

Applying the socio-cognitive framework to the BioMedical Admissions Test (BMAT)

Insights from language assessment

Also in this series:

The Impact of High-stakes Examinations on Classroom Teaching: A case study using insights from testing and innovation theory

Dianne Wall

Impact Theory and Practice: Studies of the IELTS test and *Progetto Lingue 2000*

Roger Hawkey

IELTS Washback in Context: Preparation for academic writing in higher education

Anthony Green

Examining Writing: Research and practice in assessing second language writing

Stuart D. Shaw and Cyril J. Weir

Multilingualism and Assessment: Achieving transparency, assuring quality, sustaining diversity – Proceedings of the ALTE Berlin Conference, May 2005

Edited by Lynda Taylor and Cyril J. Weir

Examining FCE and CAE: Key issues and recurring themes in developing the First Certificate in English and Certificate in Advanced English exams

Roger Hawkey

Language Testing Matters: Investigating the wider social and educational impact of assessment – Proceedings of the ALTE Cambridge Conference, April 2008

Edited by Lynda Taylor and Cyril J. Weir

Components of L2 Reading: Linguistic and processing factors in the reading test performances of Japanese EFL learners

Toshihiko Shiotsu

Aligning Tests with the CEFR: Reflections on using the Council of Europe's draft Manual

Edited by Waldemar Martyniuk

Examining Reading: Research and practice in assessing second language reading

Hanan Khalifa and Cyril J. Weir

Examining Speaking: Research and practice in assessing second language speaking

Edited by Lynda Taylor

IELTS Collected Papers 2: Research in reading and listening assessment

Edited by Lynda Taylor and Cyril J. Weir

Examining Listening: Research and practice in assessing second language listening

Edited by Ardeshir Geranpayeh and Lynda Taylor

Exploring Language Frameworks: Proceedings of the ALTE Kraków Conference, July 2011

Edited by Evelina D. Galaczi and Cyril J. Weir

Measured Constructs: A history of Cambridge English language examinations 1913–2012

Cyril J. Weir, Ivana Vidaković, Evelina D. Galaczi

Cambridge English Exams – The First Hundred Years: A history of English language assessment from the University of Cambridge 1913–2013

Roger Hawkey and Michael Milanovic

Testing Reading Through Summary: Investigating summary completion tasks for assessing reading comprehension ability

Lynda Taylor

Multilingual Frameworks: The construction and use of multilingual proficiency frameworks

Neil Jones

Validating Second Language Reading Examinations: Establishing the validity of the GEPT through alignment with the Common European Framework of Reference

Rachel Yi-fen Wu

Assessing Language Teachers' Professional Skills and Knowledge

Edited by Rosemary Wilson and Monica Poulter

Second Language Assessment and Mixed Methods Research

Edited by Aleidine J Moeller, John W Creswell and Nick Saville

Language Assessment for Multilingualism: Proceedings of the ALTE Paris Conference, April 2014

Edited by Coreen Docherty and Fiona Barker

Advancing the Field of Language Assessment: Papers from TIRF doctoral dissertation grantees

Edited by MaryAnn Christison and Nick Saville

Applying the socio-cognitive framework to the BioMedical Admissions Test (BMAT)

Insights from language assessment

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



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108439312

© Cambridge University Press 2017

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2017

20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Printed in

A catalogue record for this publication is available from the British Library

ISBN 978-1-108-43931-2

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables, and other factual information given in this work is correct at the time of first printing but Cambridge University Press does not guarantee the accuracy of such information thereafter.

Contents

Acknowledgements	vi
Series Editors' note	ix
Foreword	xi
Preface	xiii
Notes on contributors	xvi
List of abbreviations	xix
1 The Cambridge Approach to admissions testing <i>Nick Saville</i>	1
2 The biomedical school applicant: Considering the test taker in test development and research <i>Amy Devine, Lynda Taylor and Brenda Cross</i>	17
3 What skills are we assessing? Cognitive validity in BMAT <i>Kevin Y F Cheung and Sarah McElwee</i>	35
4 Building fairness and appropriacy into testing contexts: Tasks and administrations <i>Mark Shannon, Paul Crump and Juliet Wilson</i>	81
5 Making scores meaningful: Evaluation and maintenance of scoring validity in BMAT <i>Mark Elliott and Tom Gallacher</i>	114
6 The relationship between test scores and other measures of performance <i>Molly Fyfe, Amy Devine and Joanne Emery</i>	143
7 The consequences of biomedical admissions testing on individuals, institutions and society <i>Sarah McElwee, Molly Fyfe and Karen Grant</i>	181
8 Conclusions and recommendations <i>Kevin Y F Cheung</i>	216
References	233
Author index	249
Subject index	254

1 The Cambridge Approach to admissions testing

Nick Saville

Cambridge English Language Assessment

1.1 Purpose of the volume

Admission to study medicine is highly competitive and medical degree courses are among the most oversubscribed in the world. Many applicants that submit themselves to the multi-stage selection processes used for admissions have the highest school-leaving qualifications achievable. In the UK, these pressures led to development of the BioMedical Admissions Test (BMAT), which at the time of writing has been used for over a decade in the selection of students to medicine, veterinary medicine, dentistry and biomedical sciences courses. Originally commissioned for use by a small number of universities, BMAT has been increasingly employed by institutions around the world, partly due to growing trends for medical courses to use English as the medium of instruction.

In the UK, a report commissioned by the regulatory body for doctors, the General Medical Council (GMC), recommended that medical schools include standardised testing alongside other selection methods (Cleland, Dowell, McLachlan, Nicholson and Patterson 2012). Since the publication of this report, use of BMAT and other admissions tests has become more widespread in the UK, and there has also been more research focused on selection processes. This volume contributes to the body of research on selection for medical study, by presenting work on BMAT that has been conducted in over a decade of research and validation. Historically, this research and validation work has informed test construction and been used to respond to specific stakeholder queries. More recently, the results of research have been made available for medical schools and departments considering use of BMAT in their admissions process, many of whom have since adopted BMAT. Therefore, much of the research in the present volume will be familiar to members of the BMAT stakeholder group; however, this is the first time that this work has appeared together in one collection, and been made available for the wider medical education community.

The process of validating a test to show that it is fit for purpose involves a process of building an argument. As Kane (2013:1) has noted, ‘public claims require public justification’. The intention of the collection of studies and procedures in the chapters that follow are offered in the spirit in which

Kane suggests – to make public the validation arguments that support use of BMAT. The studies presented are a mixture of published research, conference papers and internal Cambridge Assessment reports. They are not an exhaustive list but represent key examples and summaries of research work (spanning the earliest years of BMAT to the present day) that covers a broad range of validity evidence. The volume has been compiled for readers involved in selecting students for medicine or biomedical courses, from policy-makers to anyone wishing to better understand the evidence base for the test.

In addition to collecting together research on BMAT, this volume articulates a multi-faceted conceptualisation of test validity – the socio-cognitive approach – and applies it to the admissions testing context of BMAT. This theoretical framework of assessment is used for in-depth analysis of validity in high-stakes language testing, and has been adopted by Cambridge Assessment Admissions Testing because of the comprehensive and structured treatment it provides of multiple aspects in the testing process. Some of the work presented in the following chapters predates Cambridge Assessment Admissions Testing's use of the socio-cognitive model in validation; however, all of the studies included map onto areas identified in the socio-cognitive approach, demonstrating the framework's suitability for the admissions testing context, and potentially for other contexts across educational assessment.

1.2 The Cambridge Approach

Cambridge Assessment Admissions Testing is part of the Cambridge Assessment Group, Europe's largest educational research and assessment agency, and a department of the University of Cambridge. We recognise that this brings with it a high level of responsibility and a requirement to ensure that our assessment systems not only deliver fair and dependable results for test takers, but also have positive effects and consequences for society at large. We work with national governments and other organisations to develop learning and testing solutions that meet their precise needs and, where these needs cannot be met using our existing services, we develop tailored solutions.

In order to ensure that all forms of assessment are of the highest quality and are appropriate to their context and intended uses, Cambridge Assessment has developed a set of common standards known as the *Cambridge Approach* (Cambridge Assessment 2009). This approach sets out an overarching framework for assessment, reflecting the University's broad goal 'to contribute to society through the pursuit of education, learning and research at the highest levels of excellence'.

While Cambridge Assessment Admissions Testing adheres to the Cambridge Approach in general, it also looks at best practice in assessment taking place within the three examination boards that make up the Cambridge Assessment Group, to ensure the quality of the assessments it delivers. In particular,

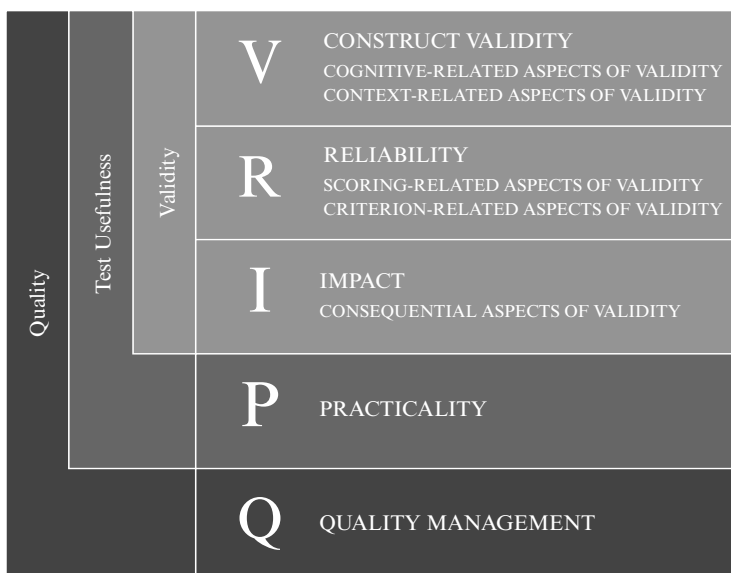
Cambridge Assessment Admissions Testing uses principles developed by Cambridge English Language Assessment (which provides over 5 million tests a year for speakers of English as a second language) to guide test development, production and research. The Cambridge English *Principles of Good Practice* document provides an accessible and concise overview of the key concepts, together with examples of how the principles are put into practice (Cambridge English 2016). The four guiding principles are: fitness for purpose, communication and collaboration, quality and accountability, validity and validation.

Below we focus mainly on the principles of fitness for purpose and validity and validation to explain why these are relevant to this volume. A brief overview of communication and collaboration is presented and we also summarise how quality and accountability impacts on the work of Cambridge Assessment Admissions Testing.

