WHAT IMPACT HAS THE UNITED KINGDOM CLINICAL APTITUDE TEST (UKCAT) HAD ON SELECTION TO UNDERGRADUATE MEDICINE AND DENTISTRY IN THE UK?

RACHEL GREATRIX

Thesis submitted to the University of Nottingham

for the Degree of Doctor of Philosophy

January 2022

Abstract

The United Kingdom Clinical Aptitude Test (UKCAT) Consortium was established in 2005 with the aims of achieving greater fairness in selection to medicine and dentistry, widening participation in medical and dental training and improving the evidence base around selection in the UK. The UKCAT was first used in selection in 2006.

This thesis aimed to determine the impact UKCAT had on selection to medicine and dentistry in the UK, to document the development of the organisation and to establish the extent to which UKCAT had achieved its original aims.

Having established the research evidence base in existence in 2005, documents were reviewed to present a picture of both organisational and test development. A further literature review specific to UKCAT was undertaken alongside a systematic review of the literature relating to the predictive validity of the test. Drawing on data collected on an annual basis from universities, how the use of the test in selection had changed was described. A qualitative analysis of interviews with admission tutors was used to further evaluate the impact of the test and in particular the extent to which UKCAT had met its original aims.

Findings indicate that UKCAT now plays a significant part in decisions made regarding most applicants to medical and dental training. The test has largely replaced the scoring of personal statements in selection processes and there is evidence that it predicts performance in medical school. The test is both an effective and efficient tool for discriminating between the many high performing

candidates. Use of the test has not widened participation and universities have increasingly used the test differently in selection for this candidate subgroup.

The landscape around selection continues to develop, impacted by developments in technology, COVID, increasing applicant numbers and the potential of post-qualification admissions. Drawing on the findings in the thesis, proposals regarding the development of the test, priorities for future research and the candidate experience are put forward to inform future developments in UKCAT. These include further evaluation of subgroup performance differences in the test, commissioning of future predictive validity cohort studies and exploring coaching effects on test performance.

Acknowledgements

I would like to thank my supervisors Professor Reg Dennick, Professor Susan Anderson and Professor Brigitte Scammell for their advice, support and good humour throughout this journey.

Professor Sandra Nicholson provided me with significant guidance and advice and collaborated on one of my publications. Other collaborators, whose contributions have been invaluable include Professor Jon Dowell and Professor Robert McAndrew.

My thanks to UKCAT for funding my PhD and for the support of the Board and in particular Mr Nigel Siesage (UKCAT Chair) in undertaking this work.

Thanks also to those who have worked in the UKCAT Office over the years and in particular to my colleagues Helen Beeston and Esther Hamilton Dick.

A significant source in this thesis have been documents produced by colleagues at Pearson VUE responsible for the development of the test and monitoring its performance. They are too numerous to name, but their thoroughness and consistency have been much appreciated as I have worked through various documents.

Finally, my thanks to the admission tutors and managers with whom UKCAT collaborates, many of whom have contributed to interviews within this thesis. Their commitment to improving medical and dental selection continues to drive the work of UKCAT.

Conflicts of Interest

Rachel Greatrix is the paid Chief Operating Officer of the UKCAT Consortium and has been in that post since 2008. The author's senior and longstanding position in the organisation presents clear advantages to the undertaking of this work both in terms of their access to materials and understanding of context. However, this position creates the potential of bias in research design, analysis and interpretation. This is commented on and explored further in Chapter 2 and throughout the thesis.

Table of Contents

List of Figures

List of Tables

Abbreviations

AAMC	Association of American Medical Colleges
ACER	Australian Council for Education Research
AMEE	Association for Medical Education in Europe
ANZ	Australia and New Zealand
ASME	Association for the Study of Medical Education
BMA	British Medical Association
BMAT	Biomedical Admissions Test
BMJ	British Medical Journal
DAT	Dental Admission Test
DSC	Dental Schools Council
EAL	English as an Additional Language
EI	Emotional Intelligence
EMA	Educational Maintenance Allowance
EPM	Education Performance Measure
GAMSAT	Graduate Medical Schools Admission Test
GDC	General Dental Council
GDPR	General Data Protection Regulation
GMC	General Medical Council
GPA	Grade Point Average
HDS	Hogan Development Survey
HIC	Health Informatics Centre
HYMS	Hull York Medical School
ITQ	Interpersonal Traits Questionnaire
MCAT	Medical College Admission Test
MEARS	Managing Emotions and Resilience Scales
MMI	Multiple Mini Interview
MSC	Medical Schools Council
MSORD	Medical Schools Outcomes Research database
NACE	Narcissism, Aloofness, Confidence and Empathy
OSCE	Objective Structured Clinical Examination
OSLERS	Objective Structured Long Examination Records
PQA	Personal Qualities Assessment
PSA	Prescribing Safety Assessment
RWP	Research Working Party
SEC	Socio-Economic Class
SJT	Situational Judgement Test
SPA	Supporting Professionalism in Admissions
TSA	Thinking Skills Assessment
UCAT	University Clinical Aptitude Test
UCL	University College London
UCLES	University of Cambridge Local Examinations Syndicate
UJC	Ucat Joint Committee
UKFPO	United Kingdom Foundation Programme Office
UKCAT	United Kingdom Clinical Aptitude Test
UKCATSA	United Kingdom Clinical Aptitude Test Special Arrangements (test with rest breaks between sections)

UKCATSEN	United Kingdom Clinical Aptitude Test Special Educational Needs (test with 25% additional time)
UKCATSENSA	United Kingdom Clinical Aptitude Test Special Educational Needs Special Arrangements (test with 25% additional time and rest breaks between sections)
UKCATSEN50	United Kingdom Clinical Aptitude Test Special Educational Needs 50% additional time (test with 50% additional time)
UKMACS	United Kingdom Medical Applicant Cohort Study
UKMED	United Kingdom Medical Education Database.
UMAT	Universities Medical Admissions Test
WA	Widening Access
WP	Widening Participation
WPG	Work Psychology Group

Publications

During the writing of this thesis the following peer reviewed publications were produced:

Does the UKCAT predict performance in medical and dental school? A systematic review

BMJ Open January 2021, Rachel Greatrix, Sandra Nicholson and Susan Anderson

https://bmjopen.bmj.com/content/bmjopen/11/1/e040128.full.pdf

UKCAT and medical student selection in the UK - what has changed since 2006?

BMC Medical Education, Article number 292 (2020), Rachel Greatrix and Jonathan Dowell

https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-020-02214-1

UCAT and dental student selection in the UK – what has changed?

British Dental Journal (accepted for publication) Robert McAndrew and Rachel Greatrix

Chapter 1: Introduction

1.1 UKCAT

The United Kingdom Clinical Aptitude Test (UKCAT) Consortium was established in 2005 to create an admission test for medicine and dentistry with the aims of achieving greater fairness in selection to medicine and dentistry and to the widening participation in medical and dental training of under-represented social groups. Through an ongoing programme of research UKCAT intended to identify the characteristics in applicants which would make them good dentists and doctors and thus improve the quality of those who enter the professions, with the ultimate aim of improving patient care.

By 2020 over 350,000 candidates had taken the test and UKCAT had become an established part of the admissions landscape in the UK. In 2021 a further 37,000 candidates took the test. In 2020, the test was used by 30 universities as part of their admission requirements for medicine and dentistry. The use of the test by those universities in their selection processes has grown in strength over that period (Adam, Dowell et al. 2011, Greatrix and Dowell 2020).

The creation of the test was a bold move for the founding Consortium universities and was introduced not without controversy. Criticisms at the time were that the UKCAT was a test without validity, that it created an additional barrier for candidates applying to these programmes and as a result would act in opposition to efforts to widen access to the professions (Cassidy 2008).

At the same time the introduction of the UKCAT represented a step change in admissions to medicine and dentistry in the UK. Up to then most medical and dental schools relied on traditional methods of selection, combining school leaving qualifications with an assessment of personal statements and references to select candidates for interview. Interviews tended to be 'traditional' and predominantly semi-structured. At the time, a small number of institutions used the Biomedical Admissions Test (BMAT) within their selection processes and some graduate entry programmes used the Graduate Medical Schools Admission Test (GAMSAT).

1.2 Medical and Dental Selection in the UK

Selection to medicine and dentistry remains a challenging and contentious issue internationally. A review of the Ottawa consensus statement on selection and recruitment to the healthcare professions reported a growing evidence based approach to selection in the UK (Patterson, Roberts et al. 2018). However, it also highlighted the continuing complexity and challenge in medical student selection, noting the often conflicting drivers around diversity, differential attainment, retention and institutional aspirations.

The drivers around medical and dental selection have changed since 2005. National policy around Higher Education has focussed significantly on widening access (Schwartz 2004) and more recently light has been shined on the professions including medicine and dentistry (Milburn 2009). This shifted the discussion around

selection from a significant focus on efficacy/fairness of selection tools, placing the debate instead in discussions around social mobility and inequality. This has resulted in an increased focus on diversifying the medical student population in particular, with attempts across the sector to raise aspirations amongst applicants from under-represented groups. At the same time the use of contextualised admission processes and development of gateway and other access programmes has sought to reduce the barriers to applicants from these groups (Curtis and Smith 2020).

At the same time there has been a focus by the General Medical Council (GMC) and others on identifying the traits required to make good doctors. Of particular importance was the publication of the Role of the Doctor Consensus Statements (Council 2008, Council 2014) supported by the GMC publication regarding Good Medical Practice (General Medical Council 2013). National reports such as that arising from the review at Mid-Staffordshire (Francis 2013) led again to questions about the *values* being tested during recruitment to training and later into the professions.

Postgraduate recruitment trends in the professions have not stood still either, with significant changes particularly around recruitment into foundation year programmes. Foundation Year doctors are required to take the Prescribing Safety Assessment (PSA). The use of situational judgement tests (SJT), testing clinical procedural knowledge alongside non-cognitive attributes are now widely used within postgraduate medicine and dentistry (Webster, Paton et al. 2020).

1.3 Background to the Development of this Thesis

Prior to this research study being formalised, the UKCAT Board had expressed a view that formally documenting elements of the development of the organisation was desirable. Whilst scoping this exercise, I identified the advantages of undertaking this work within a research degree framework, providing additional direction and rigour. The original research plan focussed explicitly on creating this historical record. As the work progressed, I identified opportunities to explore the impact the test has had on selection and to evaluate the extent to which the original aims of the organisation had been achieved were identified. As a result the research questions developed to those outlined later in this chapter.

As an assessment, UKCAT generates volumes of quantitative data, analysis of which will contribute to this research. At the same time the subjective perspective of those using the test in selection will be explored qualitatively. This thesis is underpinned by a pragmatic philosophy supporting a mixed methods approach to the research. This is discussed in Chapter 2.

I was particularly focussed on opportunities when answering the research questions to identify recommendations for action by the organisation.

My senior and longstanding position in the organisation presents clear advantages to the undertaking of this work both in terms of their access to materials and understanding of context. However, this position creates the potential of bias in

research design, analysis and interpretation. This is explored further in Chapter 2 and where relevant in subsequent results chapters.

1.4 Research Questions

To investigate the impact that UKCAT has had on selection to medicine and dentistry in the UK, the following research questions will be addressed:

- How was the UKCAT Consortium established and how has it developed?
- What is the evidence base around the utility of the UKCAT itself?
- How has the UK admission landscape changed since 2006 and what have been the drivers around this?
- What did the UKCAT Consortium set out to achieve and to what extent has it been successful?

Impact is defined broadly within the thesis. The thesis includes reflections on the original aims of the organisation and the extent to which these have been achieved. Information is presented to record the development of the UKCAT Consortium (as an organisation) and the test itself. Impact is explored within this information in terms of key achievements and challenges since 2005. In addition, this information contributes to an understanding of how the scope and reach of the test has changed.

The thesis analyses the contribution the organisation has made to improving the evidence base around selection in the UK, in itself a potential impact on selection. At the same time the research evidence base around the use of the test itself is presented. The impact UKCAT has had on actual selection processes is presented both quantitatively (using data collected on an annual basis from Universities) and qualitatively (obtained through qualitative interviews with admission tutors).

1.5 Outline of Thesis

The research undertaken in this research is underpinned by a pragmatic paradigm. The thesis uses mixed methods research to answer the research questions outlined above. **Chapter 2** describes this paradigm and why it is relevant to this research. **Chapter 2** also provides an overview of why mixed methods research is appropriate and outlines the different approaches taken to answer each of the research questions. Given the potential conflicts of interest created by my role in UKCAT, this chapter also discusses the importance of reflexivity to minimise the potential of bias in research design and the presentation of results.

It is important to understand the context around selection to medicine and dentistry, particularly in the UK, prior to the creation of the UKCAT Consortium and the initial creation of the test. This is described in **Chapter 3**, the primary focus of which is on the literature around selection, also referencing national policy drivers contributing to discussions at the time. This evidence base contributed to the original aims of the UKCAT Consortium.

The original and developing aims and objectives of the organisation are also presented in **Chapter 4** providing a basis, later in the thesis, to evaluate the extent to which these aims have been achieved.

The first evidence regarding impact of the UKCAT on selection will be presented in **Chapter 4**, which provides a detailed timeline of the key milestones in the development of the organisation, the delivery of the test and the Consortium's research agenda. **Chapter 4** also summarises how test content has developed since 2006 and presents summary statistics of candidate performance since inception.

UKCAT set out to improve the evidence base around selection to medicine and dentistry in the UK. In **Chapter 5** the research evidence relating to how the test has developed, its use in selection and the candidate experience is detailed. Many of these research papers focus on the ability the UKCAT to predict performance in medical and dental school. This chapter also includes a systematic review of the predictive validity of the UKCAT.

Chapter 6 presents a picture of the changing landscape around selection to medicine and dentistry since 2006. In this chapter, key national drivers which have influenced change are identified. Running alongside this, data will be presented which shows how selection methods, and in particular the use of the UKCAT, have changed since 2006.

Admission Tutor views on the test, its impact on local and national selection and the future of the organisation will be obtained through a series of interviews. Outcomes from these interviews are reported in **Chapter 7.**

The discussion section, **Chapter 8**, will consider the extent to which the original aims of the organisation have been met and evaluate the impact UKCAT has on selection. Drawing on the literature and data analysis, implications for the future of the organisation and the test will be identified.

1.6 Summary

This chapter sets out the research questions to be addressed in this thesis and provides an outline of the thesis structure. How the thesis evolved is presented alongside information regarding my role in the UKCAT. It will be important for me to acknowledge this role when considering appropriate research methods and evaluating outcomes. The development of the UKCAT Consortium and UKCAT test runs alongside changing priorities in selection to medicine and dentistry with a particular focus on widening access and assessment of professional attributes. The extent to which these priorities have driven change will be further explored in the thesis.

Chapter 2: Research Design and Methodology

2.1 Introduction

This chapter provides an overview of mixed methods research and describes the pragmatic paradigm that underpins the research design and methodology for this thesis. The research questions addressed within the thesis are outlined and a description provided of the different research methodologies used to address each of these questions alongside a justification for those particular methods.

2.2 Mixed methods research

Researchers adopting a quantitative paradigm (primarily) use numerical data to generate objective results with the intention of generalising their results to a broader population. In this positivist view of the world, the researcher has no influence on their research outcomes although they are required to examine and interprets their results (Cleland 2015). Other researchers utilising qualitative approaches would argue that pure objectivity is not always achievable; instead, they seek to draw conclusions from observing available evidence. It is necessary to recognise and acknowledge the researcher's role in such research activity. When reviewing research outputs from quantitative and qualitative researchers, significant differences in research design, data collection and data analysis would be observed. Qualitative research is underpinned by different theoretical frameworks to quantitative research.

In the past quantitative and qualitative researchers have argued that their different, associated paradigms exclude one another. More recently researchers have attempted to reconcile these two positions, acknowledging that both quantitative and qualitative research are important (Johnson and Onwuegbuzie 2004), seeking to position mixed methods research within a continuum of methodologies. Whilst this might be regarded as a pragmatic approach, the use of different methods and their sequencing still requires care and justification (Lingard, Albert et al. 2008). Mixed methods research allows researchers to draw upon the strengths of each approach, measuring results and exploring relationships whilst at the same time exploring behaviours and attitudes in greater depth.

Leech and Onwuegbuzie define mixed methods research as studies that involve 'collecting, analysing and interpreting quantitative and qualitative data in a single study or in a series of studies that investigate the same underlying phenomenon' (Leech and Onwuegbuzie 2009).

Lingard describes mixed methods research as combining 'elements from both qualitative and quantitative paradigms to produce converging findings in the context of complex research questions' (Lingard, Albert et al. 2008). However, he went on to stress the importance of an explicit justification of any strategy to mix methods, including sequencing, prioritisation and (where appropriate) integration of methods. By adopting a clear strategy, the researcher can use, combine and

triangulate data to support results in a way which could not be achieved with a single method.

Johnson and Onwuegbuzie describe eight distinct steps when creating a mixed methods research process model and highlight three of these for further focus (Johnson and Onwuegbuzie 2004). The *purpose of mixed method research* (or rationales) had previously been described by Greene et al as triangulation (corroborating results using different research methods), complementarity (providing additional perspectives on results from one method by using another), initiation (identifying contradictions in the data which my lead to revisiting the research questions), development (where findings from one method to inform the other method) and expansion (expanding the breadth of research using different research methods) (Greene, Caracelli et al. 1989). *Data Analysis* in mixed methods research comprises several stages requiring the researcher to give careful thought to describing data, drawing together datasets created using different methodologies and ultimately integrating these data in order to present results. *Legitimation* (validity) is the process by which the researcher considers and assesses the trustworthiness of their results.

Cresswell also advises researchers to consider how timing, weighting, mixing and theorizing impact on planning mixed methods processes (Creswell 2013). Consideration of these allows researchers to identify a preferred strategy to address their research questions. Cresswell describes six alternative strategies which researchers may adopt such as sequential explanatory design (in which initial quantitative results informing secondary quantitative data collection) and (more relevant to this thesis) concurrent triangulation design (in which analysis of data collected concurrently helps identify where results converge/diverge).

Study design may also involve use of a cyclical approach with data interpretation indicating the need for further data collection and the potential revisiting of the research questions (Johnson and Onwuegbuzie 2004).

Within the realms of medical education, mixed methods are widely used to (Lavelle, Vuk et al. 2013, Cleland 2015) with researchers adopting a pragmatic approach to address their research questions. As outlined above, different research methods should be used strategically and significant given to timing, weighting and sequencing of difference research methods. This is discussed more in the following section.

2.3 Mixed Methods and the Pragmatic Paradigm

In exploring the use of pragmatism as a research paradigm, Morgan defined research paradigms as 'social worlds where research communities exert a powerful influence over the beliefs we consider to be meaningful and the actions we accept as appropriate' (Morgan 2014). Research paradigms therefore guide both the choice of research question and the methods to be used to explore it.

Quantitative research is usually underpinned by a positivist paradigm in which researchers seek to objectively describe phenomena which can be observed and measured. This purist approach has largely moved on to a post-positivist paradigm which whilst incorporating measurement and observation, accepts an element of fallibility in this approach, in particular recognising the potential of the researcher influencing outcomes. Qualitative research is often underpinned by more subjective paradigms such as constructivism where reality can only be interpreted (not discovered) and is influenced by individuals and the researcher. Traditionally these paradigms polarized researchers. Social science researchers became increasingly critical of this 'false dichotomy' (Onwuegbuzie and Leech 2005) promoting instead a pragmatic use of both quantitative and qualitative research methods allowing researchers to draw on the strengths of each.

Positioning itself between these two paradigms is the pragmatist approach, allowing the researcher to combine different research methods in order to best answer the research question. Within mixed methods research, this pragmatic approach feels attractive, allowing the researcher flexibility in identifying methodologies appropriate to answering the research question. Morgan (Morgan 2014) argued however that use of pragmatism as a research philosophy requires researchers to go beyond its use as merely an opportunistic solution to solving a methodological challenge. Morgan built on John Dewey's concept of research inquiry in which 'actions as outcomes of inquiry serve as the basis of beliefs'. The pragmatist approach therefore is a dynamic process of inquiry where the researcher reflects on the research questions, actions taken and beliefs underpinning both. The pragmatist will be aware throughout the research process of the choices being made (and those being rejected) in regard to the research question and the methods identified to answer those questions.

Kelly and Cordiero (Kelly and Cordeiro 2020) considered the utility of the pragmatist paradigm in relation to organisational research which is of relevance to this thesis. They identified three principles of relevance to such research:

- An emphasis on actionable knowledge should ensure that research agendas are relevant, focussed on problem solving and potentially draw upon multiple and diverse experiences.
- Recognising the interconnectedness between experience, knowing and action might direct the researcher to explore (and perhaps triangulate) organisational processes/impact through both observed actions and individual experiences.
- A view of inquiry as an experiential process promotes a more inclusive research process combining perspectives across the organisation.

Kelly and Cordiero went on to describe how these principles had been applied in two projects, drawing relevant conclusions for each stage of the research journey. Research objectives were grounded in organisational priorities from the very start

and provided scope for recognising and drawing on the experiences of different stakeholders. Critically, this allowed research questions to broaden beyond those that might be identified from a narrower review of the literature. Methodological choices were again directed by pragmaticism, requiring the researchers to explore both formal documentation and the experience of stakeholders. Sequencing of research methods and identification of opportunities to triangulate findings are highlighted as further pragmatic choices to be made at this stage. The pragmatic approach, promoting a holistic view of the organisation, opened up significant opportunities in terms of research methods. The researcher is required to reflect on these choices in terms of their own influence on them and the potential impact of the inquiry process on data collected from individuals. Pragmatist principles also guided the process of analysis and identification of research outcomes, focussing on outcomes that were relevant and likely to contribute to organisational development.

2.4 Use of Mixed Methods and the Pragmatic Paradigm in this Thesis

This thesis seeks to determine the impact UKCAT has had on the UK admissions landscape whilst at the same time investigating the extent to which UKCAT has met its original aims. An overview of the thesis structure is provided in Chapter 1. It is intended to present a variety of data to feed into structuring the interviews with admission tutors (Chapter 7) and ultimately the discussion around UKCAT's impact and the extent to which is has achieved its aims. The qualitative interviews are a key part of the thesis. Admission tutors play an important role both in contributing to UKCAT's strategic and operational goals, and in implementing and developing local selection processes. Their views therefore are of critical importance in determining the impact of the organisation on selection to medicine and dentistry in the UK.

Sequencing data analysis in this way is intended to provide different perspectives and where appropriate to triangulate findings. To achieve this would not be possible without combining both quantitative and qualitative approaches to the analysis of data. Evaluation of impact will draw on several data sources including quantitative and qualitative data collected from Consortium universities (primarily regarding the use of the test in selection) since the creation of UKCAT. These data will be analysed using both qualitative and quantitative methods. A review of the literature directly relating to the UKCAT (including a systematic review of evidence of predictive validity) alongside these analyses will inform the focus of interviews with admission tutors. The qualitative analysis of these interviews will be critical in addressing the research questions. Underpinning these analyses, a documentary review will allow both the creation of an organisational timeline and provide evidence against which to judge the extent to which UKCAT has met its original and emerging objectives.

In exploring the research questions in this thesis a positivist paradigm might on occasion be deemed appropriate when presenting for example data regarding

candidate and test performance and the investigating the predictive validity of the test. On the other hand constructivism might better guide the acquisition of knowledge in relation to understanding stakeholder views for example around the utility of the test. A pragmatist approach has instead been taken, recognising this permits utilising different research methodologies. At the same time the pragmatist focus on action and problem solving leading to tangible outcomes for the organisation fits the focus of this research. Consideration has been given as to the most appropriate methodologies to answer each research question, to the sequencing of the research and to how findings are considered and presented.

2.5 Reflexivity

Given my central role in UKCAT (as described in Chapter 1) reflexivity is of importance to the credibility of the results presented in this thesis.

Haynes describes reflexivity in simple terms as 'an awareness of the researcher's role in the practice of research and the way this is influenced by the object of the research, enabling the researcher to acknowledge the way in which he or she affects the research processes and outcomes.' (Haynes 2012). She goes on to suggest that conceptualising reflexivity might change depending on the nature of the research being undertaken. On the one hand reflexivity might be used to evaluate research methods and design; where the researcher is more involved in interpreting reality than reflexivity focusses more on understanding and acknowledging the 'values and preconceptions the researcher brings to that understanding'. This is a helpful distinction in terms of this thesis where the nature of reflexivity may well vary at different stages of the research.

Reflexivity on the part of the researcher conducting qualitative research generally and interviews in particular is important. Mann and Macleod note that reflexivity in this context is not intended to eliminate any subjectivity introduced by the researcher but supports researchers in being cognisant of this very subjectivity (Mann and MacLeod 2015).

Researchers need to understand any potential influence their own role in interviews may have on outcomes and their relationship with the interviewee. Rapport is important, as is the impact of any hierarchical relationship (DiCicco-Bloom and Crabtree 2006). Reflexivity is of equal importance in minimising the impact of bias which may impact on many aspects of interviewing including the identification of the sampling and analysis (Cohen and Crabtree 2008).

As previously commented, it will be important throughout this study to be aware of potential bias in results and outcomes as I and many contributors have close associations with the organisation. UKCAT has a history of openness around publication of findings and evidence around the test and it is intended that this spirit will guide my approach.

2.6 Research Questions and Methodologies

2.6.1 How was the UKCAT Consortium established and how has it developed?

A review of the literature, including national policy documents available will help put in context the admissions landscape prior to the creation of the Consortium and implementation of the test.

Further information will be obtained from UKCAT documents available from around the time of the formation of the organisation.

One purpose of this study is to document the development of an organisation which has been in existence for over ten years. Whilst this is in no way a long time in terms of organisational history, the initial years around the creation of the organisation and establishment of the test are of particular interest. The primary source of information to draw upon will be from UKCAT's own archives. These comprise hard copy and electronic records including agendas, supporting papers and minutes for Board, Consortium, subcommittees and working groups in existence during this period.

This activity closely resembles archival research and will involve locating, evaluating and interpreting available data sources. Reviewing these materials in this way will facilitate the evaluation of decisions and actions retrospectively in the light of new information. In particular, and of relevance to this study, it will lead to a creation of a timeline of activities undertaken by the organisation. Reflecting on decisions made in the light of new evidence will help evaluate the success or otherwise of achievements against the organisation's original aims.

It is intended that an initial review and synthesis of these data will start to create a timeline of organisational development which can then be complemented with external drivers and context. Of relevance to these external drivers will be an understanding of key developments in medical and dental education that may have influenced admissions processes, reports published by the regulators (General Medical Council and General Dental Council) and wider government reports impacting in this area. It is expected that much of this reported work would be reflected in discussions within the Medical and Dental Schools Councils and in particular in their annual meetings of admissions tutors. Outcomes from these meetings will further provide context regarding issues facing universities, actions being taken and drivers for change.

The establishment of a narrative timeline around the development of the organisation will help shape the thesis. It may be possible from this to identify key 'eras' of development. For example, it is likely that there will be significant focus on the founding of the organisation and initial delivery of the test. This may be followed by a period of consolidation. UKCAT has more recently embarked on additional activity around the development of the test and an increased research focus which may be regarded as a second period of development.

This timeline will provide a record for future researchers of the development of the test itself noting where content and timing changes have occurred and the reasons for those changes.

It will be important when undertaking this piece of work for me to be aware of my position in the organisation and avoid interpreting information from documents beyond that which is formally presented. This will be assisted by a continuous review of the narrative timeline to ensure that facts (and not opinions) are presented.

Running parallel to the development of UKCAT as an organisation, the development of the test itself will be documented. Critical to this will be drawing on information contained in UKCAT annual technical reports which includes detailed psychometric analysis of candidate and item performance. This will allow for year on year presentation of key metrics within the test, particularly around candidate subgroup performance.

The timeline referred to above will include an annual overview of test content. Further, more detailed information will be presented regarding the content of each subtest including examples of item types. A Descriptive Statistics section will include information extracted from the UKCAT annual Technical Reports to highlight some of the key trends relating to candidate numbers, performance and test analysis.

In terms of the descriptive statistics these will be being presented in this form for the first time. I will need to give careful thought to the interpretation of these data given their potential links to other aspects of the research.

Quantitative Study Elements to answer this question: Presentation of key test data over time

Qualitative Study Elements to answer this question: Literature review, document review

2.6.2 What is the evidence base around the utility of the UKCAT itself?

Since UKCAT was established there has been a growing evidence base around its use. Some of this research evidence has utilised data contained in UKCAT's own research database; other studies have been undertaken locally in institutions or collaboratively between them. It is intended to record and critique this evidence base and identify to what extent such work has informed the further development of the test and its use by universities.

Given its centrality to discussions around fairness and utility of the test, a systematic review of the predictive validity of the test will take place.

In terms of both the literature and systematic reviews it will be of equal importance to recognise and report on studies with negative outcomes for UKCAT including predictive validity studies with few or no positive relationships identified. It is

intended to submit the systematic review for publication in order that this work is peer reviewed.

