Cambridge Assessment Admissions Testing
The Triangle Building
Shaftesbury Road
Cambridge
CB2 8EA

 admissionstesting.org

 [admissions-testing](https://www.linkedin.com/company/admissions-testing)

 [/admissionstesting](https://www.youtube.com/channel/UCadmissionstesting)

 [@admissionstests](https://twitter.com/admissionstests)

Copyright © UCLES 2020 | CER/5036/V1b/JAN20



* 6154508277 *