Fitness for purpose

Central to the validity of BMAT as an admissions test for biomedical study is establishing that it is fit for purpose. Fitness for purpose is a multi-faceted idea that incorporates not only the more traditional aspects of validity and reliability, but also the practicality of the test, its impact on stakeholders, and the quality management system that underpins it to ensure that the standard of the assessment is consistent over time. We discuss these issues below with

Figure 1.1 Elements of fitness for purpose in test validation



reference to practicality and impact. Figure 1.1 illustrates our approach to achieving fitness for purpose, and shows how developments in testing theory and quality management have been incorporated into a coherent framework that guides our assessment practices.

Impact by design

High-stakes assessment has important effects and consequences within an educational system and on society more widely. These effects are referred to as *impact*. Test takers in particular are affected because the results of tests are used to make important decisions about them which can affect their lives. In developing and administering our tests we adopt the principle of *impact by design*: we strive to achieve positive impact in the contexts in which our assessments are used and we undertake to investigate this through our validation processes. We seek to design and develop test features that promote positive effects on learning (Saville 2012). The principle of impact by design is useful for considering the constructs assessed by BMAT, and how preparing for the test might benefit the test taker.

The individual qualities of validity, reliability, impact and practicality cannot be evaluated independently; rather their relative importance must be determined in order to maximise the overall ‘fitness for purpose’ of the exam (see Saville 2003).

Practicality considerations

Practicality can be defined as the extent to which an examination is practicable in terms of the resources necessary to produce and administer it to the highest standard in its intended context and use. It affects many different aspects of an examination and we regularly consult relevant stakeholders during test development and revision processes on the practical aspects of using an admissions test. Test length is one such example; while longer tests can increase reliability because they capture more measurement data, they may be impractical to administer. In addition, an overly long exam could induce fatigue in candidates, which in turn could introduce error into the measurements. If some candidates are more susceptible to fatigue than others, this can result in score differences unrelated to the attribute being assessed. This example illustrates how practicality can impact the validity of the inferences based on a person’s score responses; therefore, some practical aspects overlap with contemporary conceptualisations of validity (e.g. Messick 1995).

Other practicality considerations are concerned with the locations where an examination is administered and the processes for getting materials to and from these places securely (either digitally or physically). BMAT uses an international network of test centres maintained by the three exam boards across Cambridge Assessment. We work with centres to make sure that the

systems we use are up to date and flexible enough to allow effective and efficient administration. Finally, in line with our educational mission, we wish to maintain access for the widest proportion of candidates possible, which means we strive to hold costs at a reasonable level, whilst ensuring that the test is available internationally.

Communication and collaboration

Our approach recognises the importance of communicating and collaborating with the organisations and individuals that use assessments to make decisions. This enables a test developer to have an understanding of the contextual issues that impact on a programme of assessment. Cambridge Assessment Admissions Testing works to ensure that it is receptive to the needs, opinions and knowledge of key stakeholders. In the context of BMAT, these are primarily the biomedical and dental departments that use the test to select students from their applicant pools.

Information and liaison

We constantly liaise with organisations that use our admissions tests, in order to provide support and information. For BMAT, a liaison group meets twice annually to discuss test sessions and general issues related to healthcare admissions. Although hosted by Cambridge Assessment Admissions Testing, meetings are chaired by one of the stakeholders; this changes on a rotating basis. The meeting chair occasionally uses the meeting to identify a specific agenda point, which can prompt a day of discussions and talks that focus on an area of interest. All organisations that use BMAT are invited to the meetings, and attendees typically include the admissions tutors for individual departments or courses. The forum provides an opportunity for BMAT users to discuss operational issues and topics of academic interest. Members of Cambridge Assessment Admissions Testing's research team attend these meetings to present updates and discuss possible areas for future research.

Cambridge Assessment Admissions Testing researchers also attend academic conferences that focus on a wide range of areas, such as educational assessment, higher education policy and medical education. Cambridge Assessment engages in policy discussions regarding the use of admissions tests in various international contexts, and also hosts events to lead discussions around healthcare selection and bring together institutions from the UK and overseas. Most recently, the Optimising Admissions conference held at the Royal College of Physicians in April 2017 featured presentations from the General Medical Council (GMC), medical schools from the Netherlands and healthcare educators from around the UK.

Research collaborations

Cambridge Assessment Admissions Testing works with many organisations throughout the world, including universities, government departments, major commercial organisations and many others. As part of Europe's largest assessment agency, which is also a department of the University of Cambridge, we have a research capacity that enables delivery of joint international projects. These researchers also develop and manage relationships with key stakeholders, often by collaborating with them.

Recently, Cambridge Assessment's collaboration with the GMC has been influential in advancing the UK Medical Education Database (UKMED) project, which will include BMAT data for research purposes. This initiative supports medical education research by linking data together from multiple sources and making it available securely and anonymously. Research into admissions testing is also supported using funded research programmes. A recent round awarded funding for three BMAT-related projects, to researchers from Lee Kong Chian School of Medicine (LKC), University of Leiden Medical School and Imperial College School of Medicine.

Quality and accountability

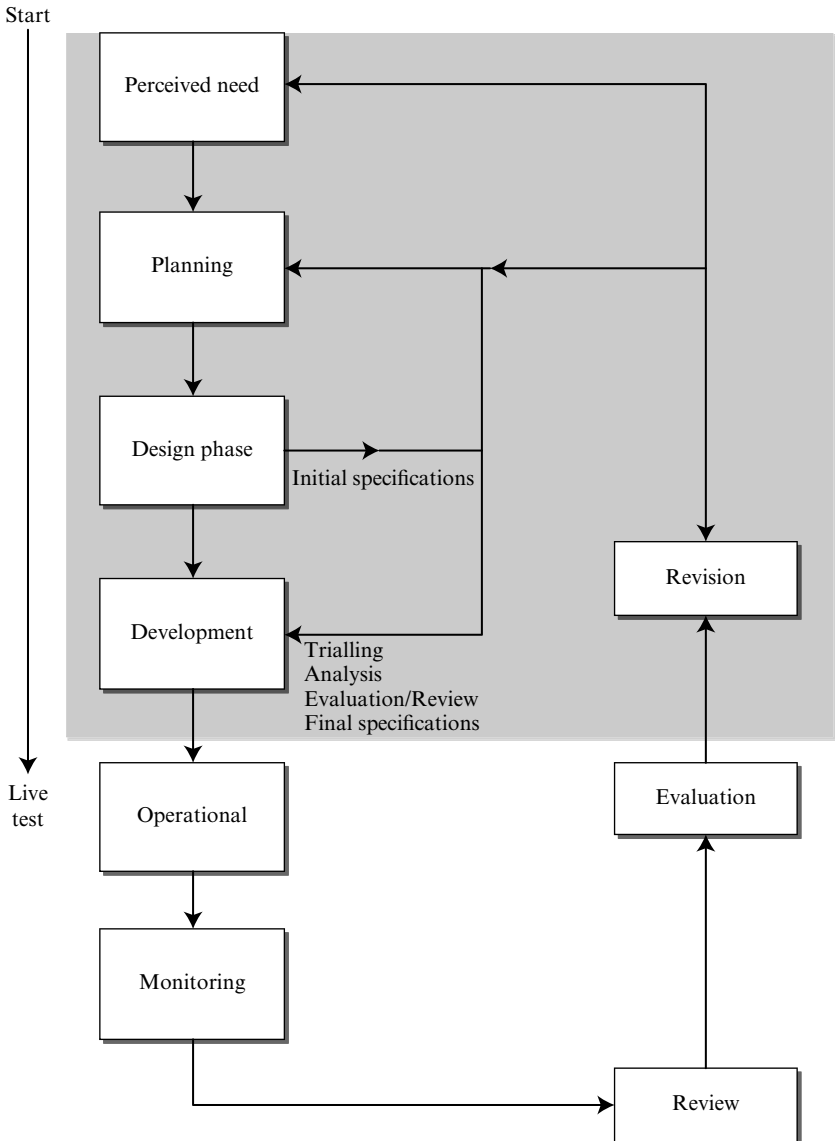
Quality control and assurance supports an assessment organisation to achieve fitness for purpose consistently. Cambridge Assessment Admissions Testing, like Cambridge English Language Assessment, adopts a process approach to ensuring quality. Processes are defined and agreed so that quality control and quality assurance procedures can be carried out. Many of the processes related to BMAT question paper production are described in Chapter 4 of this volume. These procedures are constantly reviewed as part of the test development and validation cycle.

The test development and validation cycle

To ensure fitness for purpose we employ an explicit model for the *test development and validation process* which incorporates continual improvement cycles. This is applied to all our admissions tests, including BMAT.

As shown in Figure 1.2, the process begins with a perceived need for a new or revised test, and is then broken down into three phases: planning, design and development. The first task in the planning phase is to define the intended *context and use* of the prospective test by identifying stakeholders and their needs, and considering both theoretical and practical issues in meeting these needs. The original BMAT stakeholders were the medical and veterinary schools at University of Cambridge and University of Oxford. As BMAT was designed to meet the needs of these stakeholders at their request, they were consulted throughout the planning, design and development of the test.

Figure 1.2 A model of the ongoing test development validation cycle



The output of the development stage is a set of test specifications – a document or documents defining the test, its validity argument and its operational requirements. The specifications act as a ‘blueprint’ for the operational production of tests. The most up-to-date test specification for BMAT is available

on the Cambridge Assessment Admissions Testing website. Chapter 3 of this volume describes the original planning phase of BMAT and Chapter 4 outlines how the test specification informs the way that BMAT papers are constructed.

The question paper production process begins with commissioning of draft materials and ends in the printing of the final question papers. The process for each test component is managed by an assessment manager who works with external experts, including item writers. Each component of a test has *item writer guidelines* specifying the requirements of each task type. Questions that do not meet these criteria are rejected or rewritten. Those that are accepted are taken through a rigorous editing process by experienced consultants.

In the operational phase, the process of examination administration ensures that all necessary arrangements are in place so that candidates can take the exam in the most efficient way. Key tasks include quality assurance of the test centres, delivery of exam materials and administrative documentation to centres. Details of the quality assurance of BMAT administration can be found in Chapter 4 of this volume.

The main stages of post-exam processing are marking, grading and the reporting of results. Data on test takers, test materials, and marking and grading procedures must be captured, stored and analysed for all exam sessions. Chapter 5 outlines the scoring, marking and reporting procedures for BMAT.

All Cambridge Assessment Admissions Testing assessments are reviewed and evaluated regularly. Review takes place during the routine monitoring of operational processes, and typically, improvements are implemented in an ongoing manner. All facets of the test's fitness for purpose are evaluated, including its practicality. Although the primary purpose of BMAT has changed little, some contextual factors have changed substantially. For example, the first time BMAT was administered, demand for a biomedical admissions test was limited to a small number of institutions in the UK, whereas now, BMAT is used by 17 institutions in seven countries to support their selection procedures. These developments impact on operational and theoretical considerations that are monitored and adjusted for. An example of this is presented in Chapter 4 as a case study: a revision of BMAT Section 2, which ensured BMAT's fitness for purpose in the face of science curriculum changes and an increasingly international candidature.