Qualitative Study Elements to answer this question: Review of the literature

Quantitative Study Elements to answer this question: Systematic review

2.6.3 How has the UK admission landscape changed since 2006 and what have been the drivers around this?

I have collected information and data relating to each Consortium university's use of the test has been collected on an annual basis. Quantitative analyses of these data will allow for the presentation of trends and changes in test use over the years. How the test is used in selection has previously been reported (Adam, Dowell et al. 2011) and will provide a template for the interpretation and reporting of these data. It is intended to publish this element of the research in order that this work is peer reviewed.

Additional interviews, analysed qualitatively, with existing admission tutors will ask them to reflect on how they use the test and the drivers around change in selection over time. Consideration was given to alternatives to the interviews with admission tutors including a survey (perhaps requiring both quantitative and qualitative analysis). However it was felt that quantitative data collected through this route would potentially be very similar to the data I had already collected on an annual basis. Qualitative analysis of (semi-structured) interviews would allow greater exploration of the research questions and the emergence of novel themes. (Adam, Dowell et al. 2011)

Reflexivity in terms of the qualitative interviews will be discussed in more detail in Chapter 7.

Quantitative Study Elements to answer this question: Presentation of composite data (from annual interviews) over time

Qualitative Study Elements to answer this question: Admission tutor interviews

2.6.4 What did the UKCAT Consortium set out to achieve and to what extent has it been successful?

This research question will be addressed using a variety of sources. The original and developing aims of the organisation will be presented drawing on information gleaned from the organisation archive. The admissions literature will provide background information regarding some of these themes. Critical will be the views of admission tutors, in qualitative interviews (see above), on the extent to which UKCAT has met these aims and perhaps the extent to which the original aims might need to (or indeed have already) adapt and change.*Qualitative Study Elements to answer this question: Document review, Admission tutor interviews*

2.7 Methods of Data Collection and Analysis

2.7.1 Document Review

Archival research involves the study of a collection of documents to inform research questions. How historians use and regard such research varies depending on theoretical perspectives (Mills and Helms Mills 2018). Archival materials will be used in this thesis to gain an understanding of and to record the historical development of UKCAT. Understanding how the organisation has changed and the drivers behind this can be partially achieved both by analysing internal documents and reviewing relevant external policy documents.

Documents to be analysed in this study include all agendas, minutes and (where deemed relevant) supporting papers for the meetings of the organisation since formation. These documents are stored largely electronically though in some cases only hard copies of documents remain. Where there is evidence of missing documentation these will be sort from other stakeholders to ensure a comprehensive document set is utilised. Analysis of these documents will create a historical narrative of organisational development and will also provide themes for further investigation with interviewees allowing for the triangulation of data sources (Bowen 2009). Triangulation of conclusions obtained from organisational records is important as internally generated documents may not necessarily provide firm evidence of what they purport to report.

The documents being analysed for this study create by their nature a narrative over time. This narrative will create the organisational history. At the same time, it will be possible to distil the key challenges that have been and continue to be faced by the organisation. Repetition of identified 'problems' over time might for example require further investigation through additional data sources.

2.7.2 Interviews

Di-Cicco Bloom and Crabtree focus on unstructured and semi-structured interviews (DiCicco-Bloom and Crabtree 2006) in their analysis of the qualitative research interview. Semi structured interviews are used to 'explore meaning and perceptions to gain a better understanding and/or generate hypotheses'. Potential participants are selected based on their role and understanding of the area of research. It is anticipated that they will have shared experiences of the research topic. Data collected is likely to be the unique data source to address the research question to be answered. 'Purposeful sampling' of potential participants helps ensure data is drawn from those with a breadth of roles, knowledge and experience.

In semi-structured interviews, questions are predetermined. However, it is anticipated that alternative lines of enquiry may occur in response to interviewee answers; in this way it is anticipated that the interviews become an iterative process. The first question may address the area of research directly. Further questions will have been identified but questioning may adapt as it becomes clear which line of enquiry elicit relevant information and which do not. Deviation from

set questions is anticipated depending on the direction which each interviewee takes.

Once interviews are concluded, researchers then review interview content and identify emerging themes and patterns. Responses can be identified, coded and sorted to allow the researcher to identify significant emerging themes. Analysis of data takes place by categorising data in such a way to allow the research question to be answered. Common approaches identify a hierarchy of categories grouped around main themes. Data saturation is achieved when analysis no longer generates additional data for each main category.

The quality of research data is of importance however it is generated. In relation to data quality, quantitative researchers generally focus on issues of reliability and validity. In qualitative research, credibility, transferability, dependability and confirmability of data are key issues (Whittemore, Chase et al. 2001). Researchers need to have these issues at the forefront of their minds when analysing data to ensure that criticality remains at the centre of the research process.

Interpretation of findings through the identification of themes relies on subjective judgement. Credibility can be achieved by triangulating results and conclusions using quotes and, where appropriate, allowing participants to 'check' outcomes from their contribution (Tavakol and Sandars 2014).

Interviews in this study are semi-structured individual interviews with a range of existing admission tutors and are reported on in Chapter 7. Methodologies relating to participant sampling and thematic analysis are provided in greater detail in that chapter.

2.7.3 Thematic Analysis

Thematic analysis is 'a method for identifying, analysing and reporting patterns (themes) within data' (Braun and Clarke 2006). Braun and Clarke argue that rather than a tool or process for researchers, thematic analysis is a method in its own right.

By identifying a hierarchy of themes within data, commonalties and differences can be further explored. The active role of the researcher in this process needs to be acknowledged rather than assuming that themes naturally emerge from datasets.

Themes identified through analysis should be distinct (not overlap) and important in the dataset. In some research, analysis is bottom up (inductive) with the research question emerging from the coding of the data. Thematic analysis is used in Chapter 7 to analyse responses from semi-structured interviews and therefore themes are likely to be closely linked to the specific research questions. This is not to say however that alternative and unanticipated themes might not emerge from the dataset.

Topic coding is data driven and allows the researcher to interpret data and organise it in such a way that findings can be shared and replicated. Coding templates can be

refined as the research continues requiring the researcher to return to and repeat earlier analysis. Discussions between researchers as the work progresses ensure that the approach can be critically evaluated, and alternatives considered.

Extracts from the data can be used to exemplify themes to help the reader make sense of and interpret the analysis. Whilst extracts are not a substitute for analysis, their use to support the researcher's analysis ought to be sufficient for the reader to be convinced that the themes reported and discussed are credible.

Braun and Clarke propose a phased approach to thematic analysis which are followed in Chapter 7.

Phase	Description of the process
1. Familiarising yourself with your data:	Transcribing data (if necessary), reading and rereading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking in the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic "map" of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells; generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Table 1 Phases of Thematic Analysis (Braun and Clarke 2006)

Braun and Clark expand on the criteria for conducting good qualitative research in order to produce a checklist to determine the quality of analysis undertaken. Accuracy in transcription, rigour in coding, distinctiveness in themes and a demonstration that genuine analysis (rather than description) has taken place are all highlighted.

Interviews in this study are semi-structured individual interviews with a range of existing admission tutors and are reported in Chapter 7. The methodologies

relating to sampling and thematic analysis are provided in greater detail in that chapter.

2.7.4 Literature Review

This thesis contains two reviews of the literature. The first review in Chapter 3 takes a long view of the selection literature available prior to the creation of the UKCAT Consortium. This review intends to set the scene around the establishment of UKCAT and identify the drivers around its development.

A second review of more recent literature pertaining specifically to research with reference to the UKCAT is also undertaken in Chapter 5. The aim of this review is to evaluate the evidence base specific to the test itself.

2.7.5 Systematic Review

Systematic reviews seek to assess, synthesise and critique available information in a transparent way in order to report upon the best evidence available.

A systematic review of the literature relating to evidence around the predictive validity of the UKCAT is reported on in Chapter 5. The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (Liberati, Altman et al. 2009). Greater detail regarding the use of a research protocol to underpin this piece of work is provided in chapter 5.

2.8 Summary

To address the research questions, this thesis will utilise a pragmatic mixed methods approach. Reviews of the literature will help set the scene around medical and dental selection at the time UKCAT was created. Archival research will provide information regarding the development of the organisation and the test itself. Further reviews of the literature will shine a spotlight on the validity of the UKCAT test and the extent to which the organisation has met its aims of improving the evidence base in selection. Quantitative analysis of data regarding the use of the test by universities since 2006 will allow me to describe how selection to medicine and dentistry has changed. Qualitative analysis of interviews with admission tutors will provide evidence of the impact UKCAT has had on selection locally and drivers for change where this has occurred.

Chapter 2 provides an overview to the research methods utilised in the thesis, some of which are explained in more detail in subsequent chapters.

Chapter 3: Selection to Medicine and Dentistry in the UK prior to 2006

3.1 Introduction

As outlined in Chapter 2, this chapter reviews the literature available at the time to provide a backdrop to the early discussions of the organisation (set out in Chapter 4) and associated decision making around the test, its content, future development and use in selection.

The early literature mostly relates to selection in the US, particularly around the creation and (later) validation of both the Medical College Admissions Test (MCAT) and Dental Admission Test (DAT) which became established parts of selection in the US from the late 1940s. The chapter goes on to look at the UK context prior to 2006 (when UKCAT was established) to identify and evaluate the evidence emerging in the literature around selection.

There are themes in the literature linked to national drivers around social mobility and desirable attitudes and behaviour which start to emerge more strongly later in the 20th century (e.g. diversity, assessing non-cognitive traits) which are focussed on later in the chapter. Many of these themes were drivers for change in selection at the time the test was created and continue to inform how selection is changing in the UK and internationally.

In the final section several national policy documents are highlighted which impacted on thinking around selection before UKCAT came into existence.

3.2 Literature Review - Medicine

There has been a recent, significant focus on improving the evidence base around selection to medicine and dentistry in the UK (Patterson, Knight et al. 2016). However, the challenge of recruiting to medicine has long provided a rich source for investigation by scholars. In the early twentieth century, in an article focussing largely on subject requirements and level of prior education, Davis observed that 'During the last decade no question has excited more discussion among medical teachers than what should be the requirements for admission to college' (Davis Jr 1903).

3.2.1 The United States: creating the Medical College Admission Test (MCAT)

Undoubtedly the United States has led the way in researching selection criteria and in the development of admission testing.

As early as 1918, a test of reading ability had been identified as showing the potential for enhancing admission processes (Haggerty 1918). Over the next decade or so the use of additional selection tests was explored further with the Scholastic Admissions Test trialled at a number of US medical schools (Moss 1930),

showing evidence of predicting performance in medical school. This paper recorded the deliberations of the Association of Medical Education Colleges in response to this study, concluding with members passing a motion to appoint a committee to direct the development of an aptitude test to be used specifically in selection to medical schools in the United States.

A decade later Shoemaker et al noted that the challenges around selection to medical education had shifted from a lack of demand (early in the century) to a position where there was a need for genuine and effective *selection* (Shoemaker and Rohrer 1948). Shoemaker discriminated in his discussion between the need to select those who would be successful in medical education and those who would ultimately be successful in the profession. He noted 'the fact is that it is much easier to predict grades of medical students than it is to predict a more intangible group of factors which makes for success in the practice of medicine'. Shoemaker went on to report on the outcomes of a study investigating the predictive validity of different factors on performance in the first year of medical school. He noted the success of the then trialled Professional Aptitude Test in predicting performance but also noted interesting results associated with an administered Rorschach 'Ink Blot' test. The numbers included in the study were small, but results pointed to the need to look at these personality measures rather more carefully.

By 1950 the Medical College Admission Test (MCAT) had become an established part of selection to US medical schools (Stalnaker 1950). The test at that time included separate verbal, quantitative and science tests, and a further examination of the understanding of modern society. Stalnaker noted that work was under way to investigate the possibility of measuring emotional maturity and motivation, opening up further a discussion regarding the testing of non-cognitive traits. The MCAT appeared to be providing schools with a reliable measure of educational aptitude (Hill 1959) with Hill also suggesting that schools ought to be sceptical regarding the role of college recommendations in admission processes. Parallels can be drawn here with the much later debate in the UK regarding the use of school references.

3.2.2 Validating MCAT

The validation of the MCAT as a selection tool became a significant focus of researchers over the succeeding years.

A study looking at the relationship between MCAT performance and students struggling on medical programmes (Little, Gee et al. 1960) demonstrated the difficulties in conducting longitudinal, multi-site studies in medical education, where students do not proceed regularly and where significant differences exist between institutions. MCAT appeared to help predict the types of students likely to struggle on programmes. However, the authors struggled with the fact that this factor was

inextricably linked with the nature of the medical schools themselves and the different cohorts admitted.

Positive correlations between MCAT and academic ranking in medical school were later identified (Schwartzman, Hunter et al. 1962, Conger and Fitz 1963) although this effect declined in the clinical years. Conger and Fitz concluded that 'in order to do relatively well, as opposed to simply avoiding failure, the student…needs to be younger, have higher undergraduate grades and show more of the kinds of personal qualities deemed important by admission committee interviewers.' The predictive ability of academic scores was lower in the clinical years whilst personal qualities appeared to be more critical at this stage. The study used a rather obscure formula to 'predict success' at medical school. Given the size of the sample here (700+) it is rather disappointing that the study did not look more closely at some of these individual variables; analysis which had the potential to result in more conclusive outcomes.

Roemer's single centre longitudinal study of predictors of medical school success demonstrated low to medium correlations between MCAT subtests scores and medical school assessment outcomes (Roemer 1965). Correlations were strongest in the science component of the MCAT, highest in the first year of study and declined in subsequent years.

A small study (n=82) concluded that social class did not influence medical school exam performance or MCAT scores (Fredericks and Mundy 1967). The study also concluded however, that MCAT did not predict performance in medical school. However, the restriction of range (on admitted candidates who had been successful in year 1 examinations) on such small numbers makes these findings questionable.

A significant cohort study (Julian 2005) followed 4000 students admitted in 1992 and 1993 to 14 US medical schools from admission through to residency. University grades were best predicted by a combination of MCAT scores and undergraduate course grades; MCAT scores providing incremental prediction over such grades and acting as better predictors of licensure examinations than college performance. The probability of experiencing academic difficulty or achieving distinction tended to vary with MCAT scores.

3.2.3 United Kingdom

Relevant literature is sparse in the UK until the 1980s at which point McManus published a series of papers challenging many of the accepted practices around selection to medicine in the UK.

Until the mid-70s fewer than 50% of entrants to medical schools achieved ABB at A-level (McManus 1982). In this paper, McManus commented on the recent significant rise in medical school A-level requirements, noting that whilst some of

the rise could be attributed to increased demand for places, the phenomena suggested a potential shift away from reliance on non-academic selection criteria. He expressed concern about this trend, noting the lack of evidence linking high academic achievement with clinical performance.

An audit of admissions to St Mary's Hospital Medical School in 1984 resulted in several publications (McManus and Richards 1984, McManus and Richards 1984, McManus and Richards 1984). The authors concluded that A-level achievement was the major predictor of applicants accepting an offer, with O-level achievement, early application and having a medical parent having smaller effects. Non-academic factors were the major determinate of receiving an offer following interview. The third paper reported on candidate perceptions of the process. Candidates expressed concerns regarding over-reliance on academic achievement, potential bias within the interview process and the impact of the then UCCA requirement for candidates to rank university choices. The authors' recommendations for change included the removal of university rankings within applications and a move to an early date for submission of UCAS forms for medicine (in line with the requirement for Oxbridge entries of 15th October). The authors noted the continued difficulties in discriminating between highly achieving academic candidates but warned against putting too much weight on O-level achievement.

The availability of objective data from these audits was welcomed by Horton who went on to identify actions needed to improve selection processes (Horton 1986). He concluded that advice to prospective candidates needed to improve, rank ordering of institutions should be removed, and interviews should be regarded as the central opportunity to explore non-academic qualities. Selection processes should be further investigated and monitored in order that schools could assure themselves that factors such as gender, race and social class were not impacting on decisions. Horton commented further on the need to consider 'education achievement in the context of educational opportunity'. This final call for action remains a challenge for schools even now at a time when there is a much greater understanding of educational context and significantly more reliable contextual data available.

A follow up study in 1988 reported on the performance in preclinical courses of students against a range of background variables available at selection (McManus and Richards 1986). A-level results had some correlation with pre-clinical outcomes, but these were not significant enough or sensitive enough to justify further emphasis on such grades. Personality and motivation appeared to be the most important factors.

3.2.4 United Kingdom - Broadening the Debate

In a useful summary of selection processes utilised by medical schools (alongside those used by the Regular Commission Board, Civil Service Selection Board and British Airways), Roberts et al reported that medical selectors were relying exclusively on the content of the then UCCA form; using personal statements, references and academic record (including predicted grades) to select for interview (Roberts and Porter 1989). When evaluating these different processes, Roberts critiqued the central role of references in medical selection (given the lack of validity regarding their use), the focus on science examinations (which whilst predicting performance in pre-clinical examinations moderately well, were poor predictors of clinical year performance) and the failure of interviews to provide an accurate assessment of candidate potential. Experience in other sectors led the authors to comment on the possible usefulness of personality and psychometric tests, noting that medical schools seemed largely unaware of their potential utility. Their conclusions were quite damning, suggesting that recruiters had failed to learn from experience in other sectors; and that a lack of central direction led to inconsistency and the use of methods with little or no evidence base.

The research literature then started to focus more specifically on some of these challenges. If as Goldbeck-Wood argued, academic achievement at 18 was a blunt selection instrument, it was also the case that the suitability of alternative criteria was not evidenced (Goldbeck-Wood 1996). It was possible to identify desirable qualities in medical professionals, but it was unclear how you might select on this basis. Whilst reaching no firm conclusion, the author suggested that a lottery of those meeting minimum academic criteria might be 'fairest'.

A study at a UK Medical School in 2000 (Ferguson, Sanders et al. 2000) looked at the predictive validity of various measures available to selectors including a detailed analysis of personal statements. The coding of personal statements provided a fascinating insight into the content of UCAS forms completed by applicants. The researchers concluded that the content of personal statements contributed towards selection for and success at interview. However, personal statements as categorised within the report did not predict performance in medical training (year 1) whereas previous academic ability did. The paper also looked at the role of personality variables in selection, with candidates self-reporting using the Goldberg (1995) 35 bipolar adjective markers, resulting in measures of emotional stability, surgency, intellect, agreeableness and conscientiousness (the Big Five personality traits, used widely in psychology to describe human personality) (Goldberg 1993). These personality variabilities accounted for some variance in performance in training although only conscientiousness contributed independently.

In 2001, a paper from the University of Nottingham looking at academic and non-academic predictors of success (James and Chilvers 2001) supported the use of GCSE results in selection and the addition of Biology as a compulsory subject (in addition to Chemistry) at A-level.

The establishment of the Peninsula Medical School in 2002 created an opportunity for this new school to review existing selection processes and identify best practice. A report on the journey (Searle and McHarg 2003) to determine a new selection process identified the principles which underpinned the review. The approach endorsed by Peninsula complemented academic attainment requirements with an innovative structured interview based around an ethical scenario serving as a prompt for questions. Interviewer training and benchmarking were introduced to ensure high intra-assessor reliability. Interestingly, personal statements and references were explicitly not included at any stage within selection. The school trialled an instrument to measure applicant ability to work in a team.

In order to move this discussion on, a systematic review of the literature around selection was commissioned by the Committee of Deans and Heads of Medical Schools (Ferguson, James et al. 2002), seeking to review the data on predictive validity of admission criteria, the use of non-cognitive factors and the impact of demographic factors. The review concluded that previous academic performance though not perfect was a good predictor of achievement in medical training, accounting for 23% of the variance in performance in undergraduate programmes. The authors commented on the need for long term cohort studies to examine these issues further and noted the lack of research into many aspects of admission processes for medicine. Results cast doubt on the use of personal statements and references.

A further paper looked at the development of the new medical schools between 1998 and 2014 (Hull York, Brighton and Sussex, UEA and Peninsula) (Howe, Campion et al. 2004). The report commented that these new schools had to an extent shifted from over reliance on academic achievement. However the selection processes put in place at these new schools remained fairly similar to those in operation at existing medical schools. The new schools perhaps led the way in the greater structuring of interview processes and had a greater focus on widening participation. The authors (from each of the relevant medical schools) concluded that starting with a 'clean slate' allowed schools to ensure that processes had an evidence base, informed by prevailing policies.

3.2.5 Focussing the Debate

It would appear then that towards the late twentieth century there was a growing acceptance in the UK that established approaches to selection were far from

optimal and a growing desire to explore further the selection challenges being posed by researchers.

MCAT had become an established and accepted part of medical selection in the US. The focus there on academic achievement however was starting to be questioned. In a special issue of Academic Medicine, focussed on medical selection, in 1990 the strength of the link between academic aptitude and academic achievement was questioned (McGaghie 1990). Correlations between MCAT scores/college GPA (grade point average) and medical school achievement tended to be around 0.3 and 0.4 with between 84% to 91% of the variation in medical school achievement remaining unaccounted for. McGaghie noted that whilst 'Grades predict grades, test scores predict test scores, ratings predict ratings', efforts to measure predictors of professional competence still remained elusive.

Similarly in Australia, researchers were questioning the reliance on traditional selection methods. A case report from the University of Adelaide (Marley and Carman 1999) opened a debate on selection processes, drawing upon a survey of stakeholders and interrogating matriculation scores alongside university rankings. The authors concluded that selecting candidates on academic scores alone was no longer desirable. Reliance on matriculation scores resulted in unsuitable candidates being admitted and a lack of equality of opportunity for candidates from different backgrounds. The review resulted in the inclusion of assessments of critical thinking and logical reasoning in selection processes.

Given the similar nature of medical training in Australia (a mixture of UG and PG programmes) it is not surprising that the challenges in selection are fairly similar to those in the UK. Story and Mercer describe a selection landscape with significant competition for places, the need for a greater focus on non-cognitive attributes and greater diversity amongst those recruited (Story and Mercer 2005). The authors also comment on perceived unfairness of selection processes and the potential impact of coaching effects on performance in various aspects of those processes.

Having opened up this debate the selection literature started to focus on specific challenges in selection. The next few sections focus on some of these areas of research.

Medical School Attrition and Retention

In an attempt to relate reasons for attrition back to selection (Funkenstein 1962), Funkenstein commented on the importance for selectors in having the skills 'in selecting the students most apt to succeed' in a particular school, and 'in selecting those students who will graduate'. He also commented on problems with performance at medical school affected by psycho-social issues encountered by

students. Further evaluation of factors influencing students in serious academic difficulties was required to investigate this challenge further.

Factors affecting UK medical school retention were explored in a study investigating first year drop out between 1980 and 1992 (Arulampalam, Naylor et al. 2004). The likelihood of dropping out of medical school was influenced by both choice of subjects and grades achieved at A-level. The authors commented that increasing medical student intakes and widening access might affect retention rates unless further student support measures were put in place. In a further study (Arulampalam, Naylor et al. 2007) looking a later cohorts the authors commented on changes in entry requirements and curricula over this period. Whilst dropout rates had increased between the two studies (from 3.5% to 4.9%), this could not be explained by the characteristics of applicants but appeared to be due to changes within institutions.

Identifying factors associated with struggling students was picked up as a theme by Yates et al in a study undertaken at the University of Nottingham (Yates and James 2006). The authors focussed on 5-year groups and used several metrics to identify struggling students (12.8%) and those that left the course (4%). The authors concluded that those with lower A-level grades were more likely to struggle. Other risk factors included obtaining a later offer in selection, being male, non-white ethnicity and negative comments having been included in school/college references.

Diversity in Medical Training

Demand for places on medical programmes increased in the US during the 1960s (Funkenstein 1970). Funkenstien commented that medical selection was not responding quickly enough to changes in medical training, medical applicants themselves and external demands. His paper looked in detail at the experience at Harvard Medical School over this decade. Applicant numbers increased, as did the number of those that were highly qualified. Students were becoming more diverse in terms of ethnic background due to the pro-active efforts of faculty. The paper went on to recommend the creation of complex systems around selection to create a representative class of students (focussing on minority groups and preferred speciality). Funkenstien concluded that 'The only real solution to the problem of admissions is the expansion of medical schools until every qualified applicant in the country has the opportunity to enter medical school'. Even allowing for the significant differences in higher education and medical education in the US and the UK, this recommendation seems unconventional at best and does not appear to offer the solution promised in the paper's title.

Gender in medical education appears to be of little interest to researchers until a study in 1973 looked at the academic performance of women students in medical schools (Weinberg and Rooney 1973). The authors noted an increase in female students admitted to medical programmes in the US during the 1960s with numbers tripling and the proportion doubling to 13.5% of the admitted student population. The report concluded that women did not perform as well as men during medical programmes although performance improved in later years. The authors show a distinct lack of interest in commenting on why this may have been. This study does however highlight gender performance differences in the MCAT with males outperforming females in quantitative ability and science, and women excelling in verbal ability.

A UK study in 1995 reported that ethnic minority applicants remained disadvantaged in selection to medicine (McManus, Richards et al. 1995) although disadvantage had diminished. To address this issue, the authors recommended anonymising application forms and a move to a post-qualification application system. Predicted grades ought not to be made available to selectors at the point they were measuring other traits; a proposal further endorsed in a paper looking at the relationship between cognitive traits and interviewer ratings of non-cognitive traits (Shaw, Martz et al. 1995).

A further look at the issue of admission of ethnic minorities to individual medical schools concluded that selection bias for more able candidates did not exist, but that white candidates with lower A-level scores did appear to have greater chance of being accepted to medical schools than similar candidates from ethnic minorities (Esmail, Nelson et al. 1995).

In a study commissioned by the Council of Heads of Medical Schools, McManus revisited the predictors of success in medical application, again picking up issues relating to ethnicity (McManus, Esmail et al. 1998). This comprehensive study drew on data relating to 18,943 applicants to medical school in 1996 and 1997. McManus reported several areas in which candidates were disadvantaged - in particular, applicants from ethnic minority groups, male applicants and applicants applying late in the selection cycle. The paper proved controversial, not least in relation to tables showing the relative performance in each medical school regarding admission of ethnic minority applicants. Those commenting on the paper noted that McManus was looking at the data post qualification, whilst universities acted upon a different (less rich) dataset in making their decisions. In an editorial in the same edition of the BMJ, Abbasi commented that the paper 'adds to the mounting evidence of disadvantage at all stages in the careers of ethnic minority doctors' (Abbasi 1998). She concluded that 'best practice may require a structured combination of psychometric tests and interview', calling on medical schools to

scrutinise processes and draw on evidence from within and outside medical selection.

In 2004, Powis et al explored the potential of Graduate Entry programmes to address widening access and issues regarding student diversity (Powis, Hamilton et al. 2004). Australia had led the way in developing graduate entry programmes and the authors noted the creation of around 14 programmes in the UK. The use of the Graduate Australian Medical School Admission Test (GAMSAT) had allowed Australian universities to recruit graduates from all academic backgrounds, with the test providing an assessment of relevant scientific knowledge and other skills. The authors concluded that broadening the applicant pool beyond the school leaving population might assist with diversifying the student population. They also recommend broadening requirements to include tests of non-cognitive traits and active engagement with under-represented communities. The authors cite communication skills, interpersonal skills, ethical sensitivity, personal integrity, flexibility, tolerance, conscientiousness, teamwork and management skills as traits that might be assessed within admission processes.

There is very little in the literature around widening access to medicine in the UK prior to 2006 (possibly because the recommendations arising from the Schwartz Report regarding fair access were yet to be fully considered). Seyan et al (Seyan, Greenhalgh et al. 2004) used UCAS data to look at the social class of students at medical school. They concluded that there were significant inequalities between medical schools and proposed a consistent measure be used by schools to measure changes in the applicant and student population over time.

Interviews

A review of the available evidence around the use of interviews in medical selection commented (Edwards, Johnson et al. 1990) on the shortage of studies in this area and the fact that many were more descriptive than analytic. Drawing on a review of psychology research in this area, the authors commented on the need for medical schools to determine interrater reliability for interviews and that the construct validity of interview questions should be explored. The authors went on to set out methods for adding structure to interviews with the potential that these might lead to improvements in validity and reliability.

In a complex study, Harasym et al investigated the reliability and validity of interview judgements (Harasym, Woloschuk et al. 1996). Interviewers were recruited to interview scripted, simulated candidates. The number of simulated candidates was low (6) and one would speculate as to how you can accurately script a poor interview rather than an average interview. The study did show significant variance among ratings and only moderate validity in ratings of the 'true' level of

performance. The authors concluded that greater training and consistency in questioning was needed to overcome some of this variance.

Interviews appear to have remained largely unstructured (Edwards, Elam et al. 2001) although a central component of admission processes. There was no consistent approach to the examination of desirable non-cognitive traits and as above, the authors commented that schools failed to draw upon evidence from selection to other fields.