If necessary, a major revision project is initiated which, in essence, loops back to the planning phase of the cycle. Cambridge Assessment Admissions Testing involves BMAT stakeholders in the routine monitoring described above, to include their evaluations regarding the need for major revision. The current format and structure of BMAT is outlined below, along with some context regarding the test's purpose and use.

The BioMedical Admissions Test

BMAT is a pen-and-paper test – available at schools and colleges worldwide – that assesses an applicant’s readiness for the demanding, science-based study required in medicine and other biomedical courses. It is intended to supplement, rather than replace, the information provided by prior examination results, standardised application forms (such as UCAS), and interviews. Institutions differ in their use of BMAT scores, with some using scores as a hurdle to the interview stage and others not. It is often used in a compensatory manner with other selection criteria, meaning that low-scoring applicants are not always rejected and high-scoring applicants are not always offered a place.

BMAT has three elements: a domain-general aptitude and skills section, a section based on scientific knowledge and applications, and a short communicative writing task. A summary of the sections is provided in Table 1.1.

Table 1.1 Summary of BMAT sections

Section 1 Aptitude and Skills	This element tests generic skills often utilised in undergraduate study: problem solving, understanding argument, and data analysis and inference skills. There are 35 items in 60 minutes.
Section 2 Scientific Knowledge and Applications	This element tests whether candidates have the core knowledge and the capacity to apply it, which is a pre-requisite for high-level study in biomedical sciences. Questions are restricted to material typically included in non-specialist school Science and Mathematics courses but require a level of understanding appropriate for such an able target group. There are 27 items in 30 minutes.
Section 3 Writing Task	This element tests the ability to select, develop and organise ideas and to communicate them in writing, concisely and effectively. A selection of questions on topics of general, medical, veterinary or scientific interest are available, one of which must be chosen. The response is limited to one A4 page in 30 minutes.

BMAT Sections 1 and 2 are in multiple-choice format (objectively marked) and all items are worth one mark. Scores are reported on a calibrated, 9-point scale to one decimal place. The Writing Task is marked by a team of expert markers at Cambridge Assessment. An image of the response is also supplied to each institution to which the candidate has applied. This provides the institution with an example of each applicant’s writing skill that has been completed under exam conditions, unlike other samples of writing that are commonly made available as part of the application process. Admissions tutors are therefore confident that applicants authored the Section 3 essay, which can be further reviewed qualitatively if needed. Past

Applying the socio-cognitive framework to BMAT

BMAT papers and sample materials are available on the BMAT website: www.admissionstesting.org/for-test-takers/bmat

Institutions want to be confident that they have selected the right applicants and deselected those who are least likely to succeed. The utility of BMAT is demonstrated by studies to date, which have shown that BMAT scores relate to applicants' future performance on their course of study. Despite the difficulties involved in making predictions about future behaviour, high BMAT scores are associated with high course outcomes, and low scores (especially for the Scientific Knowledge and Applications section) are associated with poor course outcomes for those admitted, despite their high A Level grades.

A further benefit to institutions is that they are able to use BMAT to deselect applicants for interview if there is sufficient evidence, from usage of the test, that applicants with low scores have very little chance of being offered a place of study (when their scores were unseen by those doing the selecting). Where the selection process begins before A Level results are available (as it does for most) we have also shown that BMAT results can predict those likely to achieve weaker A Level outcomes (importantly, these are applicants failing to meet their conditional offer grades despite high predictions).

Of course, no *single* assessment will be ideal in the selection of those for a career in medicine, dentistry or veterinary medicine and the non-academic attributes thought to be desirable in applicants, such as personal qualities, will need to be assessed in other ways. For this reason a multifaceted approach to medical selection is widely accepted and recommended (Cleland et al 2012). When used alongside other selection criteria, BMAT can usefully aid admissions tutors in choosing applicants who can cope with the academic demands of their future course of study and thrive in the intellectually rigorous environment required. This volume presents work conducted to ensure that BMAT remains fit for purpose in a complex and important environment, where an understanding of the test's validity is crucial.

Validity and validation

Validity has generally been defined as the extent to which an assessment can be shown to produce scores and/or outcomes which are an accurate reflection of the test taker's true level of ability. It is concerned with the appropriateness and meaningfulness of inferences made when using the test results within a particular social or educational context. Validation is the process of accumulating evidence to support these interpretations. We endorse this view and in doing so draw on internationally recognised standards such as the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association and National Council on Measurement in Education 2014), which is hereafter referred to as the *Standards* (2014).

Cambridge Assessment Admissions Testing's view of validation draws on a *socio-cognitive approach* to defining the abilities which are to be tested by an assessment system. This theoretical framework of learning and assessment is used for in-depth analysis of the validity of our tests, including BMAT. The model was developed by Cambridge English Language Assessment researchers in collaboration with Weir (2005), and draws on the work of Messick (1989), who stressed the interacting nature of different *types of validity* evidence (and also Bachman 1990, Bachman and Palmer 1996).

The approach is described as socio-cognitive in that carrying out tasks in real life is a social phenomenon and the underlying abilities which enable these actions are mental constructs (the cognitive dimension). A valid test seeks to engage the mental capacities in an authentic way so that appropriate inferences can be drawn from the test score.

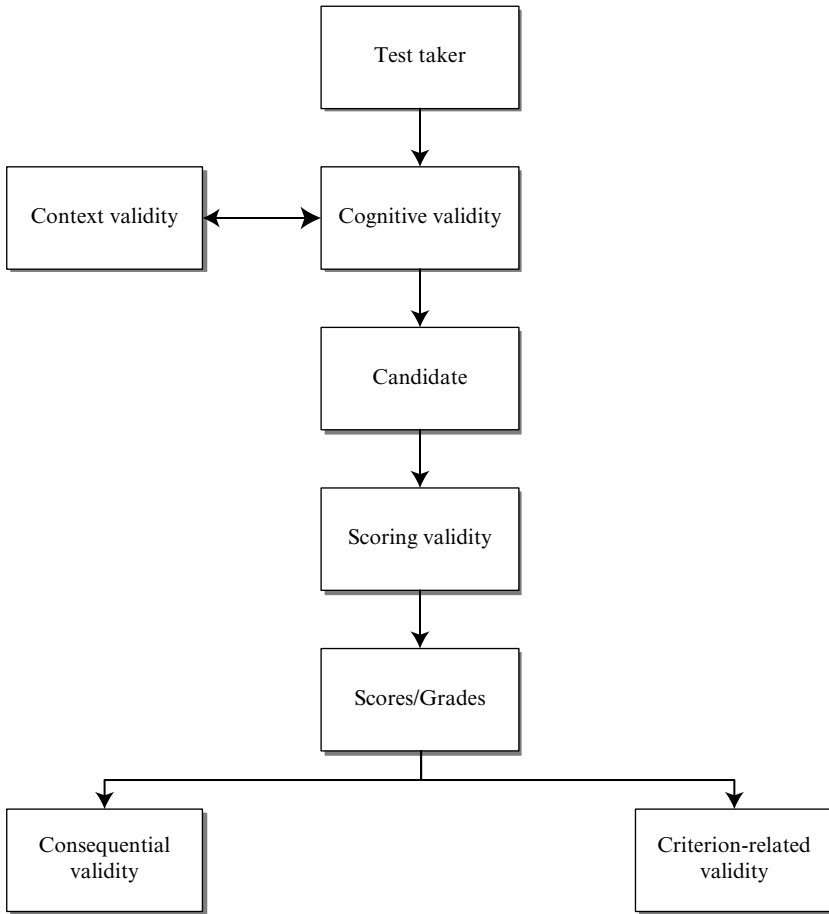
In the case of potential for biomedical study, it is important to make explicit what is meant by this notion and to account for the ways in which test results can be used to make dependable decisions about the test takers' ability. In other words, we need to have confidence that high scores reliably reflect more potential than lower scores. This point highlights a fundamental issue of validity – the nature of the *constructs* which are at the heart of our tests and how we account for them. The construct of a test is the theory that the test is based on. For BMAT, this is the theory of the cognitive skills, the core scientific reasoning and the written communication abilities that will enable a student to cope with the demands of a rigorous biomedical degree. Within the socio-cognitive approach, we account for the test construct by considering six 'aspects' of validity for which supporting evidence must be provided.

Figure 1.3 illustrates the principal direction of hypothesised relationships between elements of the socio-cognitive framework. It shows that, 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 results.

This view treats the aspects of validity outlined above as component parts of overall validity. This unitary conceptualisation of validity requires a comprehensive *validity argument* to be presented. A validity argument is a well-reasoned rationale in which the examination provider presents an overall evaluation of the intended interpretations and uses of the test which is being validated. This is consistent with the definition of validation as: 'the ongoing process of demonstrating that a particular interpretation of test scores is justified' (Bachman and Palmer 1996:22).

This approach to validation underpins contemporary work in language testing contexts, but validation of admissions tests have tended to focus on reliability, which is a narrow aspect of scoring validity, and predictive

Figure 1.3 The socio-cognitive validation framework (adapted from Weir 2005)



validity, a form of criterion-related validity (Soares 2012). There is growing consensus that the predictive and scoring validity of admissions tests should be supplemented with other forms of validity (Atkinson and Geiser 2009, Linn 2009). By applying the socio-cognitive model to BMAT, this volume provides an example of how various aspects of validity can be considered in relation to an admissions test. In building and presenting a validity argument we seek to:

- set out our claims relating to the usefulness of the test for its intended purpose
- explain why each claim is appropriate by giving reasons and justifications

- provide adequate evidence to support the claims and the reasoning behind them.

This evidence is built over time, beginning at the design and development stages, and continues to be accumulated for as long as the test remains operational. This volume presents the validity argument for BMAT in its current form, and does not preclude evidence being added to the overall case for BMAT in the future. Indeed, the Cambridge Assessment approach subscribes to a continuous model of test validation that continues throughout the lifetime of any test.

1.3 Structure of the volume

In the seven chapters that follow, evidence for the validity of BMAT is presented. Each of these chapters focuses on an aspect of the socio-cognitive model (see Figure 1.3) and then the final chapter draws this work together. The Appendix shows the questions that need to be considered by the test provider in gathering evidence for each of these aspects. We use these questions to develop standard quality procedures and operational analyses for a test as well as to design and conduct targeted research studies that generate empirical evidence. In this collection, therefore, we outline the quality procedures and routine analyses put in place for BMAT that relate to each of these aspects of validity, as well as presenting key evidence from research studies.

In designing a test for a particular context and purpose, we profile the intended test takers in terms of their characteristics: demographic features (such as gender), existing knowledge and prior learning experiences. The BMAT candidature has changed over the years that the test has been used and we continue to collect information in an ongoing manner during operational phases to make sure that the test is still fit for purpose. Additionally, various concerns about the composition of the medical student population have been raised by the Medical Schools Council (2014). Research studies have been conducted to ensure that BMAT is not contributing to diversity issues and these are presented as part of Chapter 2.

Chapter 3 discusses cognitive validity in the context of BMAT. Cognitive-related validity is concerned with the extent to which the cognitive processes employed by candidates are similar to those that will be needed in real-world contexts beyond the test. For BMAT, medical school represents the context of interest. This chapter focuses on the constructs assessed by BMAT sections and the work done to ensure that the skills assessed are relevant to successful study at medical school.

Chapter 4 is concerned with context validity. Discussion focuses on the conditions under which the test is performed and includes features of the tasks as well as the administration conditions. Some of the issues regarding task features examined in this chapter have a symbiotic relationship to those

covered on cognitive validity, as an item's characteristics can have unintended effects on the cognitive processes used to answer the question, potentially introducing construct-irrelevant variance to test scores that should be avoided. Threats to a test's validity can also arise if administration conditions are not considered and standardised, so the operational processes that safeguard the security and integrity of BMAT are included in this chapter.

It is essential to ensure that tests are scored accurately (with no processing errors) and it is important to estimate the reliability of the results – the extent to which they are stable, consistent and free from errors of measurement. In Chapter 5, the scoring procedures for BMAT are discussed, alongside some of the psychometric procedures used to monitor and evaluate test sessions. It is also essential that tests are fair and not biased in favour of one group of test takers over another. This aspect of scoring validity links closely with the issues covered in Chapter 2; knowledge of the test taker sample informs post-test analyses of group differences and this chapter presents examples of this work.

In establishing criterion-related aspects of validity the aim is to demonstrate that test scores are systematically related to another indicator of what is being measured, or a measure of some related construct, such as another established test or a predicted outcome. Much of the research on admissions testing, particularly in the medical context, is dominated by predictive validity research focused on academic achievement. In Chapter 6, published work on how BMAT scores predict course outcomes is referred to, along with some analyses relating to other outcome variables, such as A Levels. In addition, this chapter discusses some common challenges experienced when conducting predictive validity studies. In particular, work looking at the effects of range restriction in selection situations is shared and common approaches to dealing with this phenomenon are critically discussed.

Consequential validity (or 'impact') is concerned with the effects of using a test on stakeholders (including test takers themselves) and on wider society. These consequences may be positive or negative, intended or unintended, and are particularly relevant to high-stakes tests like BMAT. Stakeholder perceptions of a test and washback effects on learning are aspects of consequential validity. Chapter 7 deals with these issues and the responsibilities of the test developer to consider BMAT's impact on society more generally.

Finally, Chapter 8 outlines the key issues raised by authors throughout the volume, in order to present some conclusions related to the use and study of admissions tests such as BMAT.

1.4 Chapter summary

This introduction chapter has highlighted the need for test developers to evaluate various aspects of validity when constructing a high-stakes admission

test such as BMAT. The importance of considering various aspects of the test's candidature and administration has also been emphasised. In order to meet these requirements, we have adopted a comprehensive validation framework in the form of the socio-cognitive approach (O'Sullivan and Weir 2011, Weir 2005), and applied it to BMAT. This framework is elaborated in the following chapters to present a multi-faceted overview of BMAT, and the validity arguments that underpin the assessment. Furthermore, this collection can promote similarly comprehensive evaluations in the admissions testing literature, and potentially in other areas of educational assessment. Therefore, this volume is suitable for anybody with an interest in educational assessment.

It is hoped – and certainly intended – that those concerned with the fair, transparent, valid and reliable selection of students will find the following chapters accessible and useful. Chapter 2 addresses the importance of understanding the BMAT test taker, as the first of six components to consider from the socio-cognitive framework.

Appendix

Questions to be considered in building validity evidence for a test (adapted from Weir 2005)

Test taker characteristics	<p>What are the characteristics of the test takers (age, gender, etc.)?</p> <p>Does the test make suitable accommodations for candidates with special needs?</p> <p>Are candidates sufficiently familiar with what they have to do in the test?</p> <p>Are candidates put at ease so that they are enabled to achieve their best?</p>
Context validity	<p>Is there any evidence that the response format is likely to affect performance?</p> <p>Are the marking criteria explicit for the candidates and the markers?</p> <p>Is the timing of each part appropriate?</p> <p>Is the content knowledge suitable and unbiased?</p> <p>Are the administration conditions satisfactorily consistent and secure?</p>
Cognitive validity	<p>What are the skills/cognitive processes elicited by the test tasks?</p>
Scoring validity	<p>Are items of appropriate difficulty and do they discriminate between candidates?</p> <p>Is there a sufficient level of test reliability?</p> <p>Is there any evidence of item bias?</p> <p>Are the candidates' responses their own?</p> <p>Are there clearly defined marking criteria that cover the construct?</p> <p>Are markers trained, standardised, checked and moderated?</p> <p>Is marking reliable and consistent?</p>
Criterion validity	<p>Do test scores relate to future outcomes? (predictive)</p> <p>Do test scores relate to other tests or measurements? (concurrent)</p>
Consequential validity	<p>Are actions based on test scores appropriate?</p> <p>Is there any evidence of differential validity?</p> <p>How are candidates preparing for the test?</p> <p>Is there a washback effect in the classroom (positive or negative)?</p> <p>How is the test perceived by stakeholders?</p>

References

- Admissions Testing Service (2016a) *BMAT Section 1 Question Guide*, available online: www.admissionstestingservice.org/images/324081-bmat-section-1-question-guide.pdf
- Admissions Testing Service (2016b) *Biomedical Admissions Test (BMAT) Test Specification*, available online: www.admissionstestingservice.org/images/47829-bmat-test-specification.pdf
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1966) *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1985) *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014) *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association.
- Anastasi, A and Urbina, S (1997) *Psychological Testing*, New York: Macmillan.
- Andrich, D A (2004) Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care* 42 (1), 1–15.
- Andrich, D A (2009a) *Interpreting RUMM2030 Part I: Dichotomous Data*, Perth: RUMM Laboratory.
- Andrich, D A (2009b) *Interpreting RUMM2030 Part VI: Quantifying Response Dependence in RUMM*, Perth: RUMM Laboratory.
- Angoff, W H (1974) The development of statistical indices for detecting cheaters, *Journal of the American Statistical Association* 69 (345), 44–49.
- Arthur, N and Everaert, P (2012) Gender and performance in accounting examinations: Exploring the impact of examination format, *Accounting Education: An International Journal* 21 (5), 471–487.
- Association of American Medical Colleges (2014) *Core Competencies for Entering Medical Students*, available online: www.staging.aamc.org/initiatives/admissionsinitiative/competencies/
- Association of American Medical Colleges (2016) *Using MCAT® Data in 2017 Medical Student Selection*, available online: www.aamc.org/download/462316/data/2017mcatguide.pdf
- Atkinson, R C and Geiser, S (2009) Reflections on a century of college admissions tests, *Educational Researcher* 38 (9), 665–676.
- Bachman, L (1990) *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press.
- Bachman, L and Palmer, A (1996) *Language Testing in Practice*, Oxford: Oxford University Press.

Applying the socio-cognitive framework to BMAT

- Baldiga, K (2014) Gender differences in willingness to guess, *Management Science* 60, 434–448.
- Ball, L J (2014) Eye-tracking and reasoning: What your eyes tell about your inferences, in Neys, W D and Osman, M (Eds) *New Approaches in Reasoning Research*, Hove: Psychology Press, 51–69.
- Ball L J and Stuppel, E J N (2016) Dual-reasoning processes and the resolution of uncertainty: The case of belief bias, in Macchi, L, Bagassi, M and Viale, R (Eds) *Cognitive Unconscious and Human Rationality*, Cambridge: MIT Press, 143–166.
- Barrett, G V, Phillips, J S and Alexander, R A (1981) Concurrent and predictive validity designs: A critical reanalysis, *Journal of Applied Psychology* 66, 1–6.
- Bax, S (2013) The cognitive processing of candidates during reading tests: Evidence from eye-tracking, *Language Testing* 30 (4), 441–465.
- Bell, C (2015) A modern perspective on statistical malpractice detection, *Research Notes* 59, 31–35.
- Bell, J F (2007) Difficulties in evaluating the predictive validity of selection tests, *Research Matters* 3, 5–9.
- Bell, J F, Bramley, T, Claessen, M J A and Raikes, N (2007) Quality control of examination marking, *Research Matters* 4, 18–21.
- Bell, J F, Judge, S, Parks, G, Cross, B, Laycock, J F, Yates, D and May, S (2005) The case against the BMAT: Not withering but withered? available online: www.bmj.com/rapid-response/2011/10/31/case-against-bmat-not-withering-withered
- Ben-Shakhar, G and Sinai, Y (1991) Gender differences in multiple-choice tests: The role of differential guessing tendencies, *Journal of Educational Measurement* 28, 23–35.
- Best, R, Walsh, J L, Harris, B H J and Wilson, D (2016) UK Medical Education Database: An issue of assumed consent [Letter to the editor], *Clinical Medicine* 16 (6), 605.
- Black, B (2008) *Critical Thinking – a definition and taxonomy for Cambridge Assessment: Supporting validity arguments about Critical Thinking assessments administered by Cambridge Assessment*, Paper presented at 34th International Association of Educational Assessment Annual Conference, Cambridge, 9 September 2008, available online: www.cambridgeassessmentjobs.org/Images/126340-critical-thinking-a-definition-and-taxonomy.pdf
- Black, B (2012) An overview of a programme of research to support the assessment of critical thinking, *Thinking Skills and Creativity* 7 (2), 122–133.
- Blanden, J and Gregg, P (2004) Family income and educational attainment: A review of approaches and evidence for Britain, *Oxford Review of Economic Policy* 20 (2), 245–263.
- Bol'shev, L N (2001) Statistical estimator, in Hazewinkel, M (Ed) *Encyclopedia of Mathematics*, New York: Springer, available online: www.encyclopediaofmath.org/index.php/Statistical_estimator
- Bond, T G and Fox, C M (2001) *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Mahwah: Lawrence Erlbaum.
- Borsboom, D, Mellenbergh, G J and van Heerden, J (2004) The concept of validity, *Psychological Review* 111 (4), 1,061–1,071.
- Bramley, T and Oates, T (2011) Rank ordering and paired comparisons – the way Cambridge Assessment is using them in operational and experimental work, *Research Matters* 11, 32–35.
- Bramley, T, Vidal Rodeiro, C L and Vitello, S (2015) *Gender differences in GCSE*, Cambridge: Cambridge Assessment internal report.