In a discussion paper regarding UK admissions (Hughes 2002), the author cited evidence from Newcastle, New South Wales (Powis, Neame et al. 1988) and McMaster University (Hamilton 1972) where investment in the development of in-depth structured interviews appeared to reduce drop out from programmes. Hughes went on to comment on the potential of carefully structured interviews to predict clinical performance (Powis, Waring et al. 1992) and the possible use of group interviews. She concluded that structured interviews may provide the best possibility of reliable outcomes. She also noted the limitations on schools in adequately resourcing interviews concluding that the potential of pooling interviews between institutions ought to be considered.

Researchers from the University of Iowa (Kreiter, Yin et al. 2004) reviewed evidence from nine relevant studies published since 1990 in an attempt to evaluate the reliability of admissions processes. The studies produced highly variable results, though the authors noted this was partially due to the application of different measures of reliability. Despite the interview being the costliest and most time-consuming aspect of the application process, reliability appeared to be low to moderate. The authors proposed that information derived from interviews be used in a more psychometrically defensible fashion.

The use of assessment centres as alternatives to standard interviewing was deemed more reliable than traditional interview methods (Oosterveld and ten Cate 2004) although the authors noted the significant burden on both schools and candidates in this selection method.

McMaster University was the early developer of the multiple mini interview (MMI) in selection. MMIs are defined (Eva, Rosenfeld et al. 2004) as 'an OSCE-style exercise consisting of multiple, focused encounters……intended to assess many of the cognitive and non-cognitive skills that are currently assessed (inadequately) by the personal interview.' The use of multiple stations/scenarios were shown to reduce the effects of chance and interviewer bias. MMIs showed good reliability and were well accepted by interviewers and candidates. Whilst further analysis looking at predictive validity was required, the authors clearly saw significant potential in the use of MMIs to improve selection in the health professions.

Assessing Non-Cognitive Criteria

McGaghie asserted a consensus regarding the importance of ten personal qualities associated with excellence in medical education and practice (McGaghie 1990): character and integrity; breadth of knowledge; leadership; geographic preferences; gender, race and religious preferences; work habits and motivation to study; personality and attitude; orientation towards service; altruism; personal effectiveness. Attention needed to be paid to improving the use of qualitative variables in selection which might be achieved through significant improvements to the research evidence base.

A summary of multiple studies undertaken during the 1980s reported on predictors of basic science performance, clinical performance and performance in postgraduate assessments (Mitchell 1990). Mitchell noted differences in outcomes and methods in the many studies reviewed. However, a clear message was that traditional academic measures predicted a substantial amount of performance in medical school albeit far stronger in the early years. She advocated however, complementing such factors with non-academic measures to give an overall picture of the candidate, noting the additional importance of such measures when considering educationally disadvantaged candidates.

Hojat et al investigated the incremental effects of psychosocial measures in predicting student performance in medical school (Hojat, Robeson et al. 1993). Psychosocial tests were used (although not in a high-stakes environment) to measure traits including anxiety, neuroticism, self-esteem and extroversion. Regression analyses were used to investigate how predictive standard admission measures, psychosocial factors and a combination of both might be in relation to basic science, clinical sciences and clinical competence ratings. The results gave overwhelming support to the use of a combination of these factors at admission noting the low predictivity of standard admission measures with regard to clinical competence.

A paper published in 2000 (Carrothers, Gregory Jr et al. 2000) explored the potential of measuring emotional intelligence (EI) in medical selection. EI is defined as the ability to monitor your own emotions and those of others and to use this ability to guide and inform behaviour. A consortium of universities in Ohio had been charged with giving greater attention to personal qualities in admission. To achieve this, a group of universities used an EI instrument utilising assessor ratings. The instrument appeared to be able to identify candidates with relevant EI qualities (maturity, compassion, morality, sociability, calm disposition). The authors noted the need to investigate further whether EI changed over time and in response to medical training. Disagreements regarding the efficacy of EI are noted in the paper as is the potential to self-fake responses and the lack of predictive validity in

studies. However, it is of interest that medical schools were starting to explore options available to measure non-cognitive traits in a more systematic evidence-based way.

Edwards et al (Edwards, Elam et al. 2001) proposed an admission model for medical recruitment noting a shift towards greater use of qualitative criteria in selection processes. This shift was driven by several factors, but the authors noted in particular the development of a consensus around attributes required by medical students (led by the Association of American Medical Colleges (AAMC)). This shift mirrored a shift in selection to the profession within the USA with licensing examinations assessing professionalism, communication and interpersonal skills. The author predicted a time when medical schools would be required to similarly test such skills at entry to medical training. The theoretical admission model outlined a framework for reflecting on the different elements of selection (the role, the applicant pool, admission criteria, assessors, processes and outcomes). This model called for schools to reflect on existing processes and use this model to develop selection policies underpinned by a better evidence base.

In the UK too, medical educators were beginning to focus on how non-academic traits could be tested within selection. Hughes (Hughes 2002) noted the potential danger of using non-cognitive traits that might narrow the field and commented on the need for reassurances regarding the reliability of such methods – especially over a lifetime of practice. She highlighted potential criteria which could be tested including honesty, integrity, conscientiousness and empathy and speculated whether evaluating resilience at admission to the profession might be desirable. In considering criteria that might predict 'failure' at medical school she noted the difficulties in analysing this accurately. Hughes was however sceptical about the use of psychometric tests citing concerns regarding candidates faking behaviour in responding.

A review of the literature by Albanese et al identified up to 87 personal qualities relevant to the medical profession (Albanese, Snow et al. 2003) and the authors commented on the difficulties in identifying the most relevant of these in selection. Measuring such traits presented real difficulties and the authors discussed the extent to which such traits were 'teachable' and 'fakeable' and if so, what implications this had for selection. The authors recommended greater integration between selection requirements for medical school and the profession, suggesting it was desirable to create a continuum of required competencies. Personal qualities could be assessed through a potential combination of survey/questionnaire and interviews.

Papadakis et al looked at predictors associated with physicians subject to disciplinary action between 1990 and 2003 (Papadakis, Teherani et al. 2005)

concluding that prior unprofessional behaviour in medical school was a strong predictor. Less strong, but still significant, was performance in MCAT and academic performance in the early years of medical school.

The scrutiny of selection methods led Lumsden et al to investigate whether traits measured within the Personal Qualities Assessment (PQA) and other demographic variables impacted on selection (Lumsden, Bore et al. 2005). The PQA, developed by researchers in Australia, consisted of three separate tests designed to measure individual differences in cognitive reasoning ability, empathy vs narcissism, and ethical/moral orientation. The test had shown reasonable levels of internal reliability and correlated with other recognised tests of personality. The personal traits measured by the test appeared not to be influenced by educational background though minor differences were observed in relation to a derived measure of deprivation. The authors intended to follow up this study with predictive validity studies, monitoring student performance in medical training. They noted that such studies might lead to the incorporation of the PQA or similar tests into selection processes.

The Hogan Development Survey (HDS) (an assessment tool used to identify various personality scales for use by management in building successful teams) was found to be able to identify negative personality characteristics in medical students not usually detected within selection processes (Knights and Kennedy 2006). A follow up study reported relationships between HDS outcomes and academic performance (Knights and Kennedy 2007). The authors concluded that despite the HDS being a self-reported measure it might be a useful addition to existing admission criteria. They did not however reflect on how the tool might work in a high-stakes admissions process.

Aptitude Testing

The growing number of aptitude tests used in selection is mentioned in an article in the BMJ in 2005 (McManus, Powis et al. 2005) with the authors referring to the Biomedical Admissions Test (BMAT), GAMSAT and the then developing Thinking Skills Assessment (TSA). Drivers to use such tests included government support for alternatives to A-levels, lobby from widening participation groups and the need for universities to discriminate between large numbers of very able candidates. The authors noted existing evidence for the predictive validity of A-levels. They observed that aptitude tests did and could measure many different things but that intellectual aptitude tests ought to be designed to measure 'intellectual capabilities for thinking and reasoning, particularly logical and analytical reasoning abilities' (Schwartz 2004). The authors reported that there was little existing evidence to demonstrate that such tests had predictive validity. The authors concluded that refining A-levels (creation of A* and A** grades) was preferable to the use of tests

without an evidence base. They did suggest however that medical schools might wish *to commission their own test*, bearing in mind the need to research its use and provide evidence of its reliability.

As a member of a national working party convened by the Medical Schools Council to look at the potential of aptitude testing, Nicholson (who would go on to be Chair of UKCAT) provided a case for admissions testing (Nicholson 2005). She reported the ongoing challenge of selecting fairly from large numbers of very able applicants and the cost in time and resources of interviewing significant numbers. The solution was to reduce the number of candidates at initial screening in a fair and reliable manner. Many groups were under-represented in the workforce and additional tests could provide a tool to widen participation. Nicholson went on to note the potential use of the personal qualities assessment (PQA) which assessed verbal, numerical and spatial reasoning, contained a personality inventory, and included an ethical reasoning paper. This article clearly signalled the direction in which the group was heading in creating an admissions test for medicine.

In 2006 Parry et al (Parry, Mathers et al. 2006) provided an overview of UK medical selection methods, using information obtained from each medical school to describe the entire selection process. Whilst there was much commonality in criteria used by medical schools, the authors noted significant differences in how these were applied. Two schools were using BMAT, and one school had trialled use of the personal qualities assessment (PQA). The authors commented on the potential use of aptitude testing though noting that such a test might not discriminate further between candidates. The authors concluded that it was desirable to move towards a centralised admissions system for medicine, noting however that the lack of a robust evidence base would guide against using established mechanisms within any common approach.

3.3 Literature Review - Dentistry

Published research regarding selection for dental education is sparser than that related to medical education but again starts to emanate from the United States early in the twentieth century.

3.3.1 The United States – creating the Dental Admissions Test (DAT)

A survey of dental curricula in the US (Seccombe 1932) commented on the fact that prior to 1884, there had been no academic requirements for admission to dental programmes. A preliminary examination including 'a good English education' was introduced as a requirement in 1884 with applicants able to avoid taking this test if they were able to present evidence of other academic achievement. In 1887 a minimum entrance requirement was introduced and over the next 30 years this became more rigorous; by 1917, applicants were required to have graduated from a four-year high school programme.

The impact on the dental profession of the increased educational requirements of medical schools was significant (Rypins 1931). By 1927, leading dental schools in the US were requiring two years of college study prior to admission.

A paper in 1937 focussed on the potential of a test of 'mechanical aptitude' (Harris 1937). The authors concluded that whilst prior education and IQ correlated significantly with year 1 and year 4 performance in dental school, the test of mechanical aptitude did not. It is unclear however, the extent to which these in-course assessments measured similar skills to the test of mechanical aptitude. If in-course assessments included significant tests of knowledge, then low associations might be expected.

Whilst considering trends in dental education in the US, Dixon (Dixon 1940) reflected on the journey from a position in the nineteenth century where the requirements of a dentist were thought to be 'digital skill, native ability as a mechanic, and tact with dealing with people (and not much stress was laid on tact).' Most of the vision for higher standards in entry qualifications for dental practitioners had been achieved. During the 1930s, all US Dental Schools had raised their requirements to two college years and placed further emphasis on science background. He commented however on the need for practitioners to be able to understand the social and economic drivers influencing health, concluding that dentists required an ability to 'think clearly on all matters that bear a social relationship with the endeavours of his profession and society' alongside a good scientific background.

A paper published after WW2 reflected on the qualifications which might be required by ex-service men and women given the interruption to their education during the war years (McGrath 1945). The author drew on data collected from graduating classes from the University of Buffalo Dental School between 1932 and 1940 setting out to identify relationships between qualifications attained prior to entry and performance on the course. The most significant finding in the study was that regardless of course type (science and non-science), good college performance predicted good performance on the dental programme. The author went on to comment on the lack of correlation between prior education performance and elements of actual dental practice. The author commented that schools exploring dental aptitude tests (tests of manual dexterity at the time rather than tests of academic aptitude) claimed such tests improved selection methods.

In a wide ranging paper Peterson provided an overview of selection methods in dental schools post WW2 (Peterson 1946). Whilst many schools were using qualifications, letters of recommendation and interviews, others were utilising tests of manual dexterity (such as plaster and wax carving). He noted the use of psychological testing suggesting its purpose would be to identify those applicants

most likely to reason logically. A whole range of other tests are cited including reading accuracy, memory and visualising patterns. The author noted however that whilst use of such tests was widespread, dental schools acknowledged the need to understand significantly more about the role of such tests in selection. The paper reported the creation of the Committee on Dental Aptitude Testing (formed by the Council on Dental education) which would be working with dental schools to explore further the use of tests of aptitude in dental selection.

3.3.2 Validating the DAT

The Council of Dental Education of the American Dental Association conducted nationwide testing from 1946 onwards. A progress update in 1947 (Peterson 1947) confirmed that dentistry had its own nation-wide testing programme. One aim was to predict success or failure on dental programmes, drawing on the results of a range of tests taken on entry (rather than at the point of selection). The tests included an intelligence test, a test of science, a test on interpretation of reading materials in the natural sciences, object visualisation and a carving dexterity test. Other tests were being considered for future rounds. Dental schools administering the tests also intended to provide student assessment records in order that the value of the different tests could be explored. The scope of this exercise is incredibly impressive.

The University of Michigan's reported its experience in using admission tests in selection to the School of Dentistry in 1947 and 1948 (Travers and Wallace 1950). The tests were wide ranging including tests of both manual dexterity, critical thinking and verbal reasoning. Scores were correlated with programme scores of those admitted. The study focused on inconsistencies in findings between the two years; the tests had some predictive value in 1948 but none in the previous year, noting that changes to selection processes might have impacted on outcomes of such studies.

Following a successful trial (Peterson 1948), dental applicants in 1951 were required to sit the Dental Aptitude Test (Peterson 1951). The Committee were able to report the success of the trial and that high correlations had been identified between test scores and assessment outcomes. In addition, Peterson claimed that a large proportion of those students who failed to complete the programme would not have been admitted had the aptitude test been in use. The paper described how the test would be administered, noting that information regarding test content including sample questions should be made available to candidates.

Results from the University of Kansas (Weiss 1952), concluded that a combination of college grades and the test of science would help screen out some poorly performing students. However, Weiss highlighted the limitations in only looking at the results of those who entered dental college as motivation might affect test

performance if used in selection. In addition, he urged caution with the use of some of the tests, given variations in performance between years.

Layton went on to report on a similar exercise at the University of Minnesota (Layton 1953). He commented that the Council of Dental Education had taken a decision to use the test battery with little evidence to support its use and concluded that the tests did not predict performance on the dental programme. Whilst there was evidence that the tests performed well on a nationwide basis (Peterson 1948), local studies appeared to have more mixed findings. Whilst any conflict between research findings requires careful consideration, it is interesting that Layton felt that the local study presented more definitive results than the national programme, rather than reflecting on the limitations of a smaller local study.

A further local study at Emory University Dental School (Webb 1956) noted significant changes taking place across the period of evaluation including change to candidate demographics, test content and curriculum; highlighting these as some of the challenges faced by any attempts to conduct longitudinal studies in education.

In 1954 Tinkleman reflected on factors predicting academic success in medical and dental schools (Tinkelman 1954). In 1949, New York state had offered scholarships for studying medicine and dentistry. Competition for places on this programme had been such that the education department decided to use a written examination to aid selection. The exam covered professional aptitude and a test of science with some similarities to the Medical College Admission Test. The study sort to assess whether the scholarship winners outperformed those that had not won the scholarship during their time at medical and dental school. The results showed that scholarship winners achieved significantly higher grades in first year examinations then non-winners. Science tests predicted more than the aptitude tests. The authors went on to recommend greater weighting of the science element in future.

In a single-centre study exploring the ability of selection criteria to identify those at risk of failing to complete a dental programme, only the Dental Aptitude Test was found to independently predict successful graduation (Scheetz 1987). A national sample of 5009 dental students provided an opportunity to explore the ability of the DAT to predict performance in dental school more broadly (Kramer 1986). The findings indicated that whilst the strength of the correlation was lower, the DAT predicted performance in dental schools beyond that provided by grade point average (GPA). In addition, these two measures tended to predict different outcomes. As in other studies the authors commented on the effect of range restriction in such analyses. In conclusion (and similar to Boyd et al above), Kramer concluded that the combination of the DAT and GPA improved relationships significantly over and above using either factor independently.

3.3.3 Broadening the Debate

A Dental Aptitude Test (DAT) had been introduced in Canada in 1967 (Boyd, Teteruck et al. 1980). In their paper Boyd et al looked at the literature around the Canadian programme in order to inform Dental selectors. They noted that whilst there was no consistency in selection processes, most dental schools utilised achieved education grades, performance in the DAT, interviews and letters of recommendation. Challenges around validating the test included restriction of range, the reliability of outcome measures and differences in assessments between institutions. Existing studies looking at predictive validity showed contradictory results. Achieved grades provided the best evidence of prediction whilst the DAT did show some ability to predict across the whole dental programme. The authors therefore suggested that DAT scores ought to be used to complement or reconfirm these scores. The ability of the manual elements of the test to predict outcomes were again mixed although there appeared to be sufficient evidence to use this test to exclude poor performers. Concerns about repeat testing and coaching effects in the tests of manual dexterity and given rise to the production of manuals to assist candidates in preparing for these aspects of the test.

A review of dental selection undertaken in 1988 provides a more up to date record of selection practices in Australia, the UK and North America (Spratley 1990). In Australia, selection took place almost exclusively using candidates' academic records; interviews were not routinely undertaken. In the UK, admission had been based on academic records. Until the 1970s, interviews had been standard across dental schools alongside in some cases aptitude testing focussing primarily on tests of manual dexterity (such as wax carving or wire bending).

An evaluation of admission to dental education in 2005 reviewed the literature in relation to the predictive ability of college scores, aptitude testing and measures of manual dexterity (Ranney, Wilson et al. 2005). Based on the literature the authors concluded that college scores and DAT could be used as predictors of early academic performance in schools. Tests of perceptual ability and manual dexterity were best viewed as screening tools, assuming a baseline score could be identified. The authors went on to support the use of structured interviews.

A study in Ireland (Lynch, McConnell et al. 2006) reviewed the utility of the Irish Leaving Certificate in predicting performance in dental school and concluded that it had limited value in this respect. The authors advised consideration of alternative tools such as previous academic performance, cognitive and non-cognitive ability and manual dexterity.

An international group of dental educators presented a paper in 2002 (Gangler, de Vries et al. 2002) looking at both student selection and the impact of learning environments on educational outcomes. The paper discusses best practice and

innovations in student selection and amongst other recommendations, calls upon research to improve the predictive power, reliability and validity of existing selections methods. The authors commented on the reasonable predictive power of aptitude testing. They also note the use of a range of psychometric tests looking at the non-cognitive skills of prospective trainees, noting existing evidence suggesting that moral reasoning skills were associate with excellent clinical performance. The paper called for global solutions to selection through bringing together networks of dental academics to determine core value systems and explore common issues in selection.

3.3.4 Non-Cognitive Testing

In a study published in 1944, Thompson sought to link success in dental school to personality and interest scales (Thompson 1944). This built on previous interviews he had undertaken with dentists regarding the importance of motor and mechanical abilities. The author concluded that whilst there were some correlations between these personality measures and academic outcomes, they were not sufficiently strong for use in selection.

Personality tests trialled at a dental school were shown to have little utility in selection as they failed to discriminate between high and low performing students or accepted and rejected applicants (Kalis, Tocchini et al. 1962). The authors acknowledge however the difficulty of identifying meaningful and reliable outcome measures (in this case ratings by members of faculty were used). At the same time the adequacy of the personality tests used was questioned, and in particular the impact on results of the context in which the tests were taken. The authors also note that as failure and dropout rates were so low, predicting their occurrence would always be challenging.

The first year class at the University of Pittsburgh undertook a biographical inventory of extra- curricular achievement in 1965 (Mackenzie 1967). This recorded achievement in science, dramatic arts, literature, music, art and leadership. The authors concluded that those who had participated in such extra-curricular activities prior to admission to dental school were likely to continue participating in such activities during their time in the dental programme. The authors went on to conclude that prediction of 'leadership potential' would be of value to selection to dental school on the basis that such individuals would go on to seek leadership positions within the profession or their communities. This argument however is not very clearly made and the utility of such a tool for these purposes questionable.

The University of Bristol reported use of psychometric testing (similar to that taking place in North America) which appeared to have some utility (Deubert, Smith et al. 1975).

A report in 1976 challenged the consensus around dental selection in the US (Barkley 1976). The author comments on inflated GPA scores, unscrupulous practices in achieving test scores and the danger in recruiting only those that good at tests. He noted that despite the accepted rigour of selection, dental schools were still admitting students who would go on to struggle on the programme and in some cases drop out. The author claims that one significant issue was the focus on academic achievement at the expense of identifying those applicants with the potential to develop meaningful relationships with patients and demonstrate professional effectiveness. The author had worked with a selection firm to develop a tool which might identify the future 'outstanding dentists, aware of and concerned with relationships with people.' Whilst the tool clearly needed further development this paper suggested a potential shift in selection to take into account those skills required to produce successful practitioners.

3.3.5 The United Kingdom

Research specific to the UK context around dental selection is sparse. In 1989, Duguid reported on application trends to dental training from 1968 to 1987 (Duguid 1989). The paper focusses on applicant numbers noting the drop in application numbers observed over the period. A follow up study in 1994 (Duguid 1994) reported on reduced dental training places and a continued decline in applications although it appeared these were picking up in the later years under review. The reduction in places alongside increased proportions of mature and female students had implications for future manpower planning.

A paper published in 1997 drew attention to student dropout rates in dental education becoming a significant issue. The authors reported dropout rates in 1992-1994 of 14.8% (Drummond and Duguid 1997). Given the significant cost to the system of students not completing the course, the authors commented that a less expensive solution would be to improve selection. They did not however offer any solutions or suggestions which is disappointing as this may have been an opportunity to open up this debate.

A review of selection to the Manchester Dental programme provided an insight into the issues challenging admission tutors in the UK (Hoad-Reddick and MacFarlane 1999). The authors set out to evaluate the ability of selection data to predict first year performance on the programme. Those with high interview scores and leadership experience tended to perform better in year 1 of a PBL programme. Those students who had taken Biology at A-level performed better in the first semester.

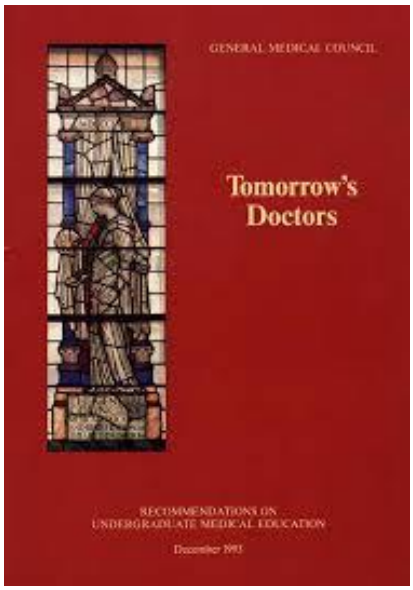
A paper published in 2000 provided a helpful overview of UK dental selection from 1983-1998 (Duguid and Drummond 2000). The paper however focussed on trends

in student demographics rather than the process of selection noting amongst other trends the increased proportion of female dental students.

A survey of applicants to Dundee and Manchester dental schools again focussed on the demographics of the applicant pool (Stewart, Drummond et al. 2004). Whilst noting the increasing representation of ethnic minorities amongst applicants, the authors commented on the need to broaden the social and cultural intake alongside existing widening participation schemes.

3.4 National Policy

Having looked at the literature around selection in the UK and internationally, this section seeks to highlight some of the national reports and policy initiatives which impacted on medical and dental selection prior to 2006.



In considering external drivers for change in selection, a sensible starting point is **Tomorrow's Doctors (General Medical Council 1993)**. Tomorrow's Doctors contained recommendations for a revised curriculum framework for undergraduate medical education. The authors noted drivers for change in medical education including an increased focus on public health, a shift between hospital based and community care and an aging population. The publication included a call to reduce the factual load on students and increase their ability to work independently – addressing ongoing concerns regarding the 'overcrowded curriculum'.

The introduction of a comprehensive list of 'attitudinal objectives' had implications for selection criteria. These objectives included respect, awareness of moral and ethical responsibilities, ability to cope with uncertainty, ability to adapt to change and the ability to work in a team. In addition, communication skills were introduced as a curriculum theme.

Comments on selection and associated criteria in the report are brief but contain some indications of future direction. The authors endorsed the approach by many schools to encourage students to take a gap year prior to entry to their programmes. Noting that selection was a matter for individual schools, the authors were encouraging of the 'trend towards liberalisation of entry requirements so that students may continue to study a broader range of subjects'. School and college advisors needed to be well informed of criteria for entry to medical schools. The authors went on to note that medical schools were 'obliged to rely more heavily on

evidence of academic achievement than they would wish' at a time 'when there is increasing emphasis on non-academic attributes' expected in members of the profession. The authors looked forward to the possibility of developing selection procedures that explored such attributes which might be 'administered on a consortium basis'.

The Report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995 (Kennedy 2001) was commissioned to inquire into the management of the care of children receiving complex cardiac surgical services at the Bristol Royal Infirmary between 1984 and 1995. This followed evidence of higher than expected infant mortality in the unit. Whilst appropriately concentrating on clinical services, the report also touched on issues relating to clinical education and selection to healthcare programmes. The report commended the growing focus on communication skills within medical curricula but noted a failure in clinical education to develop an understanding of the NHS, its values and how it operates.

In relation to selection specifically, the author commented on a heavy reliance within selection processes on the UCAS form and a brief interview, leading to criticisms of superficiality. He went on to refer to a perceived lack of transparency in some processes and limited input from the wider health service. Building on recommendations within Tomorrow's Doctors, medical schools were beginning to look at a wider range of skills required by medical professionals. The author supported the work being undertaken by the GMC to give greater definition to the qualities of a good doctor which might inform selection processes in the future.

'An understanding of science may be a necessary condition for entry to medical school, but it cannot be sufficient. The future doctor must also have demonstrated other qualities, not least a capacity to be open-minded, comfortable with uncertainty, free of preconceived views and capable of recognising and responding to ethical issues.'

The author pointed to the tension between GMC ambitions around selection and the autonomy of universities regarding selection requirements. He emphasised the need for a discussion regarding selection across a wider community including the NHS and the public.

Ten years later the GMC updated **Tomorrow's Doctors** taking account developments over the decade and lessons learnt from the implementation of the original recommendations (General Medical Council 2003). The revised document emphasised even further issues regarding attitudes and behaviours suitable for the profession – respect and communication being core to this.

Regarding selection, the authors called for 'valid, open, objective and fair selection' processes to be put in place. They noted that information regarding processes should be published and those involved in selection should receive training and include 'individuals with a range of expertise and knowledge'.

Reporting in 2004, the Admissions to Higher Education Group had been charged with developing statements of high-level principles to guide admissions to UK universities. This resulted in the **Schwartz Report – Fair Admissions to Higher Education, Recommendations for Good Practice, Department of Education and Skills 2004 (Schwartz 2004)**. The report noted the massive increase in demand for higher education over recent years and commented upon the significant diversity within the system both in terms of programmes offered and the ultimate student experience.

The authors emphasised that fairness within the admissions system was necessary because, despite diversity, some groups remained under-represented. The Group accepted that university selection processes were generally fair. However, they identified several areas for improvement, focussing on information provided to applicants regarding the assessment of applications and the reliability and consistency of information.

The report noted particular difficulties in selecting for courses with high demand and highly qualified applicants. Additional assessment was regarded as a 'burden' to applicants. The use of predicted grades created uncertainty and potential unfairness in the process.

The group considered that prior academic attainment remained the best predictor of success in higher education. However, they noted that the impact of social background on academic attainment created a potentially uneven playing field – an issue complicated further by the knowledge that, all other things being equal, students from the state sector outperform those from the independent sector at

University. There appeared to be broad consensus that applicants' educational contexts ought to be considered within selection processes.

Recommendations within the report's conclusions included the need to look at educational context; consider other methods of assessing potential (e.g. non-academic experiences, additional tests, outcomes from interviews etc.); ensure that assessment methods were reliable and valid; and to minimise barriers.

Post qualification applications was regarded as a solution to many of these issues. The authors noted the potential of the review of the 14-19 curriculum (Tomlinson report) to provide opportunities to stretch the most able applicants and provide greater granularity for the consideration of applicants applying to over-subscribed programmes.

The authors recommended that research take place to look at the *ability of admission tests to predict performance* in the undergraduate curriculum.

In a report in 2004 (**Medical Schools: Delivering the Doctors of the Future**), the Chief Medical Officer commented on the state of undergraduate medical education in the UK (Department of Health 2004). Medical education was expanding with recruitment targets set to be 57% above 1997 numbers by 2005.