- Bridges, G (2010) Demonstrating cognitive validity of IELTS Academic Writing Task 1, *Research Notes* 42, 24–33.
- Briggs, D C (2001) The effect of admissions test preparation: Evidence from NELS:88, *Chance* 14 (1), 10–18.
- Briggs, D C (2004) Evaluating SAT coaching: Gains, effects and self-selection, in Zwick, R (Ed) *Rethinking the SAT: The Future of Standardized Testing in University Admissions*, London: Routledge, 217–234.
- British Medical Association (2009) *Equality and Diversity in UK Medical Schools*, London: British Medical Association.
- Buck, G, Kostin, I and Morgan, R (2002) *Examining the Relationship of Content to Gender-based Performance Differences in Advanced Placement Exams*, College Board Research Report 2002-12, ETS RR-02-25, Princeton: Educational Testing Service.
- Butler, H A (2012) Halpern critical thinking assessment predicts real-world outcomes of critical thinking, *Applied Cognitive Psychology* 25 (5), 721–729.
- Butterworth, J and Thwaites, G (2010) *Preparing for the BMAT: The Official Guide to the BioMedical Admissions Test*, Oxford: Heinemann.
- Cambridge Assessment (2009) *The Cambridge Approach: Principles for Designing, Administering and Evaluating Assessment*, Cambridge: Cambridge Assessment, available online: www.cambridgeassessment.org.uk/Images/cambridge-approach-to-assessment.pdf
- Cambridge English (2014) *Instructions for Secure Administration of Admissions Tests*, Cambridge: UCLES.
- Cambridge English (2016) *Principles of Good Practice: Research and Innovation in Language Learning and Assessment*, Cambridge: UCLES, available online: www.cambridgeenglish.org/images/22695-principles-of-good-practice.pdf
- Cambridge International Examinations (2016) *Cambridge International AS and A Level Thinking Skills*, available online: www.cie.org.uk/images/329504-2019-syllabus.pdf
- Chapman, J (2005) *The Development of the Assessment of Thinking Skills*, Cambridge: UCLES.
- Cheung, K Y F (2014) *Understanding the authorial writer: A mixed methods approach to the psychology of authorial identity in relation to plagiarism*, unpublished doctoral thesis, University of Derby.
- Cizek, G J (1999) *Cheating on Tests: How to Do It, Detect It, and Prevent It*, London: Lawrence Erlbaum.
- Cizek, G J (2012) Defining and distinguishing validity: Interpretations of score meaning and justifications of test use, *Psychological Methods* 17 (1), 31–43.
- Cleary, T A (1968) Test bias: Prediction of grades of Negro and white students in integrated colleges, *Journal of Educational Measurement* 5, 115–124.
- Cleland, J A, French, F H and Johnston, P W (2011) A mixed methods study identifying and exploring medical students' views of the UKCAT, *Medical Teacher* 33 (3), 244–249.
- Cleland, J, Dowell, J S, McLachlan, J C, Nicholson, S and Patterson, F (2012) *Identifying best practice in the selection of medical students (literature review and interview survey)*, available online: www.gmc-uk.org/Identifying_best_practice_in_the_selection_of_medical_students.pdf_51119804.pdf
- Coates, H (2008) Establishing the criterion validity of the Graduate Medical School Admissions Test (GAMSAT), *Medical Education* 42, 999–1,006.

Applying the socio-cognitive framework to BMAT

- College Board (2015) *Test Specifications for the Redesigned SAT*, New York: College Board.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.
- Cronbach, L J (1951) Coefficient alpha and the internal structure of tests, *Psychometrika* 16 (3), 297–334.
- Cronbach, L J (1998) *Essentials of Psychological Testing*, New York: Harper and Row.
- Cronbach, L J and Shavelson, R J (2004) My current thoughts on coefficient alpha and successor procedures, *Educational and Psychological Measurement* 64 (3), 391–418.
- Department for Education (2014) *Do academies make use of their autonomy?*, available online: www.gov.uk/government/uploads/system/uploads/attachment_data/file/401455/RR366_-_research_report_academy_autonomy.pdf
- Department of Labor, Employment and Training Administration (1999) *Testing and Assessment: An Employer's Guide to Good Practices*, Washington, DC: Department of Labor, Employment and Training Administration.
- DeVellis, R F (2012) *Scale Development: Theory and Applications* (3rd edition), London: Sage Publications.
- Devine, A and Gallacher, T (2017) *The predictive validity of the BioMedical Admissions Test (BMAT) for Graduate Entry Medicine at the University of Oxford*, Cambridge: Cambridge Assessment internal report.
- Dowell, J S, Norbury, M, Steven, K and Guthrie, B (2015) Widening access to medicine may improve general practitioner recruitment in deprived and rural communities: Survey of GP origins and current place of work, *BMC Medical Education* 15 (1), available online: bmcmededuc.biomedcentral.com/track/pdf/10.1186/s12909-015-0445-8?site=bmcmededuc.biomedcentral.com
- Downing, S M (2002) Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? *Academic Medicine* 77, S103–S104.
- Downing, S M (2003) Validity: On the meaningful interpretation of assessment data, *Medical Education* 37, 830–837.
- Du Plessis, S and Du Plessis, S (2009) A new and direct test of the ‘gender bias’ in multiple-choice questions, *Stellenbosch Economic Working Papers* 23/09, available online: ideas.repec.org/p/sza/wpaper/wpapers96.html
- Dunbar, K and Fugelsang, J (2005) Scientific thinking and reasoning, in Holyoak, K J and Morrison, R G (Eds) *The Cambridge Handbook of Thinking and Reasoning*, Cambridge: Cambridge University Press, 705–725.
- Dweck, C S (2012) *Mindset: Changing the Way You Think to Fulfil Your Potential*, London: Little, Brown Book Group.
- Ebel, R L and Frisbie, D A (1991). *Essentials of Educational Measurement* (5th edition), Englewood Cliffs: Prentice-Hall.
- Eccles, J S (2011) Gendered educational and occupational choices: Applying the Eccles et al model of achievement-related choices, *International Journal of Behavioral Development* 35, 195–201.
- Eccles, J S, Adler, T F, Futterman, R, Goff, S B, Kaczala, C M, Meece, J L and Midgley, C (1983) Expectations, values, and academic behaviors, in Spence, J T (Ed) *Achievement and Achievement Motives: Psychological and Sociological Approaches*, San Francisco: W H Freeman, 75–146.

- Elliot, J and Johnson, N (2005) *Item level data: Guidelines for staff*, Cambridge: Cambridge Assessment internal report.
- Elliott, M and Wilson, J (2013) Context validity, in Geranpayeh, A and Taylor, L (Eds) *Examining Listening: Research and Practice in Second Language Listening*, Studies in Language Testing volume 35, Cambridge: UCLES/ Cambridge University Press, 152–241.
- Elston, M A (2009) *Women and medicine: The future. A report prepared on behalf of the Royal College of Physicians*, available online: www.learning.ox.ac.uk/media/global/wwwadminoxacuk/localsites/oxfordlearninginstitute/documents/overview/women_and_medicine.pdf
- Emery, J L (2007a) *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*, Cambridge: Cambridge Assessment internal report.
- Emery, J L (2007b) *A report on the predictive validity of the BMAT (2005) for 1st year examination performance on the Medicine and Veterinary Medicine course at the University of Cambridge*, Cambridge: Cambridge Assessment internal report.
- Emery, J L (2007c) *Analysis of the relationship between BMAT scores, A level points and 1st year examination performance at the Royal Veterinary College (2005 entry)*, Cambridge: Cambridge Assessment internal report.
- Emery, J L (2010a) *A Level candidates attaining 3 or more 'A' grades in England 2006-2009*, Cambridge: Cambridge Assessment internal report.
- Emery, J L (2010b) *An investigation into candidates' preparation for the BioMedical Admissions Test (2007 session): A replication involving all institutions*, Cambridge: Admissions Testing Service internal report.
- Emery, J L (2013a) *Are BMAT time constraints excessive?*, Cambridge: Cambridge English internal report.
- Emery, J L (2013b) *BMAT test-taker characteristics and the performance of different groups 2003–2012*, Cambridge: Cambridge English internal report.
- Emery, J L and Bell, J F (2009) The predictive validity of the BioMedical Admissions Test for pre-clinical examination performance, *Medical Education* 43 (6), 557–564.
- Emery, J L and Bell, J F (2011) Comment on I C McManus, Eamonn Ferguson, Richard Wakeford, David Powis and David James (2011). Predictive validity of the BioMedical Admissions Test (BMAT): An Evaluation and Case Study. *Medical Teacher* 33 (1): (this issue), *Medical Teacher* 33, 58–59.
- Emery, J L and Khalid, M N (2013a) *An investigation into BMAT item bias using DIF analysis*, Cambridge: Cambridge English internal report.
- Emery, J L and Khalid, M N (2013b) *Construct investigation into BMAT using Structural Equation Modelling*, Cambridge: Cambridge English internal report.
- Emery, J L and McElwee, S (2014) *Student perceptions of selection criteria for medical study: Are admissions tests a deterrent to application?*, Cambridge: Cambridge English internal report.
- Emery, J L, Bell, J F and Vidal Rodeiro, C L (2011) The BioMedical Admissions Test for medical student selection: Issues of fairness and bias, *Medical Teacher* 33, 62–71.
- Evans, J S B T and Ball, L J (2010) Do people reason on the Wason selection task? A new look at the data of Ball et al (2003), *The Quarterly Journal of Experimental Psychology* 63 (3), 434–441.

Applying the socio-cognitive framework to BMAT

- Evans, J S B T, Barston, J L and Pollard, P (1983) On the conflict between logic and belief in syllogistic reasoning, *Memory and Cognition* 11 (3), 295–306.
- Facione, P A (1990) *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*, California: The California Academic Press.
- Facione, P A (2000) The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill, *Informal Logic* 20 (1), 61–84.
- Ferguson, E and Lievens, F (2017) Future directions in personality, occupational and medical selection: myths, misunderstandings, measurement, and suggestions, *Advances in Health Science Education* 22 (2), 387–399.
- Field, A (2013) *Discovering Statistics Using IBM SPSS Statistics*, London: Sage.
- Field, J (2011) Cognitive validity, in Taylor, L (Ed) *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing volume 30, Cambridge: UCLES/Cambridge University Press, 112–170.
- Fisher, A (1990a) *Research into a higher studies test: A summary*, Cambridge: UCLES internal report.
- Fisher, A (1990b) *Proposal to develop a higher studies test: A discussion document*, Cambridge: UCLES internal report.
- Fisher, A (1992) *Development of the syndicate's higher education aptitude tests*, Cambridge: UCLES internal report.
- Fisher, A (2005) *'Thinking skills' and admission to higher education*, Cambridge: UCLES internal report.
- Fitzpatrick, A R (1983) The meaning of content validity, *Applied Psychological Measurement* 7 (1), 3–13.
- Furneaux, C and Rignall, M (2007) The effect of standardisation-training on rater judgements for the IELTS Writing Module, in Taylor, L and Falvey, P (Eds) *IELTS Collected Papers*, Cambridge: UCLES/Cambridge University Press, Studies in Language Testing Volume 19, 422–445.
- Galaczi, E and French, A (2011) Context validity, in Taylor, L (Ed) *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing volume 30, Cambridge: UCLES/Cambridge University Press, 112–170.
- Gale, M and Ball, L J (2009) Exploring the determinants of dual goal facilitation in a rule discovery task, *Thinking and Reasoning* 15 (3), 294–315.
- Gallacher, T, McElwee, S and Cheung, K Y F (2017) BMAT 2015 test preparation survey report, Cambridge: Cambridge Assessment internal report.
- Garner, R (2015) Number of pupils attending independent school in Britain on the rise, figures show, *The Independent*, 30 April 2015, available online: www.independent.co.uk/news/education/education-news/number-of-pupils-attending-independent-schools-in-britain-on-the-rise-figures-show-10215959.html
- General Medical Council (2009) *Tomorrow's Doctors: Outcomes and Standards for Undergraduate Medical Education*, available online: www.gmc-uk.org/Tomorrow_s_Doctors_1214.pdf_48905759.pdf
- General Medical Council (2011) *The State of Medical Education and Practice in the UK*, London: General Medical Council.
- Geranpayeh, A (2013) Detecting plagiarism and cheating, in Kunnan, A J (Ed) *The Companion to Language Assessment*, London: Wiley Blackwell, 980–993.

- Geranpayeh, A (2014) Detecting plagiarism and cheating: Approaches and development, in Kunnan, A J (Ed) *The Companion to Language Assessment Volume II*, Chichester: Wiley, 980–993.
- Geranpayeh, A and Taylor, L (Eds) (2013) *Examining Listening: Research and Practice in Assessing Second Language Listening*, Studies in Language Testing volume 35, Cambridge: UCLES/Cambridge University Press.
- Gilhooly, K J, Fioratou, E and Henretty, N (2010) Verbalization and problem solving: Insight and spatial factors, *British Journal of Psychology* 101 (1), 81–93.
- Gill, T, Vidal Rodeiro, C L and Zanini, N (2015) *Students' choices in Higher Education*, paper presented at the BERA conference, Queen's University Belfast, available online: cambridgeassessment.org.uk/Images/295319-students-choices-in-higher-education.pdf
- Goel, V, Navarrete, G, Noveck, I A and Prado, J (2017) Editorial: The reasoning brain: The interplay between cognitive neuroscience and theories of reasoning, *Frontiers in Human Neuroscience* 10, available online: journal.frontiersin.org/article/10.3389/fnhum.2016.00673/full
- 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.
- Green, A (2003) *Test impact and English for academic purposes: A comparative study in backwash between IELTS preparation and university professional courses*, Unpublished doctoral dissertation, University of Surrey.
- Green, A (2006) Watching for washback: Observing the influence of the International English Language Testing System Academic Writing Test in the classroom, *Language Assessment Quarterly* 3 (4), 333–368.
- Green, A (2007) Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses, *Assessment in Education: Principles, Policy and Practice* 1, 75–97.
- Green, A (2013) Washback in language assessment, *International Journal of English Studies* 13 (2), 39–51.
- Griffin, B and Hu, W (2015) The interaction of socio-economic status and gender in widening participation in medicine, *Medical Education* 49 (1), 103–113.
- Halpern, D F (1999) Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker, *New Directions for Teaching and Learning* 80, 69–74.
- Hambleton, R K and Traub, R E (1974) The effect of item order on test performance and stress, *The Journal of Experimental Education* 43 (1), 40–46.
- Hambleton, R K, Swaminathan, H and Rogers, H (1991) *Fundamentals of Item Response Theory*, Newbury Park: Sage Publications.
- Hamilton, J S (1993) *MENO Thinking Skills Service: Development and Rationale*, Cambridge: UCLES internal report.
- Hawkey, R (2011) Consequential validity, in Geranpayeh, A and Taylor, L (Eds) *Examining Listening: Research and Practice in Assessing Second Language Listening*, Studies in Language Testing volume 35, Cambridge: UCLES/Cambridge University Press, 273–302.
- Haynes, S N, Richard, D C S and Kubany, E S (1995) Content validity in psychological assessment: A functional approach to concepts and methods, *Psychological Assessment* 7 (3), 238–247.

Applying the socio-cognitive framework to BMAT

- Hecker, K and Norman, G (2017) Have admissions committees considered all the evidence? *Advances in Health Sciences Education* 22 (2), 573–576.
- Hembree, R (1988) Correlates, causes, effects, and treatment of test anxiety, *Review of Educational Research* 58, 47–77.
- Hirschfeld, M, Moore, R L and Brown, E (1995) Exploring the gender gap on the GRE subject test in economics, *Journal of Economic Education* 26 (1), 3–15.
- Hoare, A and Johnston, R (2011) Widening participation through admissions policy – a British case study of school and university performance, *Higher Education Quarterly* 36, 21–41.
- Hojat, M, Erdmann, J B, Veloski, J J, Nasca, T J, Callahan, C A, Julian, E R and Peck, J. (2000) A validity study of the writing sample section of the Medical College Admission Test, *Academic Medicine*, 75, 25S–27S.
- Holland, P W and Thayer, D T (1988) Differential item performance and Mantel-Haenszel procedure, in Wainer, H and Braun, I (Eds) *Test Validity*, Hillsdale: Lawrence Erlbaum, 129–145.
- Holland, P W and Wainer, H (Eds) (1993) *Differential Item Functioning*, Hillsdale: Lawrence Erlbaum.
- Hopkins, K, Stanley, J, Hopkins, B R (1990) *Educational and Psychological Measurement and Evaluation*, Englewood Cliffs: Prentice-Hall.
- Hu, L T and Bentler, P (1999) Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modelling* 6, 1–55.
- Hughes, A (2003) *Testing for Language Teachers* (2nd edition), Cambridge: Cambridge University Press.
- Hyde, J S, Lindberg, S M, Linn, M C, Ellis, A B, and Williams, C C (2008) Gender similarities characterize math performance, *Science* 321, 494–495.
- Independent Schools Council (2015) *ISC Census 2015*, available online: www.isc.co.uk/media/2661/isc_census_2015_final.pdf
- Independent Schools Council (2016) *ISC Census 2016*, available online: www.isc.co.uk/media/3179/isc_census_2016_final.pdf
- James, W and Hawkins, C (2004) Assessing potential: The development of selection procedures for the Oxford medical course, *Oxford Review of Education* 30, 241–255.
- Jencks, C and Crouse, J (1982) Aptitude vs. achievement: should we replace the SAT? *The Public Interest* 67, 21–35.
- Joint Council for Qualifications (2016a) *Adjustments for candidates with disabilities and learning difficulties: Access arrangements and reasonable adjustments*, available online: www.jcq.org.uk/exams-office/access-arrangements-and-special-consideration
- Joint Council for Qualifications (2016b) *General and vocational qualifications: General regulations for approved centres*, available online: www.jcq.org.uk/exams-office/general-regulations
- Julian, E R (2005) Validity of the Medical College Admission Test for predicting medical school performance, *Academic Medicine* 80, 910–917.
- Kane, M (2013) Validating the interpretations and uses of test scores, *Journal of Educational Measurement* 50, 1–73.
- Kaplan, R M and Saccuzzo, D P (2012) *Psychological Testing: Principles, Applications, and Issues*, California: Wadsworth Publishing Company.
- Katz, S and Vinker, S (2014) New non-cognitive procedures for medical applicant selection: A qualitative analysis in one school, *BMC Medical Education*, available online: www.ncbi.nlm.nih.gov/pubmed/25376161

- Kellogg, J S, Hopko, D R and Ashcraft, M H (1999) The effects of time pressure on arithmetic performance, *Journal of Anxiety Disorders* 13 (6), 591–600.
- Kelly, M E, Gallagher, N, Dunne, F and Murphy, A (2014) Views of doctors of varying disciplines on HPAT-Ireland as a selection tool for medicine, *Medical Teacher* 36 (9), 775–782.
- Kelly, S and Dennick, R. (2009). Evidence of gender bias in True-False-Abstain medical examinations, *BMC Medical Education*, available online: www.ncbi.nlm.nih.gov/pmc/articles/PMC2702355/
- Khalifa, H and Weir, C J (2009) *Examining Reading: Research and Practice in Assessing Second Language Reading*, Studies in Language Testing volume 29. Cambridge: UCLES/Cambridge University Press.
- Klahr, D and Dunbar, K (1988) Dual space search during scientific reasoning, *Cognitive Science* 12 (1), 1–48.
- Klein, S, Liu, O L, Sconing, J, Bolus, R, Bridgeman, B, Kugelmass, H and Steedle, J (2009) *Test Validity Study (TVS) Report*, Washington, DC: US Department of Education.
- Koenig, T W, Parrish, S K, Terregino, C A, Williams, J P, Dunleavy, D M and Volsch, J M (2013) Core personal competencies important to entering students' success in medical school: What are they and how could they be assessed early in the admission process? *Academic Medicine* 88 (5), 603–613.
- Kreiter, C D and Axelson, R D (2013) A perspective on medical school admission research and practice over the last 25 years, *Teaching and Learning in Medicine* 25, S50–S56.
- Ku, K Y L (2009) Assessing students' critical thinking performance: Urging for measurements using multi-response format, *Thinking Skills and Creativity* 4, 70–76.
- Kuncel, N R and Hezlett, S A (2010) Fact and fiction in cognitive ability testing for admissions and hiring decisions, *Current Directions in Psychological Science* (19) 6, 339–345.
- Kuncel, N R, Hezlett, S A and Ones, D S (2001) A comprehensive meta-analysis of the predictive validity of the Graduate Records Examinations: Implications for graduate student selection and performance, *Psychological Bulletin* 127, 162–181.
- Kusurkar, R A, Ten Cate, T J, van Asperen, M and Croiset, G (2011) Motivation as an independent and a dependent variable in medical education: A review of the literature, *Medical Teacher* 33 (5), 242–262.
- Lado, R (1961) *Language Testing: The Construction and Use of Foreign Language Tests. A Teacher's Book*, New York: McGraw Hill.
- Landrum, R E and McCarthy, M A (2015) Measuring critical thinking skills, in Jhangiani, R S, Troisi, J D, Fleck, B, Legg, A M and Hussey, H D (Eds) *A Compendium of Scales for Use in the Scholarship of Teaching and Learning*, available online: teachpsych.org/ebooks/compscalessotp
- Lawshe, C H (1975) A quantitative approach to content validity, *Personnel Psychology* 28, 563–575.
- Leijten, M and Van Waes, L (2013) Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes, *Written Communication* 30 (3), 358–392.
- Linacre, J M (2014) *Facets computer program for many-facet Rasch measurement*, version 3.71.4, Beaverton: Winsteps.com.
- Linacre, J M (2016) *Winsteps® Rasch Measurement Computer Program User's Guide*, Beaverton: Winsteps.com.