3.5 Key Findings

In 2005 medical selectors came together to discuss how an aptitude test might improve selection to medicine and dentistry in the UK. The original aims of the organisation (as discussed later in this thesis) included improving fairness in selection, improving the ability to discriminate between high ability candidates, exploring how desirable non-cognitive traits could be assessed within selection processes and widening access. Alongside these plans was an ambition to underpin the development of the test with a relevant programme of research into the test and wider selection issues. This chapter suggests that these ambitions were firmly underpinned by the then available research evidence.

Much of the early evidence regarding how measures of *academic achievement* predict performance in medical and dental schools relied on MCAT and DAT data. Relevant studies confirmed that both the MCAT and DAT predicted performance at medical schools although MCAT's ability to predict performance declined in the clinical years of programmes (Mitchell 1990). MCAT could be used to identify students likely to struggle in medical school e.g. (Conger and Fitz 1963) but correlations were stronger with MCAT science components (Roemer 1965). After accounting for the predictive ability of the MCAT, a significant amount of the variation in medical school performance remained unaccounted for (McGaghie 1990).

The MCAT is of course very different from the school leaver qualifications in the UK (e.g. A-levels) but interestingly many of these findings have been mirrored in more recent studies in the UK (Tiffin, Mwandigha et al. 2016). In 2005 UK selectors were able to reflect on the MCAT findings when considering options to help discriminate between high achieving applicants. Grade inflation is not a recent phenomenon in the UK (McManus 1982) with the challenge of discriminating between high achieving candidates highlighted similarly (McManus and Richards 1984). Evidence around the ability of A-levels to predict performance in medical schools has been available for some time (McManus and Richards 1986) although selectors were warned against overemphasising their use in selection given their limited utility in discriminating between candidates and the remaining lack of understanding in what actually predicts success.

Selectors in the US were urged as early as 1959 to be sceptical about the value of college recommendations (similar to school references included in UCAS forms) in selection processes (Hill 1959). In the UK context, doubt was cast on the central role of references in selection given the absence of evidence around their predictive ability (Roberts and Porter 1989). Similarly, personal statements did not appear to predict performance in medical school (Ferguson, Sanders et al. 2000). By 2005 there was unease regarding use of personal statements in selection with some of the newer medical schools choosing not to score them within their selection processes (Searle and McHarg 2003). Whilst removing the use of personal statements in selection was an ambition at the time and for some undoubtedly a driver in the creation of the UKCAT, a number of medical and dental schools continued to use personal statements and references as a central part of selection until relatively recently (Greatrix and Dowell 2020).

As early as 1948 the discussion around selection to medicine highlighted a tension between selecting those that would be successful academically and those who would be successful in the professions (Shoemaker and Rohrer 1948). The debate around how to select those with the required traits to be successful clinicians (and indeed what those traits were) dominated the literature over the next 70 years and continues to be a source of enquiry today. Similar issues are raised in the dental literature with concerns regarding reliance on academic ability at the expense of identifying applicants with the ability to be professionally effective. Later in the century a consensus emerged regarding the personal qualities associated with excellence in medical education (McGaghie 1990) but no significant solutions have been offered around how to assess these within selection processes. More recent analysis in the UK regarding the ability of a range of personal traits to predict outcomes at medical school concluded that only conscientiousness was an independent predictor – a trait which has been further explored (McLachlan, Finn et al. 2009) but proves difficult to assess at point of selection. Difficulties identified in

testing non-cognitive traits within the UK setting included the reliability of such measures, noting the challenge of assuming such traits might be constant over time and the potential for candidates to fake outcomes (Hughes 2002).

Medical schools in the UK were urged as early as 1986 to monitor selection processes in terms of diversity and to recognise academic achievement in the context of educational opportunity (Horton 1986). Little additional UK research exists beyond that until after 2006. However, the publication of the Schwartz report into fair access in 2005 (Schwartz 2004) was clearly informing thinking amongst selectors at the time the UKCAT Consortium came together. This report called for greater fairness in admissions whilst some groups remained under-represented in higher education. The authors also noted particular challenges in selection for courses with high demand and highly qualified candidates such as medicine and dentistry.

A debate regarding the use of aptitude tests within selection to medicine appears to emerge early this century following on from the Schwartz report which suggested their potential use more widely as a tool for widening access. Whilst one paper suggests that there was little evidence that genuine tests of aptitude had any predictive value, it did go on to suggest that medical schools might wish to commission their own test (McManus, Powis et al. 2005). The paper goes on the emphasise the need for such a test to be developed in parallel with a specific research agenda.

This review of literature suggests that until relatively recently research into medical selection was dominated with outputs from the United States. Outside the United States the dental literature is sparse and largely limited in the UK to studies focusing on applicant numbers.

Many studies emphasise the difficulties in undertaking research in this field and the limitations researchers face. Longitudinal studies are difficult to undertake especially across institutions where data, curricula and cohorts may vary. Taking account of such variation creates challenges in designing such studies and interpreting results (Little, Gee et al. 1960, McManus, Dewberry et al. 2013). The need for long term cohort studies to enable informed analysis of the value of selection criteria in the UK was called for by the Medical Schools Council in 2002 (Ferguson, James et al. 2002).

UK national policy drivers (predominantly relating to medicine) both inform and are informed by available research literature. The first 'Tomorrow's Doctors' document comments on the challenge of emphasizing non-academic attributes alongside academic achievement (General Medical Council 1993). Subsequent versions of the document further emphasised the needs for selection to include assessment of

attitudes and behaviours suitable for the profession. The Schwartz Report and subsequent national focus on social mobility would go on to shift the debate around selection from one of fairness/utility to widening access but little of this work had been undertaken prior to the creation of UCAT. This will be discussed further in Chapter 5.

Having set the scene for selection to medicine and dentistry prior to 2006, the next Chapter describes how both the UKCAT Consortium and the test itself has changed and developed over time.

Chapter 4: Creating and Developing UKCAT

4.1 Introduction

One of the intentions of this thesis is to record how both the test and Consortium have changed and evolved over time (see 2.6.1). This chapter includes a timeline of organisational development. Document Review (as outlined in Chapter 2) provided a tool to draw on the content of meeting papers (UKCAT Board, UKCAT Consortium, Research Group and Test Delivery Group), this chapter records significant developments in the organisation since its establishment in 2005. Given my position in the organisation, reflexivity is of importance in this piece of work to mitigate any risk of my interpreting information from documents beyond that which is formally presented.

4.2 Methods

From my unique position in the organisation I was able to access electronic records of all minutes of UKCAT's meeting from the creation of the organisation. Minutes were originally extracted into a word document in a table format. The document recorded the committee's name, date and those present. This document created a full record of formal minutes of the organisation. Over a period of time the content of this document was significantly reduced to create a chronological record of notable developments in the organisation. Content was largely limited to factual reports of specific events and significant discussions.

A number of discussions took place over the presentation of this information with the intention that the record created could be maintained in future years.

In addition to creating a brief history of UKCAT's development, this timeline provides evidence to help evaluate the extent to which UKCAT has achieved its aims, both those originally set and those that have been further developed over time.

Information regarding UKCAT's original and developing aims was separately extracted given the intention to focus on the extent to which these had been achieved. These are presented below.

Information regarding test content had originally been presented elsewhere but it was felt that incorporating high level information into the timeline would allow the reader to more easily see how the test had changed over time. A summary of test content is provided for each testing year alongside high-level comments regarding relevant developments to the test providing an information source for researchers conducting research in the future where changes that have been made to the test over time may be of interest. More detailed information regarding the development of the test can be found in Appendix C.

Information extracted from the minutes fell naturally under the headings below (governance, candidate experience and research). Research publications (where work had been supported by UKCAT) in each year is reported to give a sense of the emerging evidence

base and research activity over time. At the same time and for similar reasons presentations at Consortium Meetings are recorded.

Descriptive Statistics data provided below is largely extracted from the UKCAT annual Technical Reports e.g. (Pearson VUE 2020). These documents are produced on behalf of UKCAT by Pearson VUE. For each year of testing the Technical Reports provide a commentary on candidate performance (including that of sub-groups) alongside analysis of both test and item performance. Summary documents, containing much of the data relating to candidate performance, are made available on the UKCAT website.

Data was extracted from each of the annual Technical Reports into an excel spreadsheet from which the figures below were created. These data are presented to show trends over time for the first time. The information in this section is intended to provide additional context to some of the challenges (some of them ongoing) referred to elsewhere in this thesis.

A list of UKCAT Consortium Universities (since 2006) is provided as Appendix A.

A list of UKCAT Board Members (since 2006) is provided as Appendix B.

4.3 UKCAT's Aims

4.3.1 Original Aims

The first UKCAT Annual Report was produced in 2006 and set out the background to the creation of the organisation and development of the test:

The UKCAT was conceived to improve the fairness and objectivity of the Admissions process for clinical subjects… It arose partly from a widespread feeling that A-levels…were failing to discriminate between candidates at the upper end of the scale of academic ability. ….. Behind this dissatisfaction with A-levels as an indicator of ability was an additional worry that the qualities tested by these exams might not have been entirely appropriate as a way to select students for the clinical professions: that, increasingly, A-levels appeared to be testing an ability to learn facts rather than an aptitude for critical thinking and problem-solving. A new tool for selection for medicine and dentistry might offer the opportunity to select on the basis of characteristics that the medical and dental schools themselves thought would be more appropriate; it might also help to widen access by identifying academic potential in applicants from less-advantaged educational backgrounds. (UKCAT Consortium 2006)

It was intended that the UKCAT would help select students who would perform well in medical and dental school and eventually make good doctors and dentists. At the same time, it was hoped that the test might identify those students likely to struggle during undergraduate studies.

The 2006 Annual Report went on to note that a research programme was being established to allow longitudinal follow up of test candidates which would be important in helping

achieve the aims of the organisation. It was anticipated that the outcomes from such a study would ultimately inform the development of the test.

By 2007, the objectives of the organisation had synthesised into the following statement which has subsequently been largely replicated in UKCAT's Annual Reports:

UKCAT remains committed to achieving greater fairness in selection to medicine and dentistry and to the widening participation in medical and dental training of under-represented social groups. Through an ongoing programme of research UKCAT is seeking to identify the characteristics in applicants which will make them good dentists and doctors and thus to improve the quality of those that enter the profession with the ultimate aim of improving patient care. (UKCAT Consortium 2008)

4.3.2 Developing UKCAT's Aims and Objectives 2010 - 2013

In February 2010 the UKCAT Board reviewed actions required to continue to deliver and develop the aims of the organisation. Objectives identified were to:

- identify key stakeholders and external organisations to inform the creation of a communication strategy to further develop relationships and better promote UKCAT;
- embark on a longitudinal cohort study of test candidates using university assessment data;
- undertake analysis of the impact of UKCAT on widening participation;
- determine long term options for the development of the UKCAT research database;
- engage with the Consortium regarding their aspirations/concerns and requirements for support;
- share good practice regarding the use of the test within the Consortium;
- review 'fitness for purpose' of each of the cognitive sections and identify the best fit of the current non-cognitive tests.

In 2011 and 2012, the UKCAT Research Working Group led the development of a research strategy for the organisation. The strategy acknowledged that research outputs to date had been disappointing, but that progress was being made with some recent invitations for researchers to tender to undertake research. Data management arrangements had been put in place, which had facilitated the provision of datasets for two studies. The paper noted that the ability to now support ongoing research required the organisation to set out priorities and make decisions regarding its willingness to commit financial (and other) resources in this area. The paper emphasised the significant potential of using the unique resource that was the UKCAT database to support research into selection.

The overall research aim at that point was to create an improved evidence base for the selection of medical and dental students in the UK. To do this UKCAT needed to continue to

refine its research database into an effective resource for academic research, ensuring that data quality and quantity were maximised through further involvement of stakeholders and, in the future, linkage through to postgraduate education. UKCAT set out its ambition to utilise the database alongside research investment to position the organisation as a leader in the field of selection research.

To achieve these ambitions, the working group identified several objectives which included:

- identifying research priorities;
- increasing provision of progression data from Consortium members;
- exploring additional mechanisms to obtain candidate consent regarding the use of their data in wider research activity;
- exploring the potential to link UKCAT data to postgraduate performance data.

Priorities for research studies at that point were:

- Modelling the impact of increasing use of UKCAT on the interview pool;
- Establishing the predictive validity of non-cognitive instruments on relevant outcome markers;
- Exploring the widening access potential of UKCAT by examining differences in school attainment and UKCAT scores utilising contextual data;
- Assessing the impact of preparation on UKCAT scores.

The Board considered a few resourcing solutions to support the delivery of these objectives. These included funding additional tenders to support specific research projects, exploring the possibility of direct employment of research expertise, commissioning a larger research programme through an academic unit and utilising existing Pearson VUE statistical expertise.

4.3.3 UKCAT Strategic Plan 2013

In February 2013 the Board drew on feedback from a survey of Consortium Members to develop a strategic plan comprising six themes.

The survey took place in January 2013 and included Consortium Representatives, Deans of medical and dental schools, administrators and other stakeholders. 70% of respondents reported that the UKCAT was either an important or very important part of their selection processes. A significant majority considered UKCAT to have at least partially achieved its objectives regarding improving fairness (82% achieved/partially achieved), discriminating between candidates (90%) and widening access (70%). Respondents were less confident that the test at that stage assessed the traits required for clinical practice.

Respondents were also asked to prioritise issues which might drive test development. These were: discriminating between the many high achieving applicants to medical and dental

programmes; testing the attributes required to be a successful medical or dental student; testing the attributes required to be a successful doctor or dentist; addressing the widening participation agenda. Respondents struggled to choose between these priorities with little consensus emerging.

When asked to comment on future research priorities, most commented on the need for further work around predictive validity at postgraduate level and in medical/dental schools. The lack of predictive validity evidence was a limiting factor for schools when asked if they intended to use the test more strongly in the future.

After further input from Consortium members, a final strategy document was approved by the Board in October 2013. The plan included the following mission statement, building upon the previous statement:

UKCAT is committed to achieving greater fairness in selection to medicine and dentistry and to the widening participation in medical and dental training of under-represented social groups. Through an ongoing programme of research UKCAT is seeking to identify the characteristics in applicants which will make them good dentists and doctors and thus to improve the quality of those who enter the professions with the ultimate aim of improving patient care.

The strategic plan included a section reporting UKCAT's achievements to date, drawing on input from the Board and Consortium members. These included:

- Secure and flexible delivery of a test to around 25,000 candidates annually with the UK and with an increasing international reach;
- Use of test bursary to candidates with low incomes (at that point 10% candidates);
- Provision to universities of a reliable and stable mechanism to discriminate between high achieving candidates;
- Growth in strength of use of the test in recent years;
- Research evidence that the test could contribute to widening access;
- Research database with growing capacity to contribute to the research evidence base around admissions;
- Increasing internal reliability of the test and face validity of content;
- Improvements in information and advice to candidates;
- Anticipated publication of the first predictive validity study;
- Provision (though the consortium) of a forum for the discussion of best practice in admissions.

The following aims were identified within the strategy, with a series of objectives underpinning their delivery:

- **Widening Access**: UKCAT will contribute to the wider debate in the UK regarding widening access whilst sharing best practice within the Consortium regarding candidate preparation for and consortium use of the test.
- **Research**: UKCAT will position itself as a UK leader in undertaking and supporting research regarding admissions into medicine and dentistry, prioritising research outputs leading to improvements in the test and its use.
- **Improving and Developing UKCAT**: The UKCAT test will be improved based on research and other evidence. Opportunities to further expand use of the test (or aspects of the test) both nationally and internationally will be identified.
- **Governance and Communications**: UKCAT will have a reputation as a forward thinking, dynamic organisation at the centre of developments in admissions to medicine and dentistry in the UK.
- **Candidate Experience**: The candidate experience from initial communication to sitting the test will be the best possible.

Subsequent discussion regarding the resource implications of delivering the strategy led to additional staffing resources being identified for marketing and communications. The financial impact of investment in research capacity was also considered.

The Strategic Plan was reviewed by the UKCAT Board in December 2016. Whilst the aims remained relevant the Board felt additional focus was required around international development and the development of the test itself.

4.4 UKCAT Organisational Timeline

2005	Creating the UKCAT
	GOVERNANCE
April 2005	Meeting convened by Medical Schools Council of the '**Testing for Undergraduate Medicine Selection'** Subgroup. Interest in creating an admissions test from 12 universities. It was anticipated the test would comprise both a cognitive and non-cognitive section and be no more than three hours long. Mode of delivery of the test had not been decided. Presentations from three test developers (Australian Council for Education Research (ACER), University of Cambridge Local Examinations Syndicate (UCLES), The University of Newcastle, New South Wales).
May 2005	At a meeting of the **Medical Schools Council Admissions Deans**, Professor Chris McManus (UCL) presented findings from current selection research: A Levels were the only good predictors of academic outcomes; personality and learning styles the main predictors of non-academic outcomes; satisfactory progression in medical careers could not be predicted by personal statements or references. Professor Keith Millar (Glasgow) presented on the use of psychometric testing in admissions. Such tests could be used to assess characteristics associated with success as a doctor and to determine if new criteria would alter student demographics. Longitudinal studies were required to confirm the predictive validity of potential tests. A workshop took place exploring the non-academic qualities being looked for in selection.
July – Sept 2005	10 medical schools committed to using an admissions test in 2006 (2007 entry) with a further 12 committed to 2007 testing. In total 22 universities intended joining the Consortium with additional interest from Veterinary Schools. Dental Schools were to meet to consider the proposal. Test name confirmed as the United Kingdom Clinical Aptitude Test (UKCAT). Professor Ian Johnson (University of Nottingham) appointed as first UKCAT Chair.
Oct – Dec 2005	'Shell company' with three founding Directors established. 23 universities had agreed to join the Consortium. Launch of the test was to take place at December Medlink (a national recruitment event for medical schools). Four companies had responded to an invite to tender for the delivery and development of the UKCAT.

2006 First Year of Testing			
TEST CONTENT AND DEVELOPMENT			
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**
40 scored items **4 unscored items** **22 minutes**	36 scored items 4 unscored items 22 minutes	60 scored items 5 unscored items 16 minutes	26 scored items 30 minutes
• **Two forms of each subtest were developed using operational items which had been trialled previously. Further versions of the test were created to allow for pre-testing of additional (unscored) items. Versions of the test were to be assigned randomly to candidates. An extended test was to be available for candidates with learning difficulties.** • **Item calibration, scaling and equating based on Item Response Theory model. Early testers did not immediately receive scores. Once enough candidates had tested, forms were equated, allowing candidates to then receive scores immediately.** • **Scale scores for each subtest ranged from 300 to 900 with a mean of 600 in the reference sample. Universities received a score for each subtest plus a total score (sum of each subtest score) for their applicants.** • **The first three subtests performed satisfactorily. Amendments to Decision Analysis were needed to increase subtest reliability. DIF analysis demonstrated that the test construct did not disadvantage different candidate subgroups.**			
	GOVERNANCE		**RESEARCH**
Jan – March 2006	Contract agreed with Pearson VUE for test delivery and development. UKCAT office to be based at the University of Nottingham. Chairs of Test Development and Delivery Committees identified. UKCAT Logo adopted. Test fees: £60 UK/EU (discount for testing early), £95 International. Full fee bursaries available to eligible candidates. University of Birmingham withdrew from the Consortium.		The need to obtain candidate consent in order that data could be used in future research was noted.
April – June 2006	UKCAT website live. Agreement with UCAS to support delivery of test results to universities.		
July - Sept 2006	Following test equating, candidates testing after 8th September obtained their results immediately following testing.		
Oct – Dec 2006	Results issued to universities. University of Warwick and Imperial College London to join the Consortium.		Research Working Party established.

2007 Behavioural Test Introduced				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**	**Section 5 (MEARS, NACE, MOJAC)**
40 scored items **4 unscored items** **22 minutes**	36 scored items 4 unscored items 22 minutes	60 scored items 5 unscored items 16 minutes	26 scored items 30 minutes	 30 minutes
• **A fifth subtest was introduced (the Behavioural Test) intended to assess non-cognitive traits of empathy, integrity and robustness. Three different instruments piloted: MEARS (Managing Emotions and Resilience Scales); ITQ100 (Interpersonal Traits Questionnaire)/NACE (Narcissism, Aloofness, Confidence and Empathy); IVQ49 (Interpersonal Values Questionnaire)/ MOJAC (a measure of ethical orientation). Abridged versions of ITQ (ITQ50) and IVQ (IVQ33) were combined and piloted.** • **Anomaly identified in Abstract Reasoning with 1480 candidates incorrectly marked. The error was identified in October and required all candidate results for this section to be withdrawn. Performance in Quantitative Reasoning had shifted upwards.** • **Pearson VUE presented a paper offering advice on the use of the test in selection. The paper detailed information on the meaningfulness of both subtest and total scores noting the importance of standard errors of measurement when considering results. Universities were advised to use the test in combination with other criteria during selection.** • **The higher performance of candidates who sat the UKCATSEN (+25% additional time) was noted.**				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
Jan – March 2007	2007 test fees set at £60 and £75 (HEU) and £95 (International). Process initiated to apply for Charitable Status.	Candidates reported it would have been useful to know they might not complete subtests. Issues reported with calculator. Requirement to scroll on some items not clear. Preparation courses and books not useful; questions provided by UKCAT helpful. Marketing plan discussed focussing on reaching graduate and WP candidates.		

		Concern regarding preparation companies; improvements to UKCAT materials discussed. Pearson VUE to provide customer service to candidates in 2007.	
April - June 2007	Amendments to the company articles agreed at extraordinary general meeting. Queens University Belfast joined the Consortium. Lakin Rose appointed as company accountants appointed.		Selection process details to be requested from Consortium. Call for tenders to host the UKCAT Research database. Australian Medical Schools Outcome Database presented a feasibility study for a database enabling longitudinal studies.
July - Sept 2007	Arrangements discussed for future Board elections.		
Oct – Dec 2007	Board agreed to meet more frequently and to direct the work of sub-committees. Number of elected Board members to be reduced by one, requiring an alteration to the articles of association.	Handheld calculators to be used in 2008. Issues with scoring Section 5 (non-cognitive test) had resulted in incorrect scores communicated to candidates. Results for the abstract reasoning subtest had been withdrawn due to marking issues.	In interim, University of Nottingham to store candidate data. Initial dataset available to support study comparing UKCAT and school examination performance. Chair had written an article 'Experience with the First Year of the United Kingdom Clinical Aptitude Test'.

2008				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**	**Section 5 (MEARS, NACE, MOJAC)**
40 scored items **4 unscored items** **22 minutes**	36 scored items 4 unscored items 22 minutes	60 scored items 5 unscored items 16 minutes	26 scored items 30 minutes	 30 minutes
• **Following 2008 testing, new software allowed UKCAT to expand the item bank by increasing the number of pre-test items.** • **Lower reliability of Decision Analysis remained a concern; agreed to review construct of the subtest and consider alternatives.**				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
Jan – March 2008	Decision made to appoint a Chief Executive.	New website under development to include additional preparation materials.	Abstract to AMEE: 'UK Medical and Dental Schools use of UKCAT data; the first two years'. Funding approved to survey use of UKCAT in medical selection. Advice needed on research database specification.	
June – Sept 2008	Sandra Nicholson appointed as UKCAT Chair. Application for charitable status submitted.		Agreed to collect university assessment data for the 2006 cohort. Preliminary findings showed a correlation between UKCAT and A-Level results. Consultant engaged take forward the research database tender.	
Oct - Dec 2008	Revisions to Articles of Association approved.	Candidate appeals process to be developed for 2009.		
Dec 2008	Presentations at Consortium Meeting: • How UKCAT is used by Schools (Jane Adam, Hull York Medical School) • The Relationship between A level and UKCAT results (Sandra Nicholson, Queen Mary University of London)			

2009 First Research Outputs				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**	**Section 5 (NACE, MOJAC, MEARS, SA12)**
40 scored items **4 unscored items** **22 minutes**	36 scored items 4 unscored items 22 minutes	60 scored items 5 unscored items 16 minutes	26 scored items 30 minutes	 30 minutes
• **Decision analysis scores increased, impacting total score. Two new scenarios introduced; one did not perform as expected.** • **Time constraints most significant in quantitative reasoning; removed one set of items. To explore reducing time in Section 5.**				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
Jan – March 2009	UKCAT COO appointed. Registration with Information Commissioner completed.		Research data moved to University of Oxford. Some universities not providing progression data mainly due to student consent issues.	
April – June 2009	COO appointed as Company Secretary. Registered office moved to Nottingham.		Four papers: Resit Data; Critical Thinking A level and UKCAT; UKCAT and Language Fluency; Influence of Arts and Sciences at A-level.	
July – Sept 2009	Research Chair Jon Dowell appointed. University of Durham admitted. Charitable Status approved.	Some candidates had test scores adjusted due to a duplicated item and a poorly performing item.		
Oct - Dec 2009			Data Privacy Document approved. Progression dataset to be collected from schools agreed. Tender proposal: UKCAT's WP impact.	
Dec 2009	Presentations to Consortium Meeting • 'The UKCAT as a predictor of performance in early MBBS examinations', Dr Philip Bradley (Newcastle University)			

	• ‘A comparison of A-level and UKCAT performance in students applying to UK Medical and Dental Schools in 2006’, Dr Sandra Nicholson (QMUL), Professor David James (University of Nottingham), Dr Janet Yates (University of Nottingham) • Scottish PQA Follow up study - preliminary findings, Dr Jon Dowell (Dundee University)

2010 First Candidate Survey				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**	**Section 5 (ITQ/IVQ, MEARS, SA12)**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 23 minutes	60 scored items 5 unscored items 16 minutes	28 scored items 32 minutes	 27 minutes
• **Test reliabilities continued to increase. Speededness (the ability of candidates to complete the subtest within the allocated time) in quantitative reasoning had reduced following changes made to the section. The number of trialled items had significantly increased, and it was now possible to increase screening criteria for items to increase quality.** • **As sufficient data was available for research on the behavioural tests, it was agreed to remove the subtest in 2011 whilst considering alternatives.**				
PUBLISHED RESEARCH SUPPORTED BY UKCAT				
James, D., J. Yates, and S. Nicholson, Comparison of A level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study (James, Yates et al. 2010)				
	DELIVERING THE TEST/CANDIDATE EXPERIENCE		**RESEARCH**	
Jan - Mar	Second practice test available on website. UKCAT twitter account created. Bursary entitlement extended to include all in receipt of EMA and those in receipt of a full maintenance grant. Agreed to investigate costs of creating an 'Official Guide'.			
April – June	Key messages from candidate survey 2009: • 58% candidates felt the test to be useful but only 45% agreed that the test looked at the right attributes • Candidates did not report that their test result strongly influenced their university choices. • Candidates wanted more information about how the test was used in selection.		Call for tenders on research around UKCAT and WP; contract awarded to the University of Durham. Lead appointed to work on predictive validity longitudinal study.	

	• 52% candidates found out about the test from university websites (schools and colleges 28%) • 32% reported their school college was not at all informed about the test • 67% had purchased books to assist with preparation; 10% had attended a preparation course • 61% reported their test experience as good/very good. Actions arising from the survey included: • Engaging with the Consortium regarding information for applicants • Reviewing communications with schools/colleges • Reviewing website and preparation materials/advice • Discussions with Pearson VUE regarding test centre experience and customer services.	
June	Presentations to Consortium Meeting: • Selecting future doctors: What do we expect of our medical student selection procedures? (Sandra Nicholson, QMUL)	
Oct – Dec	Agreed to investigate creation of an on-line bursary system.	University of Dundee appointed to support Research Database.
Dec	Presentation at Consortium Meeting: • Developments in Computer Based Testing (Pearson Vue) • Current issues in admissions testing in the UK - the view from SPA (Supporting Professionalism in Admissions)	

2011 Behavioural Test Removed			
TEST CONTENT AND DEVELOPMENT			
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 23 minutes	60 scored items 5 unscored items 16 minutes	28 scored items 32 minutes
• **Analysis suggested action needed to make Quantitative Reasoning and Decision Analysis more discriminative amongst more able candidates.** • **Alternative approaches to the Decision Analysis subtest were considered but no change agreed for 2012.**			
RESEARCH SUPPORTED BY UKCAT			
Adam, J., J. Dowell, and R. Greatrix, Use of UKCAT scores in student selection by UK medical schools (Adam, Dowell et al. 2011)			
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**
Jan - Mar 2011	New Contract awarded to Pearson VUE.	Candidate views on test preparation had been obtained in a 2010 survey. Key messages: • 44% find out about the test from university websites (24% from schools/colleges) • Advice from schools/colleges good or OK (56%) • Books considered the most helpful preparation resource (40%) (UKCAT Practice tests 31%, preparation course 4%) • A few weeks preparation was sufficient (60%) • Timing rather than difficulty the major challenge (71%) Survey was to be used to improve marketing and preparation advice with specific comments to Pearson VUE regarding the testing experience.	Contract agreed with University of Dundee Health Informatics Centre (HIC) to support the management of the Research Database.