Applying the socio-cognitive framework to BMAT

- Linn, R L (2009) Considerations for college admissions testing, *Educational Researcher* 38 (9), 677–679.
- Liu, O L, Frankel, L and Roohr, K C (2014) Assessing critical thinking in higher education: Current state and directions for next-generation assessment, *ETS Research Report Series* 1, 1–23.
- Long, R (2017) GCSE, AS and A Level reform, House of Commons briefing paper Number SN06962, available from: researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06962
- Lord, F M and Novick, M R (1968) *Statistical Theories of Mental Test Scores*, Reading: Addison-Wesley.
- Lu, Y and Sireci, S G (2007) Validity issues in test speededness, *Educational Measurement: Issues and Practice* 26, 29–37.
- Luxia, Q (2007) Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China, *Assessment in Education: Principles, Policy and Practice* 1, 51–74.
- Mantel, N and Haenszel, W (1959) Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of the National Cancer Institute* 22 (4), 719–748.
- Massey, A J (2004) *Medical and veterinary admissions test validation study*, Cambridge: Cambridge Assessment internal report.
- Mayer, R E, Larkin, J H and Kadane, J (1984) A cognitive analysis of mathematic problem-solving ability, in Sternberg, R J (Ed) *Advances in the Psychology of Human Intelligence*, Hillsdale: Lawrence Erlbaum, 231–273.
- McCarthy, J M and Goffin, R D (2005) Selection test anxiety: Exploring tension and fear of failure across the sexes in simulated selection scenarios, *International Journal of Selection and Assessment* 13 (4), 282–295.
- McCurry, D and Chiavaroli, N (2013) Reflections on the role of a writing test for medical school admissions, *Academic Medicine* 88 (5), 568–571.
- McDonald, A S (2001) The prevalence and effects of test anxiety in school children, *Educational Psychology* 21 (1) 89–101.
- McDonald, R P (1981) The dimensionality of tests and items, *British Journal of Mathematical and Statistical Psychology* 34 (1), 100–117.
- McManus, I C, Dewberry, C, Nicholson, S and Dowell, J S (2013) The UKCAT-12 study: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a collaborative study of twelve UK medical schools, *BMC Medicine* 11, available online: bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-11-244
- McManus, I C, Dewberry, C, Nicholson, S, and Dowell, J S, Woolf, K and Potts, H W W (2013) Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies, *BMC Medicine* 11, available online: bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-11-243
- McManus, I C, Powis, D A, Wakeford, R, Ferguson, E, James, D and Richards, P (2005) Intellectual aptitude tests and A Levels for selecting UK school leaver entrants for medical school, *BMJ* 331, 555–559.
- Medical Schools Council (2014) *Selecting for Excellence Final Report*, London: Medical Schools Council.

- Mellenbergh, G J (2011) *A Conceptual Introduction to Psychometrics. Development, Analysis, and Application of Psychological and Educational Tests*, The Hague: Eleven International Publishing.
- Messick, S (1989) Validity, in Linn, R L (Ed) *Educational Measurement* (3rd edition), Washington DC: The American Council on Education and the National Council on Measurement in Education, 13–103.
- Messick, S (1995) Validity of psychological assessment: Validation of inferences from person's responses and performance as scientific inquiry into scoring meaning, *American Psychologist* 9, 741–749.
- Milburn A (2012) *Fair access to professional careers – A progress report by the Independent Reviewer on Social Mobility and Child Poverty*, London: Cabinet Office.
- Morris, B J, Croker, S, Masnick, A M and Zimmerman, C (2012) The emergence of scientific reasoning, in Kloos, H, Morris, B J and Amaral, J L (Eds) *Current Topics in Children's Learning and Cognition*, Rijeka: InTech, 61–82.
- Ndaji, F, Little, J and Coe, R (2016) *A comparison of academic achievement in independent and state schools: Report for the Independent Schools Council January 2016*, Durham: Centre for Evaluation and Monitoring, Durham University, available online: www.isc.co.uk/media/3140/16_02_26-cem-durham-university-academic-value-added-research.pdf
- Newble, D (2016) Revisiting 'The effect of assessments and examinations on the learning of medical students', *Medical Education* 50 (5), 498–501.
- Newble, D I and Jaeger, K (1983) The effect of assessments and examinations on the learning of medical students, *Medical Education* 17 (3), 165–171.
- Newton, P and Shaw, S D (2014) *Validity in Educational and Psychological Assessment*, London: Sage.
- Nicholson, S and Cleland, J (2015) Reframing research on widening participation in medical education: using theory to inform practice, in Cleland, J and Durning, S J (Eds) *Researching Medical Education*, Oxford: Wiley Blackwell, 231–243.
- Niessen, A S M and Meijer, R R (2016) Selection of medical students on the basis of non-academic skills: is it worth the trouble? *Clinical Medicine* 16(4), 339–342.
- Niessen, A S M, Meijer, R B and Tendeiro, J N (2017) Applying organizational justice theory to admission into higher education: Admission from a student perspective, *International Journal of Selection and Assessment* 25 (1), 72–84.
- Norris, S P (1990) Effect of eliciting verbal reports of thinking on critical thinking test performance, *Journal of Educational Measurement* 27 (1), 41–58.
- Novick, M R (1966) The axioms and principal results of classical test theory, *Journal of Mathematical Psychology* 3 (1), 1–18.
- Nowell, A and Hedges, L V (1998) Trends in gender differences in academic achievement from 1960 to 1994: An analysis of differences in mean, variance, and extreme scores, *Sex Roles* 39 (1/2), 21–43.
- O'Hare, L and McGuinness, C (2009) Measuring critical thinking, intelligence and academic performance in psychology undergraduates, *The Irish Journal of Psychology* 30, 123–131.
- O'Hare, L and McGuinness, C (2015) The validity of critical thinking tests for predicting degree performance: A longitudinal study, *International Journal of Educational Research* 72, 162–172.
- O'Sullivan, B and Weir, C J (2011) Test development and validation, in O'Sullivan, B (Ed) *Language Testing: Theories and Practices*, Basingstoke: Palgrave Macmillan, 13–32.

Applying the socio-cognitive framework to BMAT

- Palmer, E J and Devitt, P G (2007) Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? *BMC Medical Education* 7, bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-7-49
- Papp, S and Rixon, S (forthcoming 2017) *Assessing Young Language Learners: The Cambridge English Approach*, Studies in Language Testing volume 47, Cambridge: UCLES/Cambridge University Press.
- Patel, V L, Arocha, J F and Zhang, J (2005) Thinking and reasoning in medicine, in Holyoak, K J and Morrison, R G (Eds) *The Cambridge Handbook of Thinking and Reasoning*, Cambridge: Cambridge University Press, 727–750.
- Patterson, F, Knight, A, Dowell, J S Nicholson, S., Cousans, and Cleland, J. (2016). How effective are selection methods in medical education? A systematic review, *Medical Education* 50, 36–60.
- Paul, R and Elder, L (2007) *Critical Thinking Competency Standards (For Educators)*, Tomales: Foundation for Critical Thinking.
- Pearson VUE (2017) *UK Clinical Aptitude Test (UKCAT) Consortium UKCAT Examination Executive Summary Testing Interval: 1 July 2016–4 October 2016*, available online: www.ukcat.ac.uk/media/1057/ukcat-2016-technical-report-exec-summary_v1.pdf
- Pelacia, T and Viau, R (2017) Motivation in medical education, *Medical Teacher* 39 (2), 136–140.
- Plass, J A and Hill, K T (1986) Children's achievement strategies and test performance: The role of time pressure, evaluation anxiety and sex, *Developmental Psychology* 22 (1), 31–36.
- Powis, D A (2015) Selecting medical students: An unresolved challenge, *Medical Teacher* 37 (3), 252–260.
- Quality Assurance Agency (2002) *Subject Benchmark Statement: Medicine*, available online: www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-Medicine.pdf
- Quality Assurance Agency (2015) *Subject Benchmark Statement: Biomedical Sciences*, available online: www.qaa.ac.uk/en/Publications/Documents/SBS-Biomedical-sciences-15.pdf
- Ramsay, P A (2005) *Admissions tests (Cambridge TSA and BMAT) and disability*, Cambridge: University of Cambridge internal report.
- Rasch, G (1960/1980) *Probabilistic Models for Some Intelligence and Attainment Tests*, Chicago: University of Chicago Press.
- Rasch, G (1961) On general laws and meaning of measurement in psychology, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (4), Berkeley: University of California Press, 321–333.
- Rasch, G (2011) *All statistical models are wrong!*, available online: www.rasch.org/rmt/rmt244d.html
- Reibnegger, G, Caluba, H-C, Ithaler, D, Manhal, S, Neges, H M and Smolle, J (2010) Progress of medical students after open admission or admission based on knowledge tests, *Medical Education* 44, 205–214.
- Röding, K and Nordenram, G (2005) Students' perceived experience of university admission based on tests and interviews, *European Journal of Dental Education* 9 (4), 171–179.
- Rodriguez, M C (2003) Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations, *Journal of Educational Measurement*, 40(2), 163–184.

- Ross, J A, Scott, G and Bruce, C D (2012) The gender confidence gap in fractions knowledge: Gender differences in student belief–achievement relationships, *School Science and Mathematics* 112 (5), 278–288.
- Sackett, P R and Yang, H (2000) Correction for range restriction: An expanded typology, *Journal of Applied Psychology* 85, 112–118.
- Sam, A, Hameed, S, Harris, J, Meeran, K (2016) Validity of very short answer versus single best answer questions for undergraduate assessment, *BMC Medical Education* 16 (1), available online: bmcmededuc.biomedcentral.com/articles/10.1186/s12909-016-0793-z
- Saville, N and Hawkey, R (2004) The IELTS impact study: Investigating washback on teaching materials, in Cheng, L, Watanabe, Y and Curtis, A (Eds) *Washback in Language Testing: Research Context and Methods*, London: Lawrence Erlbaum, 73–96.
- Saville, N (2003) The process of test development and revision within UCLES EFL, in Weir, C J and Milanovic, M (Eds) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*, Studies in Language Testing volume 15, Cambridge: UCLES/Cambridge University Press, 57–120.
- Saville, N (2012) Applying a model for investigating the impact of language assessment within educational contexts: The Cambridge ESOL approach, *Research Notes* 50, 4–8.
- Scardamalia, M and Bereiter, C (1987) Knowledge telling and knowledge transforming in written composition, in Rosenberg, S (Ed) *Advances in Applied Psycholinguistics, Volume 2: Reading, Writing and Language Learning*, Cambridge: Cambridge University Press, 142–175.
- Schwartzstein, R, Rosenfeld, G, Hilborn, R, Oyewole, S and Mitchell, K. (2013) Redesigning the MCAT exam: balancing multiple perspectives, *Academic Medicine* 88 (5), 560–567.
- Scorey, S. (2009a) *Investigating the predictive validity of the BMAT: An analysis using examination data from the Royal veterinary College BVetMed course for the 2005, 2006 and 2007 BMAT cohorts*, Cambridge: Cambridge Assessment internal report.
- Scorey, S (2009b) *Investigating the predictive validity of the BMAT: An analysis using examination data from the University College London course for the 2003 to 2007 BMAT cohorts*, Cambridge: Cambridge Assessment internal report.
- Seyan K, Greenhalgh T and Dorling D (2004) The standardised admission ratio for measuring widening participation in medical schools: analysis of UK medical school admissions by ethnicity, socioeconomic status, and sex, *British Medical Journal* 328, 1,545–1,546.
- Shannon, M D (2005) *Investigation of possible indicators of excessive time pressure in BMAT*, Cambridge: Cambridge Assessment internal report.
- Shannon, M D and Scorey, S (2010) *BMAT Section 3 marking trial March 2010 – Marker reliability analysis*, Cambridge: Cambridge Assessment internal report.
- Shannon, M D (2010) (Ed) *Preparing for the BMAT: The Official Guide to the BioMedical Admissions Test*. Oxford: Heinemann.
- Sharples, J M, Oxman, A D, Mahtani, K R, Chalmers, I, Oliver, S, Collins, K, Austvoll-Dahlgren, A and Hoffmann, T (2017) Critical thinking in healthcare and education, *BMJ* 357, available online: www.bmj.com/content/357/bmj.j2234.long
- Shaw, S D (2002) The effect of standardisation on rater judgement and inter-rater reliability, *Research Notes* 8, 13–17.

Applying the socio-cognitive framework to BMAT

- Shaw, S D and Weir, C J (2007) *Examining Writing: Research and Practice in Assessing Second Language Writing*, Studies in Language Testing volume 26, Cambridge: UCLES/Cambridge University Press.
- Shea, J and Fortna, G (2002). Psychometric methods, in Norman, G R, van der Vleuten, C P and Newble, D I (Eds) (2012) *International Handbook of Research in Medical Education (Vol. 7)*, New York: Springer Science and Business Media, 97–126.
- Shultz, M M and Zedeck, S (2012) Admission to law school: New measures, *Educational Psychologist* 47 (1), 51–65.
- Simon, H A and Newell, A (1971) Human problem solving: The state of the theory in 1970, *American Psychologist* 12 (2), 145–159.
- Sireci, S G (1998) The construct of content validity, *Social Indicators Research* 45, 83–117.
- Sjitsma, K (2009) On the use, misuse, and the very limited usefulness of Cronbach's alpha, *Psychometrika* 74 (1), 107–120.
- Soares, J A (2012) The future of college admissions: Discussion, *Educational Psychologist* 47 (1), 66–70.
- Stegers-Jager, K M, Steyerberg, E W, Lucieer, S M and Themmen, A P N (2015) *Medical Education* 49 (1), 124–133.
- Stemler, S E (2012) What should university admissions tests predict? *Educational Psychologist* 47 (1), 5–17.
- Steven, K, Dowell, J S, Jackson, C and Guthrie, B (2016) Fair access to medicine? Retrospective analysis of UK medical schools application data 2009–2012 using three measures of socioeconomic status, *BMC medical education* 16 (1), available online: bmcmmeduc.biomedcentral.com/articles/10.1186/s12909-016-0536-1
- Stevens L, Kelly M E, Hennessy M, Last J, Dunne F, O'Flynn S (2014) Medical students' views on selection tools for medical school – a mixed methods study, *Irish Medical Journal* 107 (8), 229–231.
- Stoet, G and Geary, D C (2013) Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 Years of PISA data, *PLOS ONE*, available online: journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0057988&type=printable
- Stuppel, E J N, Maratos, F A, Elander, J, Hunt, T E, Cheung, K Y F and Aubeeluck, A V (2017) Development of the Critical Thinking Toolkit (CriTT): A measure of student attitudes and beliefs about critical thinking, *Thinking Skills and Creativity* 23, 91–100.
- Tai, R H, Loehr, J F and Brigham, F J (2006) An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments, *International Journal of Research and Method in Education* 29 (2), 185–208.
- Taylor, L (Ed) (2011) *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing volume 30, Cambridge: UCLES/Cambridge University Press.
- Thissen, D, Steinberg, L and Wainer, H (1993) Detection of differential item functioning using the parameters of item response models, In Holland, P and Wainer, H (Eds) *Differential Item Functioning*. Hillsdale: Lawrence Erlbaum, 67–113.
- Thomson, A and Fisher A (1992) *MENO: A validation study of informal reasoning items*, Norwich: University of East Anglia internal report.
- Tiffin, P A, McLachlan, J C, Webster, L and Nicholson, S (2014) Comparison of the sensitivity of the UKCAT and A Levels to sociodemographic

- characteristics: A national study, *BMC Medical Education* 14, available online: bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-14-7
- Tighe, J, McManus, I C, Dewhurst, N G, Chis, L and Mucklow, J (2010) The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP (UK) examinations, *BMC Medical Education* 10, available online: bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-10-40
- Trainer, S (2015) Student data privacy is cloudy today, clearer tomorrow, *The Phi Delta Kappan* 96 (5), 13–18.
- Tsai, M-J, Hou, H-T, Lai, M-L, Liu, W-Y and Yang, F-Y (2012) Visual attention for solving multiple-choice science problem: An eye-tracking analysis, *Computers and Education* 58 (1), 375–385.
- Universities and Colleges Admissions Service (2016) *Applicant numbers to 'early deadline' university courses increase by 1%, UCAS figures reveal today*, available online: www.ucas.com/corporate/news-and-key-documents/news/applicant-numbers-%E2%80%99early-deadline%E2%80%99-university-courses-increase
- Weigle, S C (1994) Effects of training on raters of ESL compositions, *Language Testing* 11 (2), 197–223.
- Weigle, S C (1999) Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing* 6 (2), 145–178.
- Weigle, S C (2002) *Assessing Writing*, Cambridge: Cambridge University Press.
- Weir, C J (2005) *Language Testing and Validation: An Evidence-based Approach*, Basingstoke: Palgrave Macmillan.
- Weir, C J and Taylor, L (2011) Conclusions and recommendations, in Taylor, L (Ed) *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing Volume 30, Cambridge: UCLES/Cambridge University Press, 293–313.
- Wilhelm, O and Oberauer, K (2006) Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks, *European Journal of Cognitive Psychology* 18 (1), 18–50.
- Willmott, A (2005) *Thinking Skills and admissions: A report on the validity and reliability of the TSA and MVAT/BMAT assessments*, Cambridge: Cambridge English internal report.
- Woolf, K, Potts, H W W, Stott, J, McManus, I C, Williams, A and Scior, K (2015) The best choice? *The Psychologist* 28, 730–735.
- Wouters, A, Croiset, G, Galindo-Garre, F and Kusrkar, R A (2016) Motivation of medical students: Selection by motivation or motivation by selection, *BMC Medical Education* 16 (1), available online: www.ncbi.nlm.nih.gov/pubmed/26825381
- Wouters, A, Croiset, G, Schripsema, N R, Cohen-Schotanus, J, Spaai, G W G, Hulsman R L and Kusrkar, R A (2017) A multi-site study on medical school selection, performance, motivation and engagement, *Advances in Health Sciences Education* 22 (2), 447–462.
- Wright, S (2015) Medical school personal statements: a measure of motivation or proxy for cultural privilege? *Advances in Health Sciences Education* 20, 627–643.
- Yeager, D S and Dweck, C S (2012) Mindsets that promote resilience: When students believe that personal characteristics can be developed, *Educational Psychologist*, 47(4), 302–314.

Applying the socio-cognitive framework to BMAT

- Yu, G, He, L and Isaacs, T (2017). *The Cognitive Processes of taking IELTS Academic Writing Task 1: An Eye-tracking Study*, IELTS Research Reports Online Series, British Council, IDP: IELTS Australia and Cambridge English Language Assessment, available online: www.ielts.org/-/media/research-reports/ielts_online_rr_2017-2.ashx
- Zeidner, M (1998) *Test Anxiety: The State of the Art*, New York: Plenum.
- Zimmerman, C (2000) The development of scientific reasoning skills, *Developmental Review* 20, 99–149.
- Zimmerman, C (2007) The development of scientific thinking skills in elementary and middle school, *Developmental Review* 27, 172–223.
- Zinbarg, R E, Revelle, W, Yovel, I and Li, W (2005) Cronbach's α , Revelle's β , and McDonald's ω^2 : Their relations with each other and two alternative conceptualizations of reliability, *Psychometrika* 70 (1), 123–133.
- Zohar, A and Peled, B (2008) The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students, *Learning and Instruction* 18 (4), 337–352.
- Zumbo, B D and Rupp, A A (2004) Responsible modelling of measurement data for appropriate inferences: Important advances in reliability and validity theory, in Kaplan, D (Ed) *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, Thousand Oaks: Sage Press, 73–92.
- Zwick, R (Ed) (2004) *Rethinking the SAT: The Future of Standardized Testing in University Admissions*, London: Routledge.
- Zwick, R and Ercikan, K (1989) Analysis of differential item functioning in the NAEP history assessment, *Journal of Educational Measurement* 26, 55–66.
- Zwick, R, Thayer, D T and Lewis, C (1999) An empirical Bayes approach to Mantel-Haenszel DIF analysis, *Journal of Educational Measurement* 36 (1), 1–28.