April – June 2011		'Fit to test' policy to be further emphasised on the website. On-line bursary application system introduced.	Basic analysis of data by HIC to include test/retest reliabilities. Durham to explore links between behavioural test and professionalism. HIC to explore correlations between behavioural test and Multiple Mini Interviews. Hull York Medical School to develop a common OSCE as an outcome measure in longitudinal studies.
Oct – Dec 2011		Issue with presentation of an abstract reasoning item. Late delivery of results created issues for some universities. Possibility of allowing candidates a booklet of white boards considered.	UKCAT to meet with the GMC to discuss database linkage. Data transferred to support the longitudinal study. UKCAT Board agreed to fund schools to support provision of progression data for the UKCAT Research Database.
Dec 2011	Presentations at Consortium Meeting: • UKCAT and Widening Participation – Paul Tiffin, University of Durham • UKCAT Database Update – Duncan Heather, Dundee Health Informatics Centre		

2012 Situational Judgement Test Trialled				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**	**Situational Judgement**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 23 minutes	60 scored items 5 unscored items 16 minutes	26 scored items 32 minutes	 65 (approx.) unscored items 27 minutes
• **Trial of confidence ratings within Decision Analysis agreed for 2013 with 2 minutes added to this subtest.** • **New systems allowed for a further increase in pre-testing allowing for increase in the statistical criterion applied.** • **Analysis supported an increase in standalone items in Quantitative Reasoning together with an increase in multiple choice items in Verbal Reasoning. New item types to be pre-tested in Abstract Reasoning. In Decision Analysis, additional items to be pre-tested and an additional scenario introduced.** • **Analysis of standard and extended test takers suggested performance differences had reduced following timing adjustments.** • **SJT trialled; it was agreed not to use scaled scores but to adopt a four band system of reporting.**				
RESEARCH SUPPORTED BY UKCAT				
Tiffin, P., J. Dowell, and J. McLachlan, Widening access to UK medical education for under-represented socioeconomic groups: modelling the impact of the UKCAT in the 2009 cohort (Tiffin, Dowell et al. 2012)				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
Jan - Mar 2012	Browne Jacobson appointed as solicitors.	Additional criteria introduced for bursary eligibility following the removal of the EMA in England. A web based Official Guide was developed which candidates could download for free.	The Chair had met with the GMC to discuss the possibility of linking UKCAT data, to data collected at entry to the profession and beyond. Priorities for future research were identified. Draft paper 'Comparison of the sensitivity of the UKCAT and A Levels to sociodemographic characteristics: a national study' considered by members (Tiffin, McLachlan et al. 2014). Professor Chris Dewberry presented findings from the UKCAT12 (Longitudinal Study) (McManus, Dewberry et	

			al. 2013), reporting that the test had a small, significant positive correlation with medical school performance during the first year of study. A paper from Peninsula Medical School had reported local findings regarding preparation effects on test performance (Lambe, Waters et al. 2012). It was agreed to invite the author to draft a proposal for a more substantive study in this area.
June 2012	Presentations to Consortium Meeting: • UKCAT and Widening Participation – Implications of Recent Research (Jon Dowell, University of Dundee) • Use of Situational Judgement Tests in the UKCAT (Work Psychology Group)		
Oct – Dec 2012		Booking and rescheduling deadline to be introduced to address capacity issues. Contingency day to allow candidates impacted in the final days of testing to re-test.	Noted a potential collaboration with the GMC and Health Education England to study the performance of the first cohort of postgraduate students in their foundation year assessments linked to UKCAT performance. Dr Paul Lambe (Plymouth University) reported findings from a survey looking at preparation effects on performance (Lambe, Greatrix et al. 2016).

2013 First Major Predictive Validity Study				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**	**Situational Judgement**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 23 minutes	50 scored items 5 unscored items 14 minutes	26 scored items 2 unscored items Confidence Rating 34 minutes	66 scored items 5 unscored items 27 minutes
• **Two Decision Analysis scenarios administered. Performance in one scenario different from anticipated, requiring re-scaling and adjustments to a small number of scores. Mean average for this subtest shifted upwards, impacting total scores and creating operational issues for universities.** • **Analysis of confidence rating in Decision Analysis showed reasonable distribution. Agreed to continue trialling these in 2014.** • **New item types trialled in abstract reasoning had performed well and it was agreed to continue pre-testing in 2014. The new 4-option multiple choice items in verbal reasoning discriminated better and it was agreed to increase the number of these items in the subtest.**				
RESEARCH SUPPORTED BY UKCAT				
Tiffin, P., L. Webster, and J. McLachlan, Forecasting the impact of the mode of use of the UKCAT on medical school entrant demographics (Tiffin, Webster et al. 2013) **McManus, I., et al., 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 (McManus, Dewberry et al. 2013)** **McManus, I., et al., Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: meta-regression of six UK longitudinal studies (McManus, Dewberry et al. 2013)**				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
Jan - Mar 2013		Research from the University of Plymouth regarding test preparation	Information Commissioner's advice on data protection act and de-identified data used to inform the UKCAT Data Privacy Statement. The	

		reviewed to provide key messages to candidates.	scope of potential research widened to include 'research related to admissions to medicine and dentistry that relates to the core objectives of UKCAT'. SJT Concurrent Validity Study proposal being developed. HIC had undertaken analysis regarding trends in socio economic status. Options considered to support delivery of research objectives including the need for support from an academic unit.
April - June 2013	Noted contents of the MSC Selecting for Excellence (Medical Schools Council 2014) report.		Research meeting focussed on the management of and development of the UKCAT Research Database. Areas of discussion included: MSC intending to create a Medical Schools Outcomes Research database to support evaluating medical school selection and education at UG and PG levels. Approved a pilot to trial matching of UKCAT and UKFPO data. The UKCAT-12 longitudinal study paper (McManus, Dewberry et al. 2013) presented for Board approval.
July - Sept 2013		A small number of candidates had decision analysis scores rescaled due to a scoring issue.	
Oct - Dec 2013	The University of Leeds, Imperial College London and	Two significant security issues resulted in additional reviews of booking data to mitigate this risk.	The Chair of the Research Panel appointed to develop documentation for the UK Medical

	Brighton and Sussex Medical School left the Consortium, intending to use the Biomedical Admission Test (BMAT) in future selection rounds.	Official Guide printed in hardcopy and provided to university WP teams. Four candidate focus groups had taken place, in addition to visits to WP teams. Recommendations from the focus groups included: • Develop the candidate toolkit to help schools/colleges better advise their students. • Increase prominence of bursary information on the website. • Amend preparation advice and improve practice materials in light of comments • Obtain advice and tips from actual candidates • Communicate with candidates prior to test day • Increase social media presence The website was being redesigned to better direct candidates. Preparation toolkit created incorporating mobile app, official guide, practice tests and candidate video.	Education Database (UKMED). The Board endorsed involvement in this project. Approved 'linked cohort study' using matched UKCAT and Foundation Year data to inform the UKMED development. Award for the research capacity building contract to the University of Durham to undertake a range of studies including a Predictive Validity Study, a Non-Cognitive Data Study and to investigate the use of Dynamic Testing. University of Aberdeen requested to submit proposals to undertake work around Widening Participation in admissions.
Dec 2013	Presentations at Consortium Meeting: • The UKCAT-12 study: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of twelve UK medical schools (Chris McManus, UCL)		

	• Findings from the 2012 Post UKCAT survey (Paul Lambe, University of Plymouth) • Predicting performance in the clinical years (Jen Cleland, University of Aberdeen) • Contextual Data in Medical Admissions (Kathryn Steven, University of St Andrews, Jon Dowell, University of Dundee) • Performance in repeat testing (Jon Dowell, University of Dundee)

2014				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**	**Situational Judgement**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 25 minutes	50 scored items 5 unscored items 14 minutes	26 scored items 2 unscored items Confidence Rating 32 minutes	62 scored items 5 unscored items 27 minutes
• **Time reduction in Decision Analysis resulted in slight increase in candidates not reaching end. Increase in time for Quantitative Reasoning had reduced speededness (75% of candidates reaching the final item compared to 71% in 2013).** • **New item types trialled in abstract reasoning in 2013 and 2014 to be used operationally in 2015.** • **Standalone items had been trialled in quantitative reasoning and had performed better than existing testlets. It was agreed to investigate incorporating these into the test in 2015.** • **Analysis of confidence ratings suggested they measured something different to the cognitive tests. However, it remained difficult to see how the ratings could be used operationally. It was agreed to continue use in 2015 whilst this was reviewed.** • **Distribution of the 2014 SJT results not as expected. This had been caused by incorrect scoring and results were withdrawn. Alternative options to pre-equating considered but delivering results immediately to candidates of equal importance. Test Development Group tasked with implementing a solution that ensured 2015 results were delivered to existing timelines.**				
RESEARCH SUPPORTED BY UKCAT				
McAndrew, R. and R. Greatrix, The UKCAT test: developments, research and its use by dental schools in the UK (McAndrew and Greatrix 2014) **Tiffin, P.A., et al., Comparison of the sensitivity of UKCAT and A Levels to sociodemographic characteristics: a national study (Tiffin, McLachlan et al. 2014)**				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
Jan - Mar 2014		Universal Credit added to the list of eligibility criteria for bursary applications.	The GMC took ownership of the UKMED development. To provide data to support the project, UKCAT to review data security and governance arrangement.	

April - June 2014			Agreed to support projects at the University of Aberdeen focussing on widening participation. UCAS had changed governance requirements requiring UKCAT to enter into agreements with universities to obtain the annual dataset. Agreed to support a PhD studentship at the University of Durham focussed on 'Evaluating and extending methods for estimating construct-level predictive validity: Implications for the UKCAT for student selection.'
July - Sept 2014		Scaling issue led to withdrawal of the SJT scores, affecting schools intending to use the test in selection. Most candidates had not been impacted, but there was some criticism around time taken to prepare for the sub-test.	
Oct - Dec 2014	Financial Advisers appointed. Nigel Siesage appointed Chair.	Photo capture of candidates introduced to increase security.	Agreed to review the contents of the Selecting for Excellence Report (Medical Schools Council 2014) to identify research study opportunities arising from recommendations.
Dec 2014	Presentation at Consortium Meeting • Similarities and Trends in Medical School Admission Tests (Steve Barkley) • Comparisons of MMI, PS and UKCAT scores (University of Keele) • Predictive validity of the UK Clinical Aptitude Test: Preliminary findings from a national study (Paul Tiffin, University of Durham) • Ability of the UKCAT to predict performance in years 3 and 4 of medical school: a national study (Paul Tiffin, University of Durham)		

2015				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Analysis**	**Situational Judgment**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 25 minutes	50 scored items 5 unscored items 14 minutes	26 scored items 2 unscored items Confidence Rating 32 minutes	63 scored items 5 unscored items 27 minutes

- **The selection for excellence group (Medical Schools Council 2014) had identified traits to be tested within medical selection. Existing test content already tested some of these traits; other traits more suited for consideration elsewhere in selection.**
- **Decision Making to be trialled in 2016 (replacing Decision Analysis) to test problem solving, dealing with uncertainty, managing risk, dealing effectively with problems. Selecting for Excellence report had driven development. Results not to be reported in first year. Quantitative reasoning items to be included as anchor items to assess candidate motivation/preparedness.**
- **Quantitative reasoning remained the most speeded subtest although new 'standalone' items had not made this worse.**
- **Abstract Reasoning new item types performed well. Agreed to introduce two item types noting potential impact on timing.**

	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**
Jan - Mar 2015	University of Birmingham and University of Liverpool joined the Consortium.	Free school meals accepted as bursary evidence. Bursary entitlement to be passed to universities as an early WP indicator. End of testing statistics to be released earlier to assist candidates in making university choices. Agreed to explore provision interim score data.	Agreed three proposals for studies linked to the Selecting for Excellence Report. My proposal to enrol on a research degree approved.
July 2015	Presentations to Consortium Meeting: • Predictive Validity of the UKCAT (Paul Tiffin, University of Durham)		

July - Sept 2015	University of Bristol and University of Buckingham joined the Consortium.		Five tenders received for work on Graduate Applicants and WP; contract awarded to Aberdeen. Proposal from University of Newcastle to look at WP in dentistry. Proposal to UKMED relationships between UKCAT and fitness to practice outcomes. Leaflet produced for Consortium members summarising existing research activities. RWP identified study priorities.
Oct - Dec 2015	Poznan University became an associate member of the Consortium.	Identified two forged score reports; options to improve report security considered. Following pilot agreed to collect university choices from candidates in 2016, replacing the current arrangement with UCAS. A survey of Bursary candidates from 2014 testing to review impact of improved communications. Key messages: • only 37% on a WP programme • these candidates less likely to receive guidance from school/college and rely on advice from friends • Majority using official practice tests and questions, supporting the development of the toolkit	
Dec 2015	Presentations to Consortium Meeting:		

	• New Developments in Situational Judgement Tests (Professor Janneke Oostrom, University of Amsterdam) • Report on the outcome of the evaluation of the relationship between medical school admissions data, demographics and UKCAT performance with UKFPO performance (Professor Jen Cleland, University of Aberdeen) • Dynamic Testing (Professor Jens Beckmann, Durham University)

2016 Decision Analysis removed/Decision Making Trialled				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Making**	**Situational Judgement**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 25 minutes	50 scored items 5 unscored items 14 minutes	26 trial (scored) items 3 trial (unscored) items 32 minutes	64 scored items 5 unscored items 27 minutes
• **Quantitative Reasoning remained most speeded subtest. New Abstract Reasoning item types had minimal speededness.** • **It was agreed overtime to address scale score drift in Quantitative Reasoning.** • **No changes to Decision Making despite concerns about candidate motivation. Scale scores anticipated to shift upwards in 2017.**				
RESEARCH SUPPORTED BY UKCAT				
Fair access to medicine? Retrospective analysis of UK medical schools application data 2009-2012 using three measures of socioeconomic status BMC Kathryn Steven, Jon Dowell, Cathy Jackson and Bruce Guthrie (Steven, Dowell et al. 2016) **Longitudinal assessment of the impact of the use of the UK clinical aptitude test for medical student selection Jonathan Mathers, Alice Sitch, Jayne Parry (Mathers, Sitch et al. 2016)** **Does the UKCAT predict performance on exit from medical school? A national cohort study R K MacKenzie et.al., (MacKenzie, Cleland et al. 2016)** **Predictive validity of the UKCAT for medical school undergraduate performance: a national prospective cohort study Paul A. Tiffin, Lazaro M. Mwandigha, et al. (Tiffin, Mwandigha et al. 2016)** **Do personality traits assessed on medical school admission predict exit performance? A UK-wide longitudinal cohort study MacKenzie, R.K., Dowell, J., Ayansina, D. et al. (MacKenzie, Dowell et al. 2017)**				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
Jan - Mar 2016		Universities to take on the role of determining exemptions from the test.	Agreed tender to evaluate SJT. Schools to be encouraged to identify assessment outcomes to investigate.	
June 2016	Contract awarded to Pearson VUE after tender.	Key messages from 2015 survey:		

		• 36% candidates finding out about the test from school/college (22% university websites) • Websites used for UKCAT included Student Room (43%) and preparation companies (most used, Kaplan 30%) • 42% not using social media for careers; the most popular site Facebook • Those using social media for UKCAT favoured Twitter and YouTube. • Candidates that registered late did not perform as well as early registrants • 90% used UKCAT practice tests; 75% using books and 20% paid courses • Average reported preparation time 26 hours; 29 hours for higher scorers.	
July 2016	Presentations to Consortium Meeting: • Developing the Situational Judgement Test (Stuart Martin from the Work Psychology Group) • Use of Contextual Data in Medical School Admissions (Sandra Nicholson, Queen Mary, University of London)		
July - Sept 2016		Interim test statistics delivered to candidates for the first time.	Further research priorities identified.
Oct - Dec 2016	American University of the Caribbean became an associate member of the Consortium.	Agreed to explore version of the test with rest breaks between sections. Course choices collected from candidates. Issues with data quality caused issues with some candidates entering incorrect PIDs. Further exploration required.	A proposal around widening access and dental selection was considered. Agreement to collaborate with Dr Kath Woolf (UCL) around the UK Medical Applicant Cohort Study (UKMACS).

		Standalone SJT being created to be used by a graduate entry medicine programme.	Tender to undertake further evaluation of the SJT was awarded to Queen Mary University of London.
Dec 2016	Presentations to Consortium Meeting: • Exploring whether graduate applicants and graduate entry programmes widen access to medical and dental education (Prof Jen Cleland, University of Aberdeen)		

2017 Decision Making Introduced				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Making**	**SJT**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 25 minutes	50 scored items 5 unscored items 14 minutes	26 scored items 3 unscored items 32 minutes	63 scored items 6 unscored items 27 minutes
• **Agreed to rescale quantitative reasoning scale scores downwards in 2018 recognising upwards drift in scores in recent years.** • **Decision Making performed as expected; some upward drift in scale scores. 2017 test results to set the norm for future scaling.** • **New item type was to be trialled in the SJT in 2018. SJT scores bands revised for 2018 based on actual distributions in 2017.**				
RESEARCH SUPPORTED BY UKCAT				
The relationship between school type and academic performance at medical school: a national, multi-cohort study Kumwenda, B., et al., (Kumwenda, Cleland et al. 2017) **Do personality traits assessed on medical school admission predict exit performance? A UK-wide longitudinal cohort study MacKenzie, R.K., et al., (MacKenzie, Dowell et al. 2017)** **The Predictive Validity of a Text-Based Situational Judgment Test in Undergraduate Medical and Dental School Admissions Patterson, F., et al., (Patterson, Cousans et al. 2017)** **Evaluating and extending statistical methods for estimating the construct-level predictive validity of selection tests Mwandigha, L.M., (Mwandigha 2017)**				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
March 2017	Discussions regarding collaboration with a Consortium of Australian and New Zealand universities using UMAT (Universities Medical Admissions Test). St George's University (Grenada) became an associate member of the Consortium.	Separate test cycle delivered for Humanitas University (Milan).	Application to undertake a study focussing on contextual data and dental selection (University of Cardiff) was approved. Proposal to focus on UKCAT coaching effects approved.	
April - June 2017	Discussions ongoing with the Council of Deans for Health regarding the potential	Re-design to UKCAT website.		

	of developing an admission test to be used for selection to the allied health professions. Non-Disclosure agreement signed with UMAT universities.		
June 2017	Presentations to Consortium Meeting: • Using UKMED to predict Fitness to Practice declarations at provisional registration in UK medical undergraduates (Lewis Paton, University of York) • The role of a secondary school's performance in predicting an entrant's academic outcomes at medical school- or 'How low do we go when 'discounting' grade offers to candidates from disadvantaged schools? (Lazaro Mwakesi Mwandigha, University of York)		
July 2017		Two new version of the test made available to candidates: UKCATSEN50 (+50% time); UKCATSA (5-minute rest breaks between subtests).	
Oct – Dec 2017	Anglia Ruskin University and Aston University joined the Consortium.	UCAS re-contracted to support matching of candidates to university choices. UKCATSENSA (+25% time and 5-minute rest breaks) introduced.	Further research priorities identified.
Dec 2017	Presentations to Consortium Meeting: • Selective Admission: What Are the Trends? (Dr Anna Mountford-Zimdars, University of Exeter) • Gatekeepers or facilitators? Engaging schoolteachers as advocates for widening access to medicine. (Kirsty Alexander, University of Aberdeen)		

2018				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Making**	**SJT**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 25 minutes	50 scored items 5 unscored items 14 minutes	26 scored items 3 unscored items 32 minutes	63 scored items 6 unscored items 27 minutes
• **The rescaling of quantitative reasoning had worked as anticipated. No further changes to scaling this subtest to be made in 2019 to review stability.** • **Changes made to the SJT blueprint in advance of 2019 testing introducing new item types operationally.**				
RESEARCH SUPPORTED BY UKCAT				
What is the effect of secondary (high) schooling on subsequent medical school performance? A national, UK-based, cohort study Mwandigha, L.M., et al., (Mwandigha, Tiffin et al. 2018) **Predictors of fitness to practise declarations in UK medical undergraduates Paton, L.W., et al., (Paton, Tiffin et al. 2018)** **The ability of 'non-cognitive' traits to predict undergraduate performance in medical schools: A national linkage study Finn, G.M., et al., (Finn, Mwandigha et al. 2018)** **Are efforts to attract graduate applicants to UK medical schools effective in increasing the participation of under-represented socioeconomic groups? A national cohort study Kumwenda, B., et al., (Kumwenda, Cleland et al. 2018)**				
	GOVERNANCE		**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	
March 2018	University of Kurdistan Hewler admitted as an associate member. The Board considered alternative options to support the delivery of the UKCAT in Australia and New Zealand.		Agreed to explore process to centralise approval for the UKCATSEN. Reviewed GMC guidance to universities regarding reasonable adjustments and aptitude testing. Further practice test to be available for candidates preparing for 2018. Separate test window delivered for Humanitas University.	

July 2018	Edge Hill University and the University of Sunderland joined the Consortium. UKCAT to be delivered in Australia and New Zealand in 2019. Arrangement to be managed by a licence agreement with ANZ allowing them to use the test, a collaboration agreement between UKCAT and ANZ (setting out contractual and financial arrangements) and an extension to the current contract with Pearson VUE (to include additional services). ANZ to have a separate contract with Pearson VUE covering the delivery of the test in ANZ. UKCAT privacy policy updated to reflect GDPR requirements.	
July 2018	Presentations to Consortium Meeting: • Health and Disability (Clare Owen, Medical Schools Council) • Conversion Rates and Additional Student Numbers (Austen Spruce, University of Birmingham)	
Oct to Dec 2018	Contracts with ANZ signed and the first meeting of the UCAT Joint Committee (UJC) had taken place.	A percentile look up facility introduced for candidates to check how their performance rated against all candidates. Pearson VUE to outsource customer services from 2019. Web based score reports to be introduced from 2019 allowing candidates to access results through their Pearson VUE account. Test Delivery Group given an overview of the Pearson Vue on-line proctoring system (OnVUE).
Dec 2018	Presentations to Consortium Meeting: • Professionalism: More than a word (Marina Sawdon, University of Sunderland) • Widening Access and Workforce Planning (Ben Kumwenda, University of Aberdeen)	

2019 Delivery in Australia/New Zealand				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Making**	**Situational Judgement**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 25 minutes	50 scored items 5 unscored items 14 minutes	26 scored items 3 unscored items 32 minutes	63 scored items 6 unscored items 27 minutes
• **Agreed to make minor downwards adjustment to scale scores in both abstract reasoning and quantitative reasoning.** • **New item types introduced into operationally into the Situational Judgement Test.** • **Delivery of the test in Australia and New Zealand for the first time. Test content reviewed with a view to further globalisation of content.**				
RESEARCH SUPPORTED BY UKCAT				
Relationship between sociodemographic factors and specialty destination of UK trainee doctors: a national cohort study. Kumwenda, B., et al., BMJ open, 2019 (Kumwenda, Cleland et al. 2019) **Does 'online confidence' predict application success and later academic performance in medical school? A UK-based national cohort study Tiffin, P.A. and L.W. Paton, (Tiffin and Paton 2019)**				

	GOVERNANCE	DELIVERING THE TEST/CANDIDATE EXPERIENCE	RESEARCH
Jan - March 2019	Change of name to UCAT Consortium for day-to-day business. UKCAT Consortium remained the legal name of the organisation. New Data Privacy Policy agreed.	New systems in place to deal with access arrangements including on-line approval of eligibility to sit the UKCATSEN. Registration in ANZ opened. Arrangements to support those applying to UK and ANZ universities. New website launched. Additional test cycle for Camillus University (Rome) to be delivered.	
June 2019	Pearson Vue attended a Board meeting to discuss future developments to the test.		Contract agreed with the University of York

			for ongoing research support.
June 2019	Presentations to Consortium Meeting: • Diversifying the medical workforce and issues of widening participation (Sandra Nicholson, Queen Mary University of London) • UK Medical Applicant Cohort Study (Kathryn Woolf, UCL)		
Sept 2019	Conflicts of Interest Policy agreed. Test fees lowered in 2020 recognising additional income streams from ANZ. Critical Incident Plan approved.	Significant increase in applications for access arrangements and increased complexity of cases noted.	
Oct - Dec 2019	Dr Amanda Hampshire appointed to Chair of Test Delivery; Professor Jayne Parry appointed as Research Lead. Amendments to company articles agreed for further discussion with Consortium. Trademark for UCAT logo to be investigated. Kent Medway Medical School joined the Consortium.	Firmer 'fit to sit' policy to be developed. Transition of customer services to HCL delayed to 2020. Plan to promote Bursary scheme agreed. Use of on-line proctoring for delivery of the test continued to be explored.	
Dec 2019	Presentations to Consortium Meeting: • Selection for Medical School: The case of the Netherlands and Maastricht University (Sanne Schreurs, Maastricht University) • An exploration of the impact of formal preparation activities on candidate performance in the UKCAT (UCAT) (Jayne Parry, University of Birmingham) • The UK Medical Applicant Cohort Study: Update (Kath Woolf, University College London) • MSC Selection Alliance Update (Clare Owen, Medical Schools Council)		

2020 COVID19 Pandemic - Test partially delivered online				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Making**	**Situational Judgement**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 25 minutes	50 scored items 5 unscored items 14 minutes	26 scored items 3 unscored items 32 minutes	63 scored items 6 unscored items 27 minutes
• **Analysis of performance differences between UK and ANZ candidates did not support any need for separate calibration of the two tests.** • **Decision to rescale Abstract Reasoning and Quantitative Reasoning put on hold due to COVID19 impact.** • **Advice provided to candidates anticipating performance differences between online and test centre candidates.** • **Candidates testing in test centres outperformed those testing online.**				
Research Supported by UKCAT				
UKCAT and medical student selection in the UK - what has changed since 2006? Rachel Greatrix and Jonathan Dowell (Greatrix and Dowell 2020)				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
March 2020	Early discussion regarding COVID-19 impact. Agreed to delay registration and testing to allow more time to assess impact. Future virtual summer Consortium meetings with face-to-face meeting in Nov/Dec.		UKCAT Research and Development Group created combining the roles of the Research Group and the Test Development Group.	
April – June 2020	Focus on the impact of the COVID-19 pandemic on testing. Principles to underpin decision making agreed. On-line proctoring to provide a contingency in the event of impact on test centre delivery being. Dual	Anticipated dual delivery and COVID19 would result in increase in mitigating circumstances. Candidates testing online would have a limited set of accommodations.		

	delivery model (online and test centre) was approved in June.	Agreed to allow candidates use of a whiteboard during online testing.	
July-September 2020		Significant increase in candidate numbers observed with 30% opting to test online. Online testing impacted by significant number of technical incidents requiring candidates to reschedule and/or resit. Customer Service impacted by COVID19 and unable to support the significant increase in issues from candidates taking the test online.	
October – December 2020		Overall candidate performance increased. 2020 testing experience led to improvements in the online product.	Discussions with ANZ regarding collaborative research to focus on qualitative research looking at widening access experience.

2021				
TEST CONTENT AND DEVELOPMENT				
Verbal Reasoning	**Quantitative Reasoning**	**Abstract Reasoning**	**Decision Making**	**Situational Judgement**
40 scored items **4 unscored items** **22 minutes**	32 scored items 4 unscored items 25 minutes	50 scored items 5 unscored items 14 minutes	26 scored items 3 unscored items 32 minutes	63 scored items 6 unscored items 27 minutes
Review of test content initiated driven by issues relating to the ability of difference tests to predict performance in medical/dental training, speededness in the test, the need to evaluate the Decision Making subtest and consider alternative traits that could be measured, the continuing utility of abstract reasoning, further development of the SJT and the desirability of reducing sub-group differences in the test.				
Research Supported by UKCAT				
Does the UKCAT predict performance in medical and dental school? A systematic review BMJ Open January 2021, Rachel Greatrix, Sandra Nicholson and Susan Anderson (Greatrix, Nicholson et al. 2021)				
	GOVERNANCE	**DELIVERING THE TEST/CANDIDATE EXPERIENCE**	**RESEARCH**	
March 2021	On behalf of the Consortium, the Board responded to the government consultation on post-qualification admissions broadly welcoming the consultation and noting the ability of the organisation to be flexible in terms of timing of test delivery. The need to clarity about timelines was highlighted. The potential challenge of shifting test timings resulting in test delivery clashing with applicant exam preparation was noted.	2021 testing to be delivered in test centres for most candidates with an approval process in place for candidates requesting to sit the test at home. Contingency planning to continue anticipating future lockdowns. Practice tests/question banks moved to the Pearson Vue test platform to give candidates a better preparation experience. The mobile app would be retired.		

	Brunel and Worcester were admitted to the Consortium.	UKCAT Office attendance at online candidate events increasing with priority given to events with a WA focus.	
September 2021	University of Chester joined the Consortium.	Majority of candidates were able to test in test centres. Customer Service impacted by pandemic. ANZ required to extend testing period into September for online testing required due to lockdowns across the two countries.	
October - December 2021	The Board approved revised organisational aims which had been included in the 2020 Annual Report.	In light of the 2021 experience customer service support to be a significant focus in 2022. Malpractice Policy to be developed with input from the Consortium.	Discussions had been initiated with the UKMACS study to see whether their data might help UKCAT understand the candidate journey better. Proposal from Pearson VUE to publish findings regarding the relative performance of test centre and online proctored candidates was approved.
November 2021	UCAT SJT Masterclass delivered for Consortium members. The event included the following presentations: • Development and introduction of the AAMC SJT (Dana Dunleavy and Rebecca Fraser, AAMC) • The UCAT SJT Journey (Fiona Patterson, Work Psychology Group)		
December 2021	Presentations to Consortium Meeting: • Future Test Development (Pearson VUE) • Demographic changes in UKCAT Candidates (Paul Tiffin, University of York)		

4.5 UKCAT Test Content

In 2021 the UKCAT comprised 5 separately timed and separately scored subtests: verbal reasoning, decision making, quantitative reasoning, abstract reasoning and the situational judgement test. Information is provided in Appendix C regarding each of these subtests including the different item types presented to candidates. Information is also provided regarding subtests removed from the test since 2006.

4.6 Descriptive Statistics

Since 2006, scaled scores have been reported for each of the cognitive subtests. Subtest scale scores range from 300 to 900 with a mean of 600 in the reference sample. At registration, UKCAT has collected demographic information from candidates on an annual basis to monitor the performance of different subgroups taking the test. Those demographic variables have in some cases changed over time.

4.6.1 Overall Candidate Numbers and scores by subtest

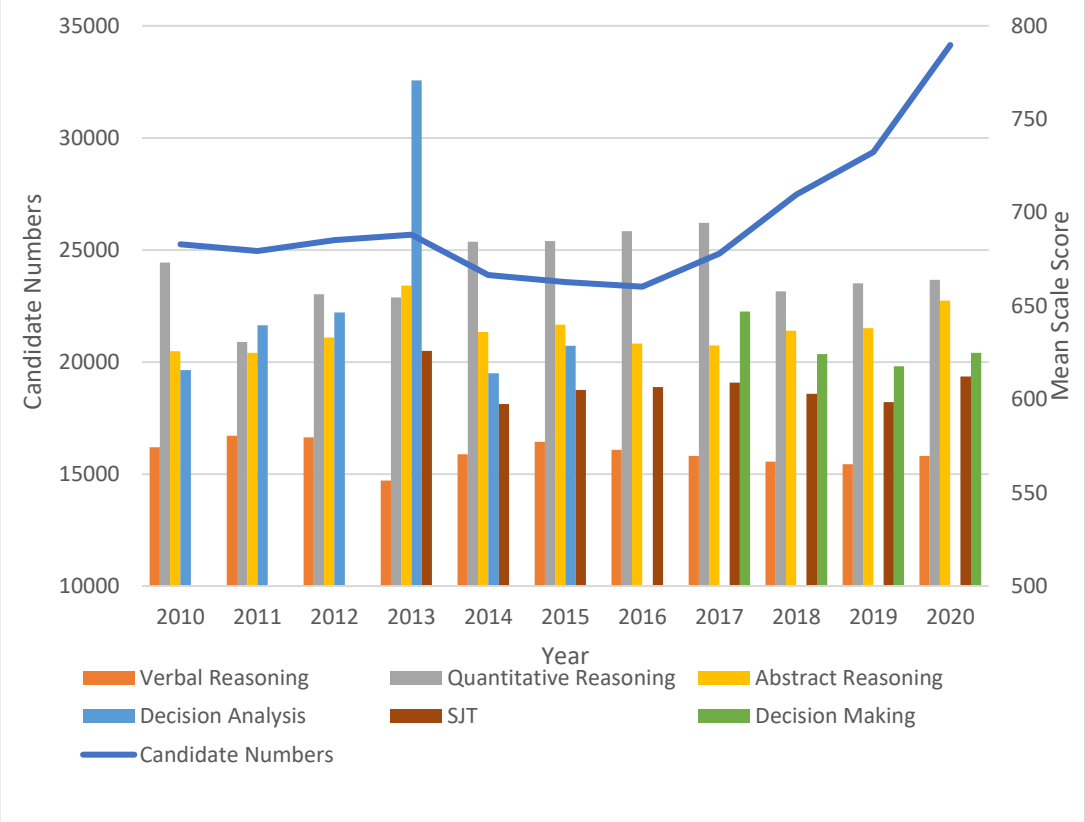


Figure 1 Candidate Numbers and Subtest Scores

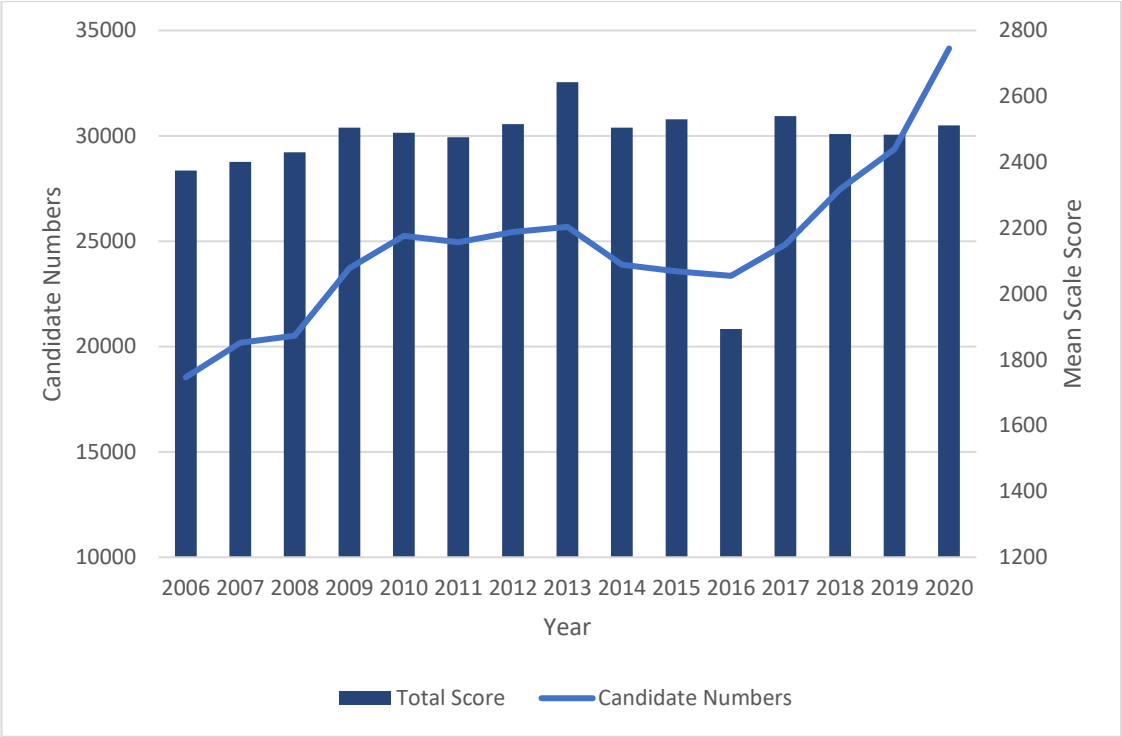


Figure 2 Candidate Numbers and Total Score

Figure 1 (from 2010 onwards) and Figure 2 provide information regarding candidate numbers alongside mean scaled subtest and total scores.

Candidate numbers broadly increased from 2006 to 2013 (from 18,540 to 25,679). The reduction in candidates in 2013 can be accounted for by the departure of three universities from the Consortium. New Consortium members joining in 2015 and 2016 did not appear to impact on the number of test takers, perhaps suggesting that many of the applicants to those new Consortium universities were already taking the test (in order to apply elsewhere). Between 2017 and 2020 large increases in candidate numbers were observed (from 24,842 to 34,144). In 2021 37,230 candidates took the test.

Mean test scores are influenced by a whole range of factors. Test content and timings have changed overtime, the candidate population has changed and perhaps most importantly increased substantially. Candidate familiarity with test content has increased overtime, with some evidence perhaps that this impacts on performance in some subtests than others.

In Figure 1 the following can be observed:

- A steady small decline in performance in verbal reasoning (from 588 in 2006 to 570 in 2020).
- Performance in quantitative reasoning has increased over time (from 597 in 2006 to 664 in 2020) to some extent impacted by changes in this subtest, including a reduction in numbers of items, increased time (e.g. 2010 and

2014) and more recently a rescaling downwards (in 2018) to bring the mean scale score more in line with the intended mean.

- A steady increase in performance in abstract reasoning (from 596 in 2006 to 653 in 2020).
- The impact of the issues associated with the Decision Analysis subtest (discussed earlier in this chapter) can be observed in significant shifts in scale scores over time. The large shift in performance in 2013 is referred to above and impacted on selection in that year. The test was withdrawn in 2015.
- Decision Making was introduced operationally in 2017 and rescaled to the candidate population in 2018.
- SJT performance declined after the first year of delivery and has been reasonably steady since then.
- The Total Test Score (Figure 2) has increased steadily overtime from 2375 in 2006 to 2511 in 2020). Significant shifts are observed in 2013 (impacted by the shift in performance in decision making) and in 2016, the trial year for Decision Making when only three subtests were scored operationally.

4.6.2 Candidate Performance by Test Taken

	Candidate Numbers							**Mean Total Score**	
	All	UKCAT		SEN		Other		**UKCAT**	**SEN**
		N	%	N	%	N	%		
2014	23884	23042	96.5%	842	3.5%	0	0.00%	**2502**	**2590**
2015	23565	22668	96.2%	897	3.8%	0	0.00%	**2529**	**2575**
2016	23359	22485	96.3%	874	3.7%	0	0.00%	**1892**	**1917**
2017	24841	23863	96.1%	949	3.8%	29	0.12%	**2537**	**2616**
2018	27469	26298	95.7%	1093	4.0%	78	0.28%	**2481**	**2584**
2019	29366	27993	95.3%	1162	4.0%	211	0.72%	**2479**	**2561**
2020	34144	32297	94.6%	1501	4.4%	346	1.01%	**2507**	**2584**

Table 2 UKCAT and UKCATSEN Candidate Numbers and Mean Total Scores

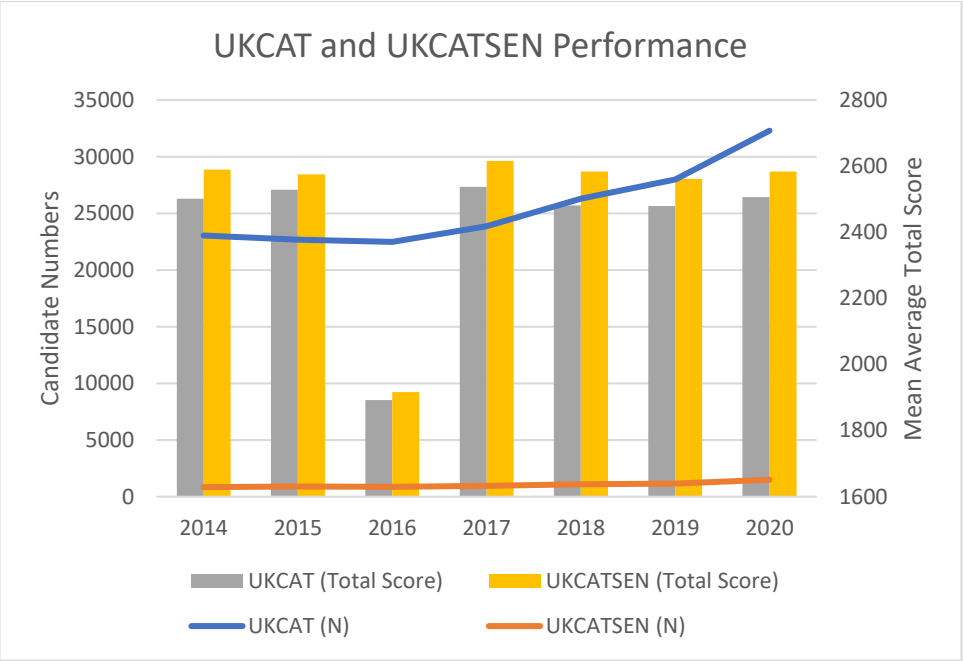


Figure 3 UKCAT and UKCATSEN Candidate Numbers and Mean Total Scores

An extended test for candidates with learning difficulties has been available since 2006. Candidates have had to evidence their eligibility to take this test which provides them with 25% additional time in all subtests. Since 2017, a number of additional versions of the test have been introduced to better meet the needs of the candidate population:

- UKCATSA/UKCATSENSA (additional 5-minute rest breaks between sections)
- UKCATSEN50 (50% additional time in all subtests)

The proportion of candidates taking extended versions of the test has increased over time from 3.5% in 2014 to 5.5% in 2020. This possibly reflects the better advertising of the availability of these tests and an increase in school and university students entitled to accommodations for examinations.

Data regarding the relative performance of these two main candidate groups is only available since 2014 and illustrated in Table 2 and Figure 3. The numbers taking the UKCATSA, UKCATSENSA and UKCATSEN50 are very small, and their results are not presented here.

UKCATSEN candidates have always outperformed UKCAT candidates (in 2020 total score by 77 scale score points) although differences in test performance have changed between years. The demographics of these two candidate groups are very different with, for example, very few international candidates take the SEN version. The performance gap between standard and SEN UK candidates is smaller than observed in Table 3 (in 2020 total score by 66 scale score points) though is still not fully understood.

Similar differences between performance in these two groups are observed in the Situational Judgment test although these may be at least partially explained by the relatively lower performance of international candidates in this subtest.

Discussions around these performance differences have focussed both on the demographics in these two groups (and the extent to which this can explain the difference) and the effect of speededness on subtest outcomes. The provision of 25% additional time across all subtests potentially advantages candidates whose main difficulties are around reading at speed.

4.6.3 Candidate Subgroups: Gender

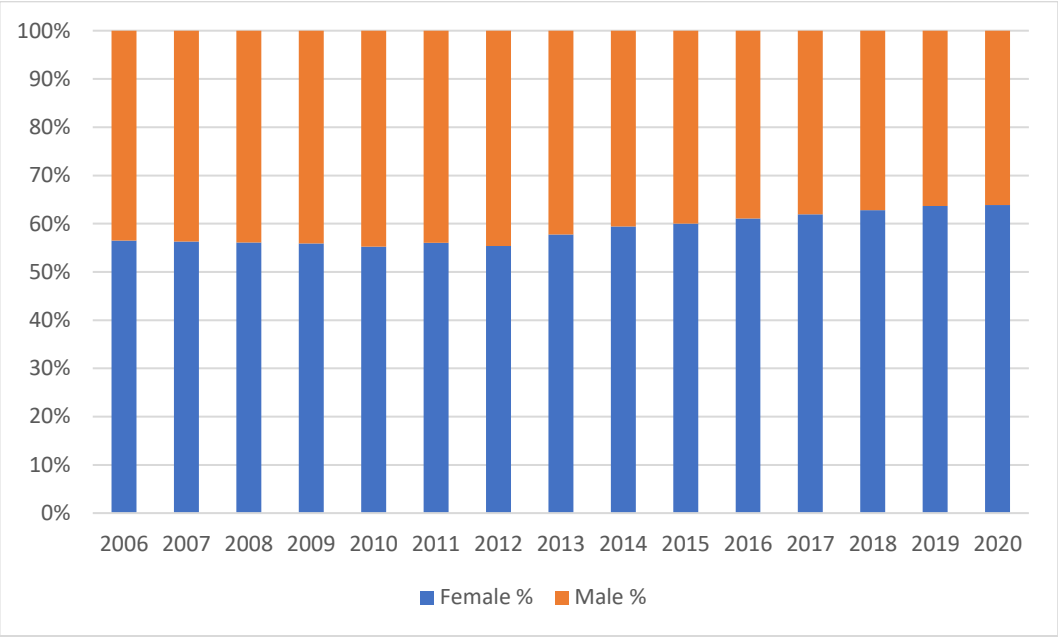


Figure 4 Candidates (%) by Gender

The proportion of female test candidates has grown over time from 56.5% in 2006 to 64% in 2020.

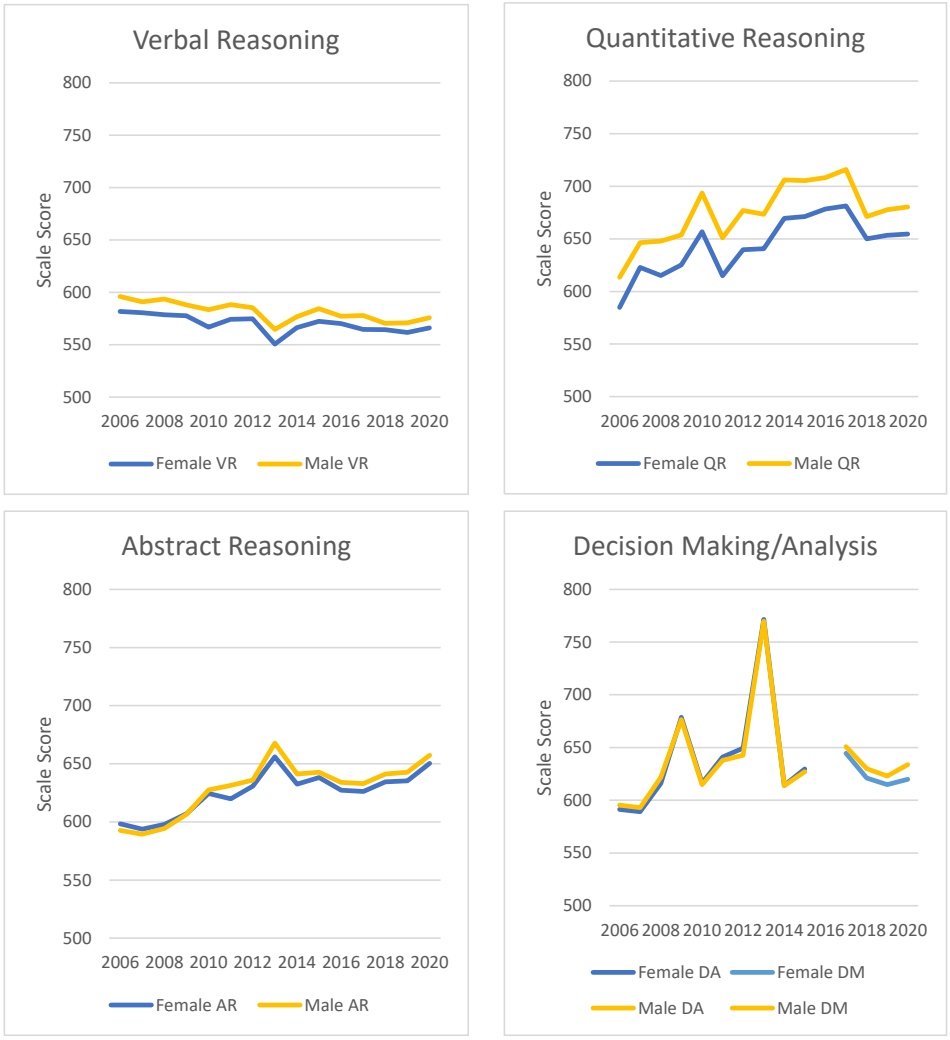


Figure 5 Gender Performance by Cognitive Subtests

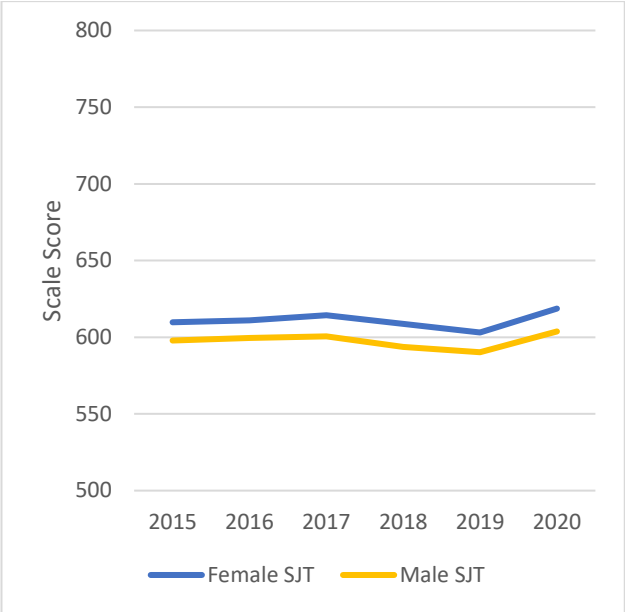


Figure 6 SJT Scores by Gender

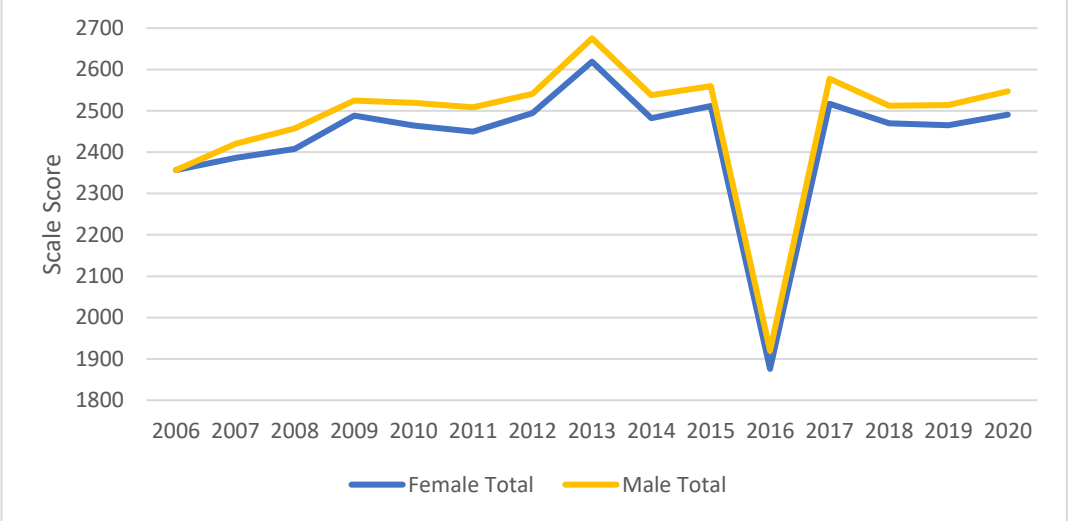


Figure 7 Total Scores by Gender

Figures 4-7 illustrate the relative performance in the test between males and females. Since 2010 males have outperformed females in all the cognitive subtests. Gender differences are most pronounced in quantitative reasoning. In 2020 the mean average difference between total scores for male and female candidates was 56.

Females outperform male candidates in the Situational Judgement test.

4.6.4 Candidate Subgroups: Socio-Economic Class

The socio-economic class (NS-SEC) is derived from occupation and employment status information. UKCAT uses a self-coding method where candidates provide information regarding the occupation, employment status and nature of the employer for their parents. This approach whilst the most convenient is almost certainly not the most accurate. The Office for National Statistics reports that in

comparisons of the self-coded and interviewer-coded five-class NS-SEC there was 75% agreement in classification. This will be further compounded by the fact that candidates are responding on behalf of their parents. SEC1-5 are defined as follows:

- SEC1 Managerial and professional
- SEC2 Intermediate
- SEC3 Small employers and own account workers
- SEC4 Lower supervisory and technical
- SEC5 Semi-routine and routine

Data is shown below from 2014 onwards as prior to this the data included non-UK candidates for whom this derivation is not appropriate.

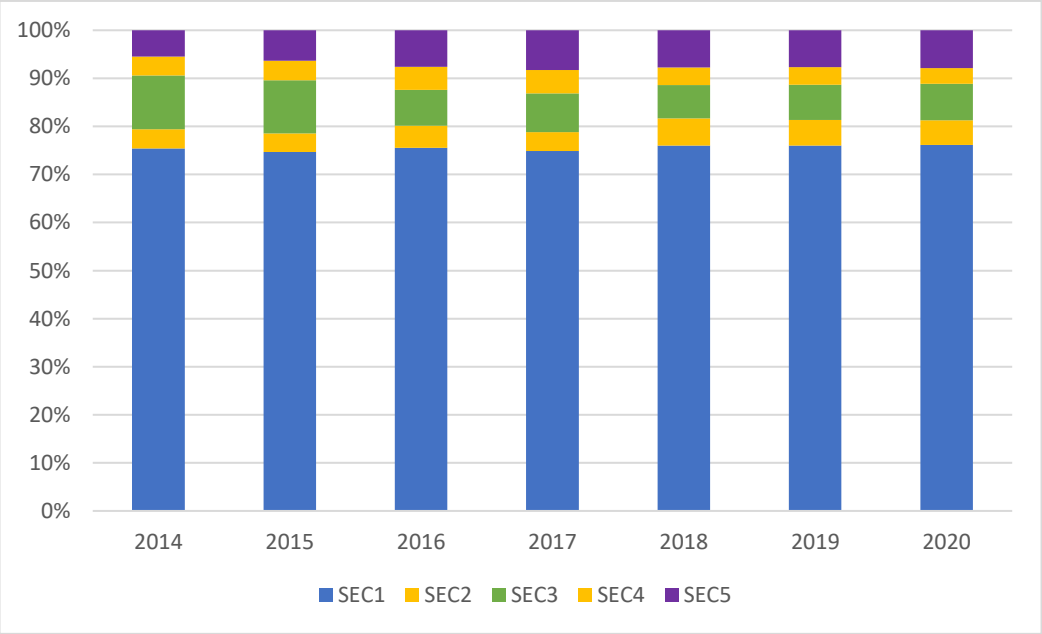


Figure 8 Proportion of Candidates by SEC

Figure 8 shows that in 2020 76.2% of candidates were from SEC1. This proportion has shifted very little from 2014 when the proportion was 75.4%. Whilst the proportion of candidates from SEC4&5 remains very low, the absolute number of candidates from these two groups has increased from 1582 in 2014 to 2651 in 2020.

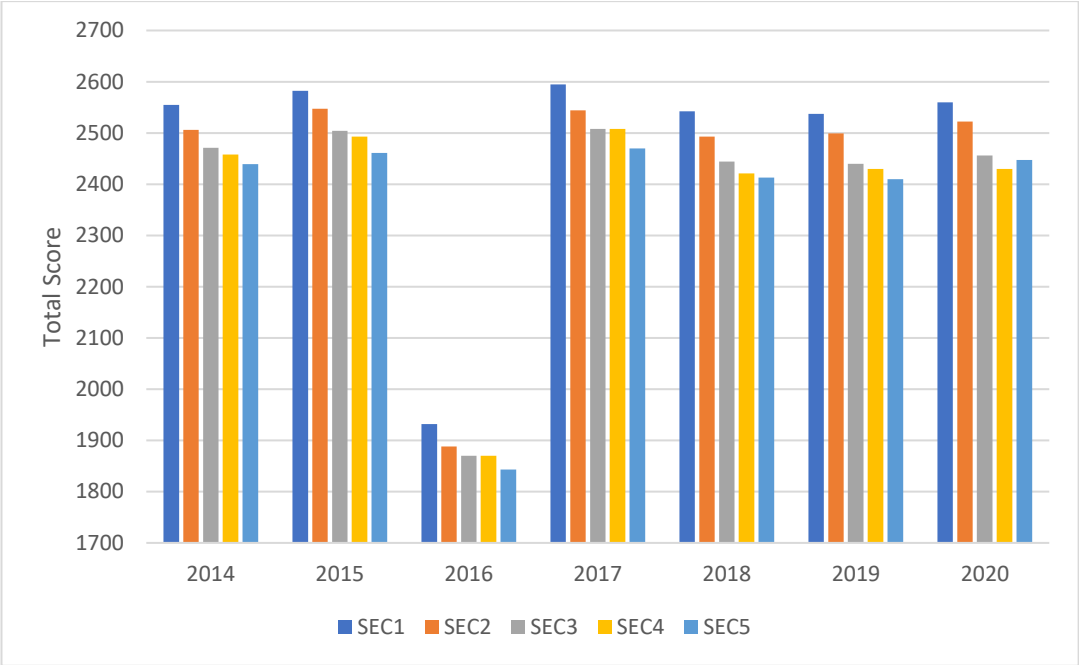


Figure 9 Mean Average Total Score by SEC

Figure 9 describes differential performance between socio-economic classes by total score. Candidates from lower socio-economic classes perform less well in the test. In 2020 those in SEC5 scored on average 113 scale points lower (total score) than candidates in SEC1.

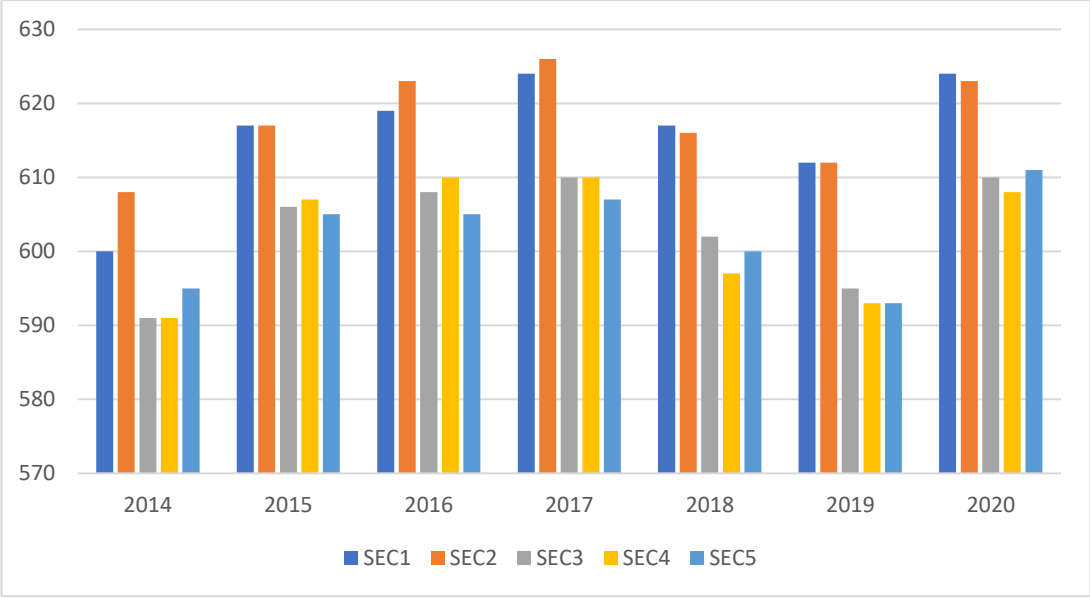


Figure 10 SJT Scale Score by SEC

Candidates in SEC1 and SEC2 also outperform other candidate in the SJT although the differences observed here are smaller than in other subtests.

4.6.5 Age and Highest Qualification

Test performance is monitored by age and its interaction with qualifications. At registration, candidates provide information as to whether their highest qualification (on entry to medical/dental school) is expected to be school leaver qualifications or degree level.

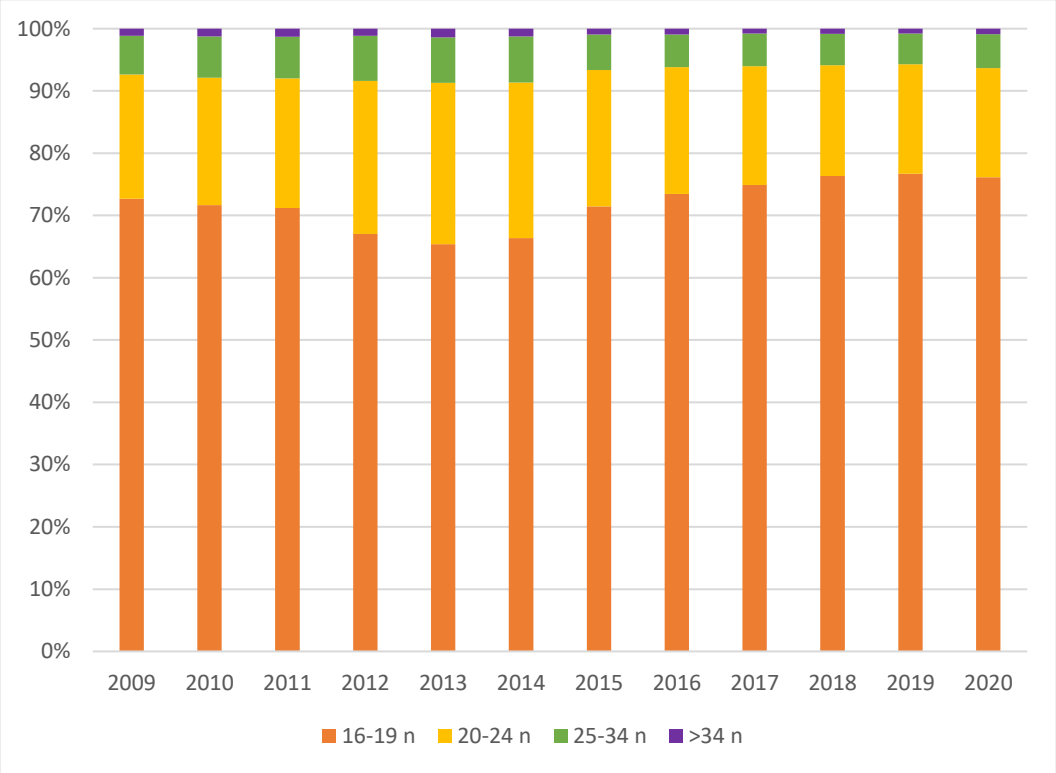


Figure 11 Distribution of candidates by age

The vast majority of candidates are in the 16-19 age range; 76% candidates in 2020.

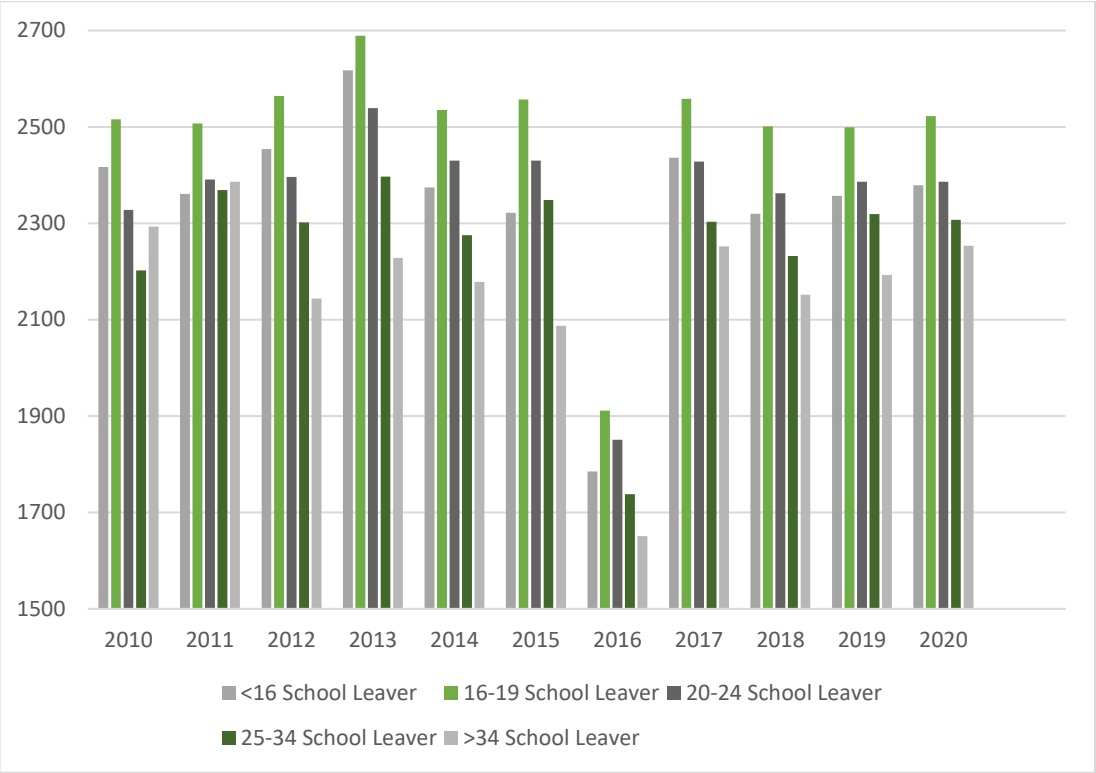


Figure 12 Mean Average Total Score by Age (highest qualification school Leaver)

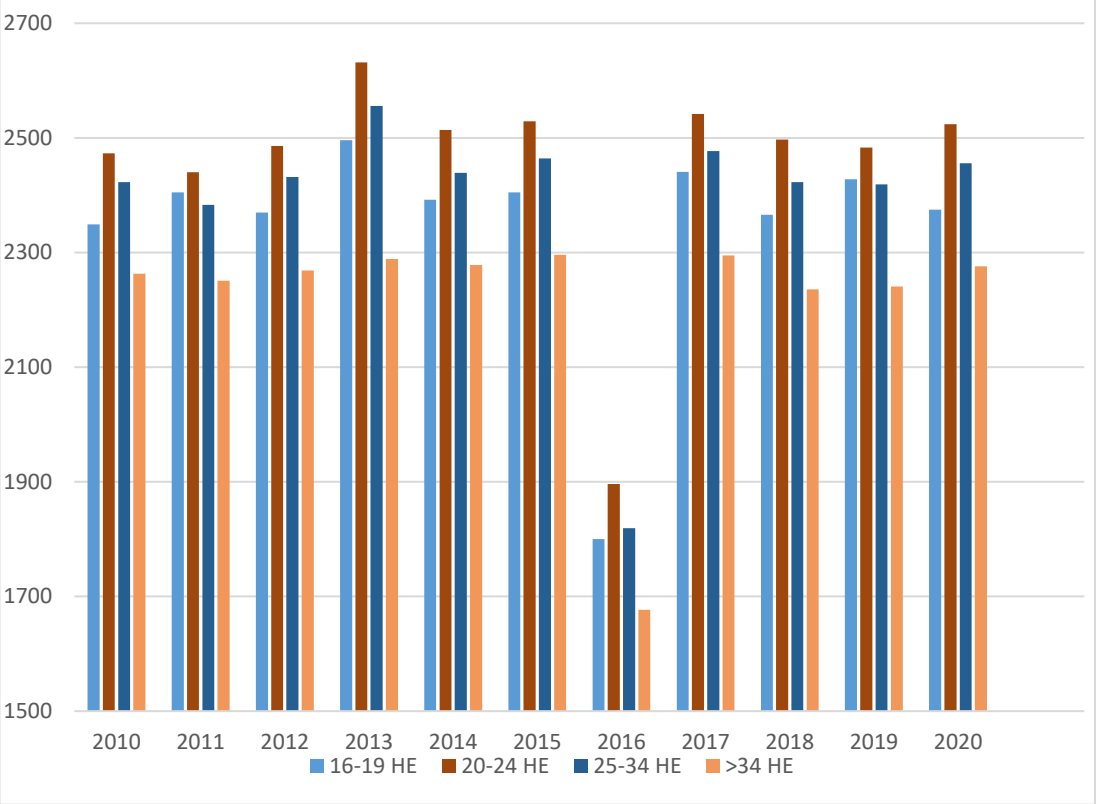


Figure 13 Mean Average Total Score by Age (highest qualification graduate)

The highest performing candidates are school leavers in the 16-19 age group although their scores are very similar to graduates in the 20-24 age group.

4.6.6 Ethnicity

At registration, UK candidates are asked to identify which ethnic group they identify with. The categories used over time have changed and for simplicity, data is presented below from 2015 onwards. The UK-other category includes candidates who declined to provide this information. Non-UK includes candidates from the EU and rest of the world.

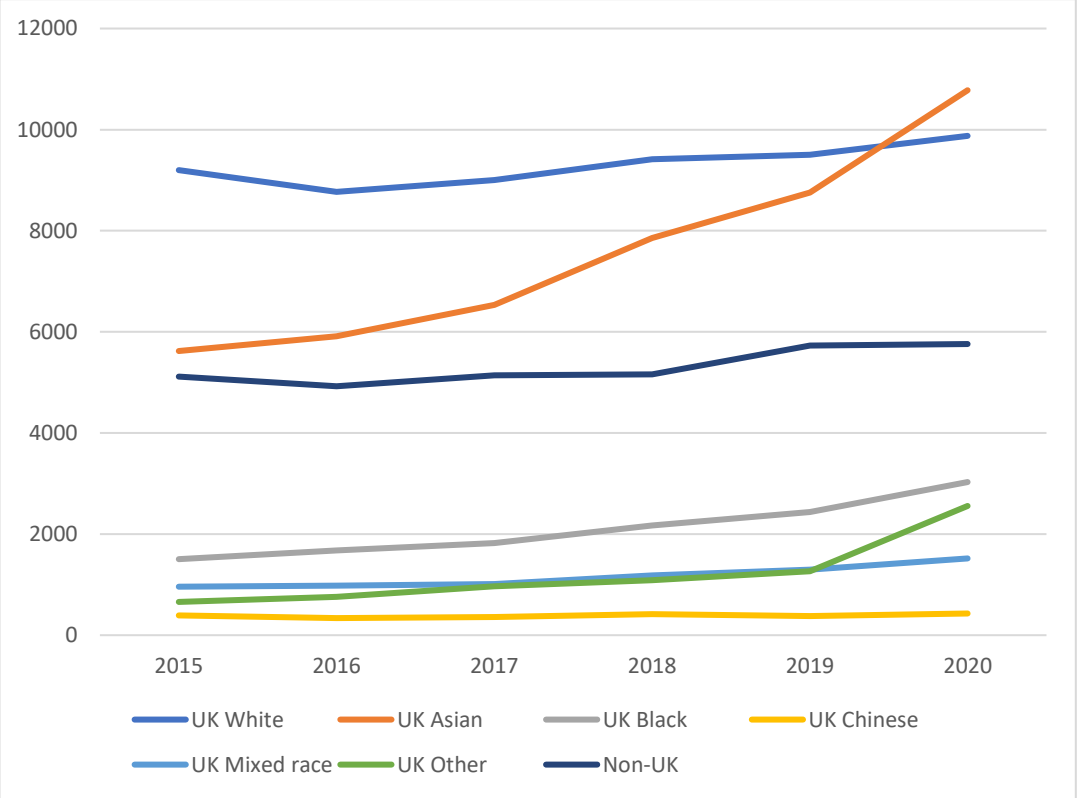


Figure 14 Candidate Numbers by Ethnic Group

Figure 14 suggests that the recent significant increases in candidate numbers might be attributed to increases in UK-Asian and UK-Black candidates. UK-Asian candidates are now the largest group of test takers.

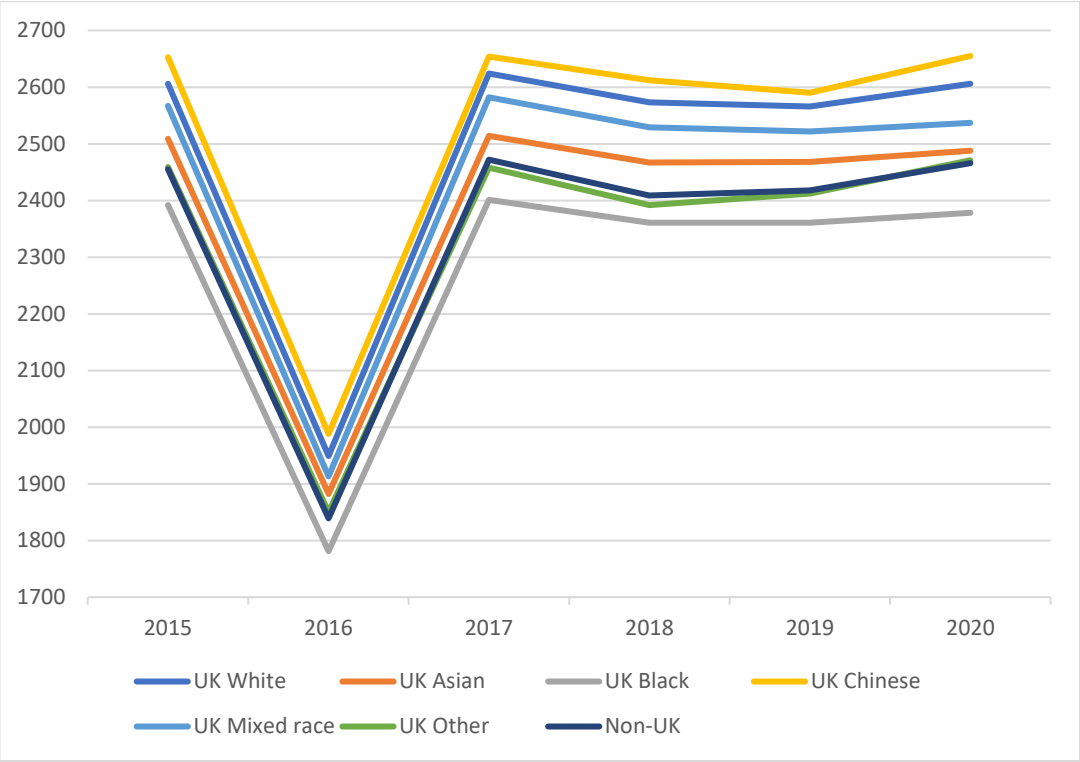


Figure 15 Mean Total Score by Ethnic Group

There is a consistent pattern to candidate performance by ethnic group with UK-Chinese candidates the highest performing group, followed by UK-White and UK-Black the lowest performers. Differential performance appears to be widening as the numbers taking the test from some of these groups increase. In 2020 the difference in total score performance between UK white and UK black candidates was 228 scale score points.

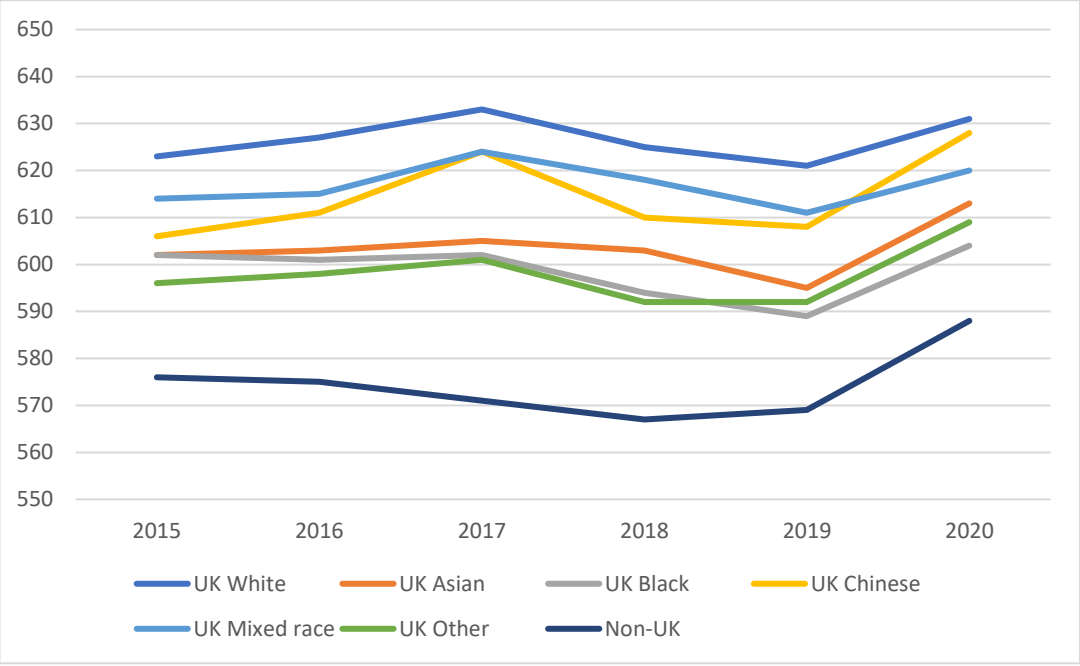


Figure 16 SJT Scale Score by Ethnic Group

Patterns in SJT performance are less consistent although UK-White candidates are the highest performers and non-UK candidates the lowest. Performance differences for UK candidates in this subtest are smaller than in the other subtests. The SJT is calibrated in a different way so direct comparisons are not appropriate. However, to provide additional context the next figure provides information regarding the relative performance of ethnic groups across subtests in 2020.

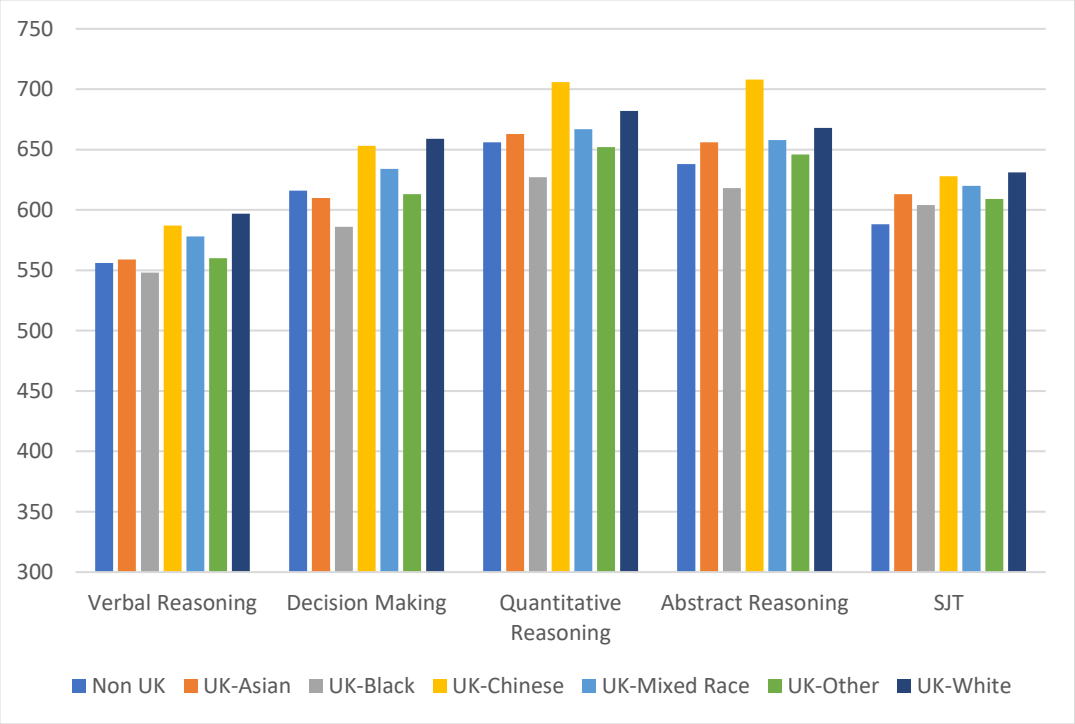


Figure 17 Mean Subtest Scores by Ethnic Group (2020)

4.6.7 English Language

Candidates provide information as to whether English is their first language.

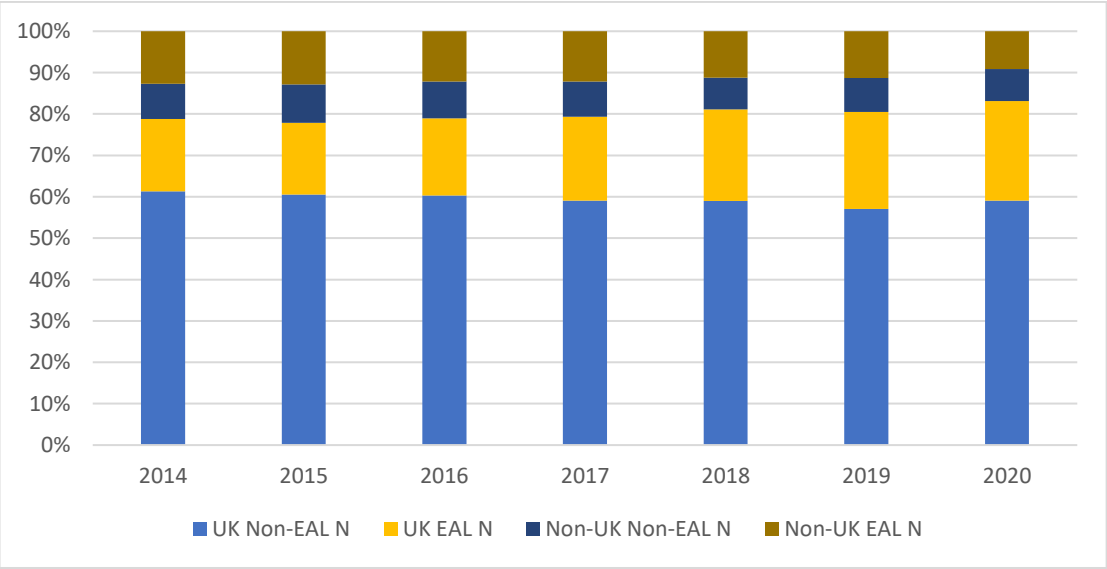


Figure 18 Number of Candidates UK/Non-UK by First Language

The proportion of UK EAL (English as an additional language) candidates has increased over time. In 2020 this accounted for 24% candidates compared with 17% in 2014.

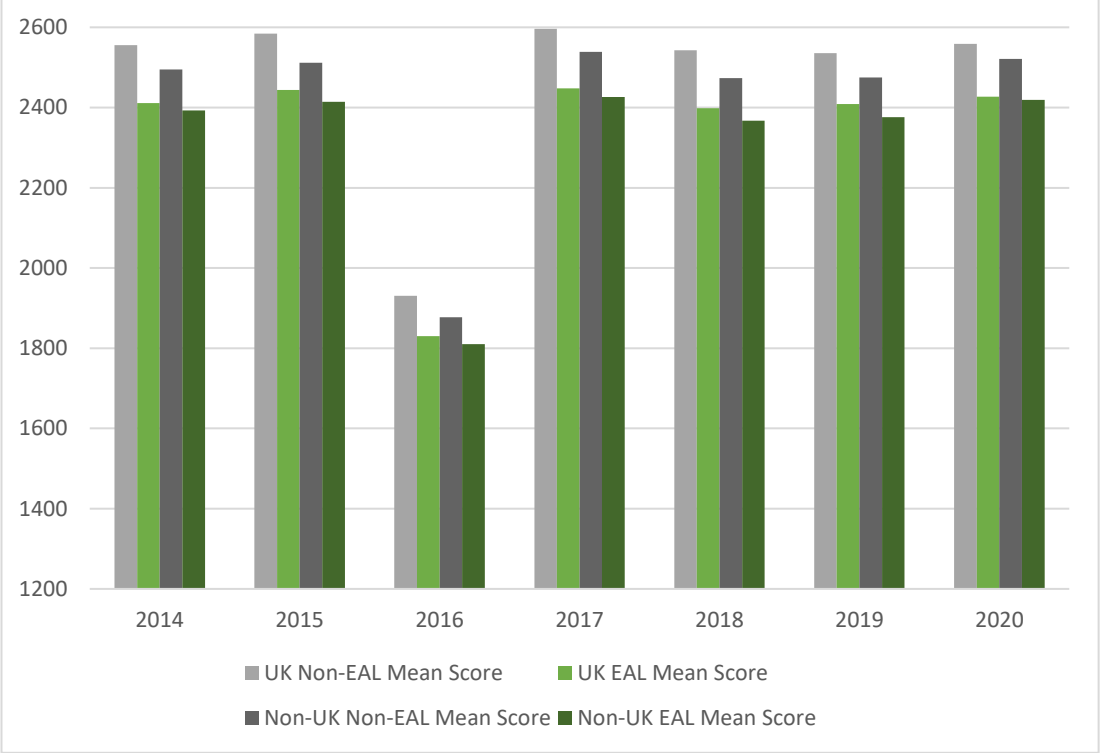


Figure 19 Mean Total Score by First Language

In terms of test performance, UK non-EAL candidates have the highest mean total scores. Non-UK Non-EAL candidates perform better than UK EAL candidates. The difference in performance for this group in 2020 was 132 points (total score).

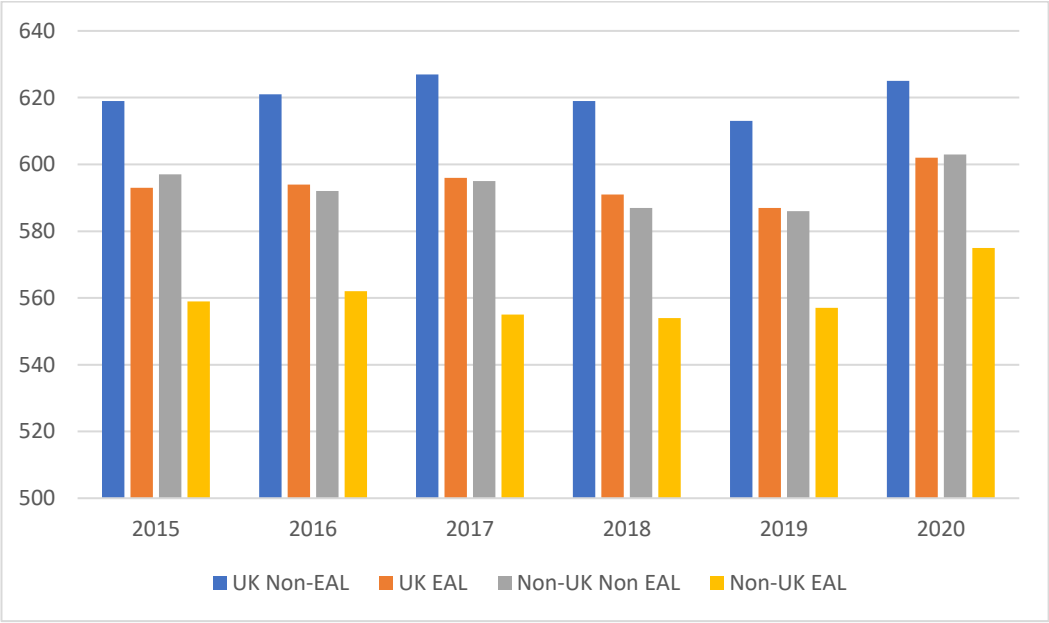


Figure 20 SJT Scale Score by First Language

In the SJT a similar pattern is observed across scores.

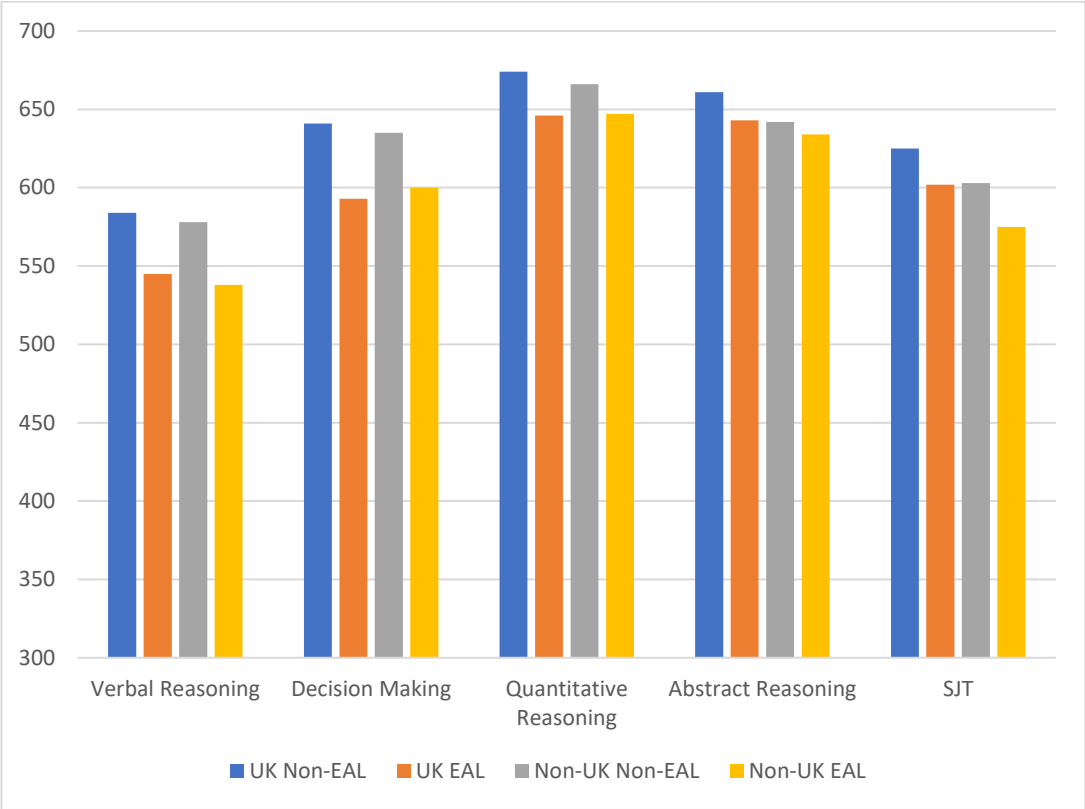


Figure 21 Mean Subtest Scores by First Language (2020)

Performance patterns are different across the subtests. The cognitive subtests with the greatest verbal load (verbal reasoning and decision making) show larger score differences between subgroups. The smallest differences occur in abstract reasoning which has minimal verbal content.

4.6.8 Reliability and Standard Error of Measurement for the Total Scale Score

The marginal reliabilities for the total scaled score are presented below alongside standard errors of measurement (SEM). Scale score reliabilities vary between subtests generally in relation to test length (number of items). SEMs also vary between subtests. In 2020 the SEM was lowest for quantitative reasoning (36) and highest for decision making (47).

	Reliability		**SEM**	
	Range	Mean	Range	Mean
2008	0.84-0.87	0.86	100.57-106.60	103.47
2009	0.86-0.91	0.89	92.38-101.70	96.89
2010	0.85-0.91	0.88	93.57-106.34	97.83
2011	0.87-0.91	0.89	91.38-103.49	97.44
2012	0.88-0.92	0.90	90.47-101.62	96.03
2013	0.89-0.90	0.90	99.63-105.09	101.77

2014	0.88-0.89	0.89	83.28-85.91	84.12
2015	0.88-0.89	0.89	84.54-87.14	85.82
2016	0.87	0.87	71.85-73.46	72.84
2017	0.89-0.90	0.89	81.50-81.93	81.71
2018	0.90	0.90	94.02-95.80	95.80
2019	0.89-0.90	0.90	79.64-80.09	79.87
2020	0.89-0.92	0.90	81.82-84.4	83.20

Table 3 Reliability and SEM for UKCAT Total Scale Score

4.6.9 UKCAT and the Impact of the COVID 19 Pandemic

Whilst preparations for 2020 testing were underway it became increasingly apparent that testing would be significantly impacted by the COVID 19 pandemic. At the point at which plans were being put in place it was impossible to predict the extent to which the candidate population would be affected by the virus and how the text centre network would operate internationally. Plans were put in place to create an online version of the test using a platform already development by Pearson Vue. Candidates were advised wherever possible to test in test centres, but the online version was made available to those who could not test in person, who did not want to test in person or had a personal preference for testing online.

The UCAT Online provided a secure solution to delivery in candidate homes utilising virtual and actual identity checks and human proctors. The delivery solution was not without challenges. Taking the test online required candidates to have the correct IT equipment and connectivity. Not all access arrangements could be supported online. Candidates under 18 needed to have a parent/guardian present to verify their identity. There were significant delivery issues relating in part to connectivity, leading to large numbers of candidates needing to retest. The product offered candidates a great deal more flexibility in terms of test dates resulting in large numbers choosing to test in the final days.

32% candidates in 2020 took the test online. Given the delivery issues around this version of the test apparent performance differences between test centre and online candidates was a concern. Post-test analysis suggested that almost all this performance difference could be explained by the different demographics of the two candidate groups.

At the same time and perhaps as a direct result of the pandemic, candidate numbers reached a record high with 34,153 candidates testing (compared with 29,375 in 2019). In 2021 this trend in increased candidate numbers continued with a record number of 37,000+ candidates.

4.7 Summary of Findings

The overall aims of the organisation have changed little since it was established. Evidence regarding the extent to which UKCAT has achieved these aims will be

further investigated in the reviews of research literature (Chapter 5) and by ascertaining the views of admission tutors (Chapter 7).

This chapter provides a high-level timeline of the development of the UKCAT Consortium and the test itself. From a loose collaboration between universities, the initial creation of the organisation was rapid, with progress from an initial outline proposal to delivery of the test taking a little over a year.

The organisation's research infrastructure was a priority from the early days of the organisation although creation of the data archive, associated research governance and commissioning of actual research and analysis did take some time. Once in place however there was rapid progress in research activity focussed on the evidence base around the test.

The UKCAT Consortium emphasis on the candidate experience is clear with developments evident to provision of online information and preparation materials. Reaching those candidates entitled to a UKCAT bursary appears to have been a priority with processes, systems and communications improving over time. Provision of accommodations to support candidates with special educational needs and medical conditions has grown substantially over the years.

The candidate view of the test with a particular focus on preparation has been obtained on several occasions. It is interesting to see how this has developed and changed over time. Schools and colleges have become more able to advise candidates on the test although the quality of this advice remains variable. Reported use of commercially available coaching has increased from 10% in 2009 to 30% of respondents in 2015. An unpublished study drawing on information obtained from a survey of candidates in 2017, reported 56% candidates had paid for commercial materials (Kulkarni S. 2020). Surveys have fed into developments in marketing and communications to candidates. Candidates would appear however to be increasingly attracted to paid for resources, perhaps reflecting their increasing utility, but also an intrinsic value placed on paid by candidates on such materials.

Delivery of the test has not always gone smoothly. The need to withdraw results from subtests on two occasions is noted above and there is a further occasion requiring the rescoring of some candidate tests.

Recurring concerns can be identified from discussions within the Consortium, and it has been helpful to me to observe these discussions over a period of time. From the outset universities were concerned about inequities created by candidate access to services offered by commercial preparation companies. Performance differences between candidate groups are regularly discussed and this data is presented elsewhere in this chapter. A particular ongoing concern is the performance difference between standard test takers and those taking the extended test linked to discussions about speededness (the extent to which candidates are able to complete each subtest) in the test.

There is evidence of the organisation looking outward to other professions and internationally. The test has been used by individual universities for other health programmes (e.g. dental hygiene, physician associates) although these have not to date led to more significant developments. In 2021 the only international associate members were the American University of the Caribbean and the University of Kurdistan Hewler. The relationship with the Australia and New Zealand universities however, represented a major development, increasing testing numbers significantly and driving a greater focus on the internationalising of test content.

Whilst the number of universities in the Consortium and the number of candidates has increased over time, the test itself remains little changed. Change has been largely iterative with the exception being the replacement of the Decision Analysis subtest with Decision Making and the introduction of the SJT.

Test and performance statistics, drawing on data in the UKCAT annual technical reports, are presented above to show trends over time. Subgroup differences in performance are of particular interest and as highlighted in this chapter have been regularly discussed within the Consortium. Of particular interest are the following observations:

- Performance in the quantitative reasoning and abstract reasoning subtests has improved over time whilst there has been a small decline in performance in verbal reasoning.
- The proportion of candidates taking the extended version of the test has increased from 3.5% in 2014 to 5.5% in 2020; UKCATSEN candidates have always outperformed UKCAT candidates (in 2020 by 66 scale score points).
- The proportion of female test candidates has grown over time from 56.5% in 2006 to 64% in 2020. Since 2010 males have outperformed females in all the cognitive subtests. Gender differences are most pronounced in quantitative reasoning. Females outperform male candidates in the Situational Judgement test.
- In 2020, 76.2% UK candidates were in SEC1 – this percentage has not shifted over time. In 2020 candidates in SEC5 scored on average 113 scale points lower than candidates in SEC1.
- The highest performing candidates are school leavers in the 16-19 age group although their scores are very similar to graduates in the 20-24 age group.
- Recent significant increases in candidate numbers can be at least partially attributed to increases in UK-Asian and UK-Black candidates. UK-Chinese candidates are consistently the highest performing group, followed by UK-White and UK-Black the lowest performers. Performance differences between these groups are significant. Differentials appear to be widening as the number of candidates from some of these ethnic groups increase.

- The proportion of candidates for whom English is a second language has increased over time to 24% in 2020. Performance differences between these sub-groups are significant.

This chapter throws some light on both candidate demographics and performance differences within the test. Whilst many discussions have taken place over the years regarding these themes, the data is presented here comprehensively for the first time. I believe it will be of benefit to the organisation to review these trends over time. Standard errors of measurement are also provided to help make judgements around observed differences. Some of these differences are within the standard error of measurement. However, these differences are consistent over time and therefore likely to be genuine performance differences. The mean average difference between some candidate groups is significant and not well understood. This is of particular concern with regards to the performance of candidates from different ethnic groups.

It is likely that there is significant interaction between some demographic markers (e.g. ethnicity and language fluency). Some subtests appear to be more sensitive to group differences than others. Greater understanding of these interactions is required and analysis of this ought to be undertaken annually.

At the same time however, some subgroups are significantly overrepresented in the candidate population in relation to the UK population. For example, in 2020 64% candidates were female. The difference in performance between male and female candidates may be partially explained by the female group representing a larger range of genuine ability than the male group. Similarly, ethnic minority candidates are significantly overrepresented in the dataset. In 2020 35% candidates declared themselves to be white British. The 2011 census data puts the UK population figure at 87%. Even accounting for shifts in demographics and missing UKCAT data this is a large difference.

Without greater understanding of these interactions and performance differences, it is not possible to identify potential changes to candidate advice and test content that might influence these differences.

Test delivery in 2020 was clearly challenged by the impact of the covid pandemic and perhaps drove the Consortium towards the use of online proctored testing rather more quickly than ever anticipated.

This chapter provides a useful resource for the UKCAT Consortium in documenting the development of the organisation and the test. At the same time this may be valuable to researchers using UKCAT data in future and wanting to understand how changes in the test might impact on research outcomes. There may therefore be value in keeping this timeline updated beyond the completion of this thesis.

Chapter 5: UKCAT and the Research Evidence Base

5.1 Introduction

One of the central aims of UKCAT was to support the development of the test through a programme of research which would also contribute to improving the evidence base around selection to medicine and dentistry in the UK. To achieve this, UKCAT set out to underpin its work with a research infrastructure which included both the creation of a Research Working Party (charged with advising the Board on research priorities and monitoring activity) and the development of a research database. The database was intended to house both test and candidate data, to allow work to take place around the validation of the test and potentially to answer broader questions regarding selection.

This chapter provides an overview of research which has taken place since 2006 in relation to the test and includes work undertaken independent of the UKCAT Consortium (though often in Consortium Medical and Dental Schools) and that directly supported by the UKCAT Consortium. As outlined in Chapter 2 both literature reviews and a systematic review are used in this Chapter.

It will explore the extent to which UKCAT can claim to have contributed to widening the evidence base around selection and comment where appropriate on the impact research outputs have had on changes to selection processes.

Many studies focus on the predictive validity of the UKCAT. In Section 1, these data are synthesised through a systematic review of the ability of the test to predict performance in medical and dental school. Study 1 draws on information from the following journal article published in 2021:

Does the UKCAT predict performance in medical and dental school? A systematic review
BMJ Open January 2021, Rachel Greatrix, Sandra Nicholson and Susan Anderson (Greatrix, Nicholson et al. 2021)

https://bmjopen.bmj.com/content/bmjopen/11/1/e040128.full.pdf

Section 2 includes a discussion around what the literature has to say about the use of the test in selection, broader issues of predictive validity (beyond performance in medical and dental school), testing of non-cognitive traits, widening access and the candidate experience. It also describes how UKMED (United Kingdom Medical Education Database) is continuing to contribute to improving knowledge about the evidence base of the test and wider selection issues for medicine.

5.2 Section 1: Systematic Review

5.2.1 Existing Reviews of the Selection Literature

Predictive validity studies are critical to help establish confidence in the use of selection tools to inform university selection processes. At the same time, such

studies ought to reassure candidates of the legitimacy of measures which might otherwise be regarded as a further hurdle in selection.

A systematic review of factors associated with success in medical school was undertaken prior to the introduction of UKCAT (Ferguson, James et al. 2002). Prior academic performance accounted for 23% of variance in undergraduate performance and 6% of that in postgraduate. The study utilised meta-analysis where sufficient quantity of data was available. This allowed them to examine impact on both undergraduate and postgraduate achievement. Interestingly this meta-analysis (and bearing in mind this was an international review of the literature) only resulted in a sample size of 21,905 participants at undergraduate level and 2,487 at postgraduate level. The authors concluded that more work was needed to identify predictivity on other factors used in selection, at the same time noting that studies looking at prediction of postgraduate performance were limited. The authors also commented on the need for more subtle examination of prediction noting that both predictors and outcome measures were likely to be intercorrelated.

Whilst not a systematic review, the approach taken by McManus et al in investigating construct validity in selection (McManus, Dewberry et al. 2013) used meta-regression to analyse the results of 6 cohort studies. The authors noted that smaller scale studies lack the statistical power to undertake such analysis and therefore chose only to include large scale cohort studies. How studies were selected for inclusion is not explained in detail, though it is undoubtedly the case that, as the authors state, longitudinal studies of selection methods are rare. The two aptitude tests included in the relevant studies (UKCAT and one other), were reported to have a relatively low overall construct-level predictive validity for undergraduate performance of .181; appreciably lower than for A-levels (.723) and GCSEs/O-levels (.359). Incremental validity (over and above A-levels) was reported as small.

A review in 2016 commissioned by the Medical Schools Council[1], investigated evidence underpinning selection to medicine in the UK (Patterson, Knight et al. 2016). The review focussed separately on different selection criteria, exploring issues relating to effectiveness, procedural issues, acceptability and cost-effectiveness. The main finding in relation to the use of aptitude tests was one of conflicting evidence. Results varied between different tests, making generalised conclusions regarding their use difficult. Sub-group differences in performance were also noted (such as gender, age, and socio-economic status), raising issues relating to fairness.

A more recent review of the literature around selection to dental training looked at European literature over the last 30 years (Cunningham, Patterson et al. 2019) and

[1] The Medical Schools Council is the representative body for UK Medical Schools (https://www.medschools.ac.uk/)

concluded that outputs were limited in number, study design often poor and reported relationships small.

5.2.2 UKCAT and Predictive Validity

Given that the UKCAT is now an established part of selection to medicine and dentistry in the UK, it is critical to understand the ability of the test to predict performance on medicine and dental programmes and indeed professional performance beyond undergraduate training. Whilst several predictive validity studies have been undertaken, no attempt had been made to synthesise these data. This systematic review seeks to provide a better understanding of the literature to assist end users of the test in making more informed decisions regarding selection.

Performance in the UKCAT is not the only criteria used by universities in selection. Predicted and achieved academic measures, personal statements and references are (or have been) routinely used to identify applicants to be interviewed. The interview itself (whether structured, semi structured or Multiple Mini-Interview (MMI)) will predominantly seek to identify those applicants with the correct personal qualities to pursue a successful healthcare career. It is therefore also of interest as to how each of these measures predict outcomes, the extent to which different criteria interact and overlap and, critically for UKCAT, its ability to predict outcomes over and above other criteria.

School leaving qualifications predict elements of performance in medical/dental school and later postgraduate performance (McManus, Woolf et al. 2013). However, a combination of grade inflation and significant competition for places has led to a reduced ability to use these grades to discriminate between applicants (McManus, Woolf et al. 2008). UKCAT (and other admission tests) therefore provide an opportunity to differentiate between high performing applicants with very similar academic records.

Longitudinal studies are difficult to undertake and cannot be undertaken until the desired outcomes are available for the relevant cohort or group. In the early days of UKCAT's development, a number of studies were undertaken looking at the first cohorts of test takers in the early stages of their programmes e.g. (Lynch, MacKenzie et al. 2009, Wright and Bradley 2010). More recently, studies looking at later performance in medical school and the foundation year application stage have taken place (Husbands, Mathieson et al. 2014, MacKenzie, Cleland et al. 2016). Whilst some studies were single or perhaps dual centre, others used data extracted from the UKCAT database[2] to look at much larger cohorts across many universities (McManus, Dewberry et al. 2013, Tiffin, McLachlan et al. 2014).

Most studies focussed on the ability of the test to predict performance in university assessments. The foundation year application process includes an educational performance measure (EPM) which has also provided a useful outcome for analysis.

[2] UKCAT has created a research database of candidate test scores and demographics. The database is held at the University of Dundee Health Informatics Centre.

Different studies have used a variety of outcome markers, at different stages of education and training, with researchers having access to different cohorts and different demographic variables.

The primary aim of this review was to evaluate existing evidence regarding the predictive validity of the UKCAT. Secondary outcomes envisaged included identification of more optimal approaches to future studies, identifying how cohorts might be best identified and outcome markers defined alongside appropriate methodologies. It was hoped that information from the review might provide additional information to medical/dental schools in evaluating selection processes.

5.2.3 Methods

The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (Liberati, Altman et al. 2009) (checklist attached as Appendix D).

5.2.3.1 Selection of Studies

The initial search and abstract screen took place in early 2018 and included searches on EBSCO, EMBASE, Educational Resources Information Center (ERIC), SCOPUS, Web of Knowledge (WoK) using the following search terms:

((UKCAT)OR(UK CAT)OR(United Kingdom Clinical Aptitude Test)OR(UK Clinical Aptitude Test)AND((valid*)OR (predict*)OR(criteri*))

The search was restricted to studies after 2006 (the first year of delivery of UKCAT).

All identified titles/abstracts were collated and reviewed for relevance in relation to inclusion/exclusion criteria. Full-text papers were accessed for relevant studies and a further decision made regarding inclusion.

Studies were included if they contain predictive validity studies around selection to medical and dental education which included the UKCAT; where all or some of the analysis focussed on the predictive validity of the UKCAT; where the target population was UKCAT test takers subsequently enrolled on medicine and dental programmes in the UK. Inclusion/exclusion criteria are summarised in Table 4.

Table 4 Inclusion/Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Empirical data	Not empirical data
Study population includes UKCAT test takers	Study population does not include UKCAT test takers
Study includes predictive validity of selection criteria including some element of the UKCAT	Study not focussed on predictive validity of selection criteria
	Selection criteria do not include any element of the UKCAT

Specific searches using the search terms of the e-journal versions of Medical Education (2006-2017), Medical Teacher (2006-2017), Advances in Health Sciences Education (2006-2017) and BMC Medical Education (2006-2017) took place as well as a review of the available published abstracts of conference proceedings of the Annual Scientific Meeting of ASME (Association for the Study of Medical Education) and the AMEE (Association for Medical Education Europe) annual conferences.

Anticipating that universities might have undertaken local (unpublished) analysis, a request to access such reports/analysis was made to UKCAT Consortium universities. This did not result in additional data sources being identified.

5.2.3.2 Search Results

Initial searches were undertaken with outcomes verified independently. The initial search took place in March 2018 with a further search in July 2018. Outcomes from the search of databases are included in Figure 22.

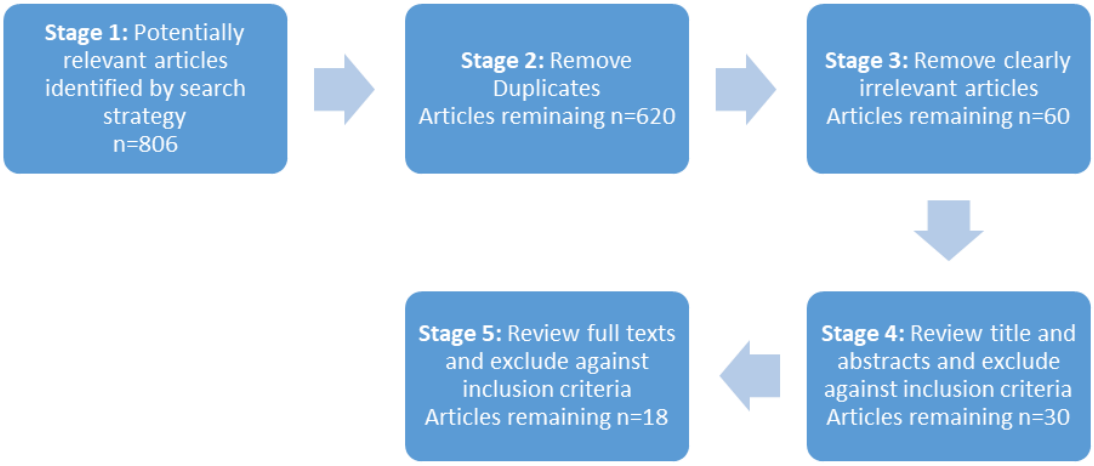


Figure 22 Database Search Results

Outcomes of a review of the grey literature (AMEE abstracts 2007 – 2017, Ottawa abstracts 2014, 2016, ASME abstracts 2009 – 2017, INRESH programmes, UKCAT Consortium Agendas) is included in Figure 23.

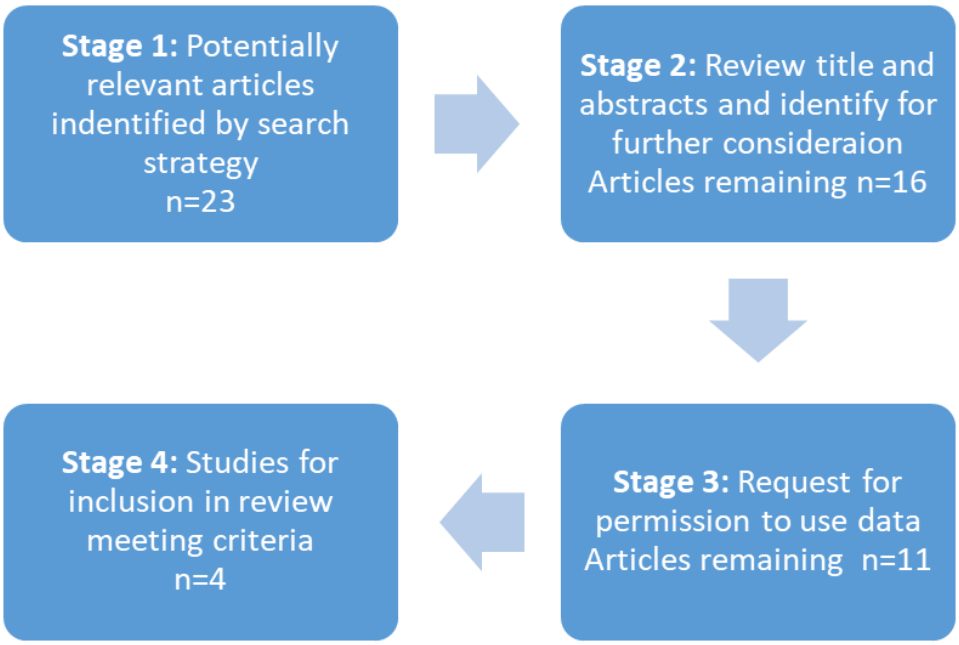


Figure 23 Grey Literature Search Results

5.2.3.3 Data Extraction

Data were extracted into a table and included article title, year of publication, sample size, number of universities included in the study, programme (medicine or dentistry) and year of admission. Correlations between predictor and outcome variables were extracted and recorded. Effect sizes and confidence intervals were calculated to create forest plots to illustrate relationships.

Similar to other studies (MacKenzie, Cleland et al. 2016), when commenting on outcomes, correlation coefficients in the range 0.00-0.29 were defined as weak, r=0.30-0.49 as moderate. No strong correlations (r=0.5+) were observed in any of these studies.

5.2.3.4 Results

22 papers were identified for inclusion in the study; 18 published articles and four articles sourced from the grey literature.

5.2.3.5 Predictor Variables

Predictor variables used were UKCAT total and subtest scores. The four subtests of the UKCAT (verbal reasoning, quantitative reasoning, abstract reasoning, decision analysis) are each reported using scaled scores (in the range 300-900) creating a UKCAT total score in the range of 1200 – 2600. Total Score has been most commonly used in selection. The SJT is reported to candidates and universities as a band (Bands 1-4 with Band 1 being the highest).

Many of the studies looked at a range of other selection parameters. Academic scores (some locally derived (Adam, Bore et al. 2012, Adam, Bore et al. 2015), personal statement/reference scores and interview scores were often included in studies to build up an authentic model of selection. Many studies considered a range of demographic factors alongside these selection criteria. Commonly used

factors were gender, age, social class, ethnicity and school type. Other predictor variables included a test of non-cognitive traits which contained content similar to that trialled within the UKCAT in earlier years (Adam, Bore et al. 2012, Adam, Bore et al. 2015).

5.2.3.5.1 Outcome Measures

Outcome variables extracted from these studies were year of assessment, assessment outcome measure and a coded proxy of the exam type (skills, knowledge or mixed assessments). The coded proxy drew on information presented in each paper as to the nature of assessment outcomes used. End of year and final examinations often include a combination of knowledge and skills assessments. The details of the weighting of the two elements were rarely reported. It is likely however, that the knowledge element had greater weighting in 'mixed' assessment outcomes.

Outcome measures varied significantly between studies. Many single/dual centre studies utilised tutor ratings and summative assessments alongside measures of non-cognitive traits such as objective structured long examination records (OSLERs) and objective structured clinical examinations (OSCEs). Some studies stratified student performance in different ways, using measures such as grade boundaries (McAndrew, Ellis et al. 2017), graduation with honours (Adam, Bore et al. 2012, Adam, Bore et al. 2015) and fitness to practice penalty points (Adam, Bore et al. 2015). Studies using UKCAT's own national dataset used aggregated end of year knowledge, skills and total marks. For these studies, researchers did not have access to information to allow more detailed interpretation of these assessments. One study used bespoke supervisor ratings against which to measure the performance of the UKCAT SJT (Patterson, Cousans et al. 2017).

5.2.4 Dental Outcomes

Four papers using dental data were identified for inclusion, reporting only 5 statistically significant results which are summarised in *Table 5*.

Table 5 Dental Schools: UKCAT predictive validity coefficients (r) with assessment outcome measures (K knowledge, S skills, M mixed)

Study Source	N	University (Year of Admission)	UKCAT Predictor (r)					Outcome Predictor Yr. K, S, M
			VR	QR	AR	DA	Total	
Lambe et al, 2016 (Lambe, Kay et al. 2016)	44	Peninsula (2014)	ns	ns	ns	ns	0.32	Yr. 1 K
			ns	ns	ns	ns	0.38	Yr. 1 K
Foley et al, 2015 (Foley and Hijazi 2015)	71	Aberdeen 2010, 2011, 2012, 2013, 2014					0.077	All M
Lala et al, 2013 (Lala, Wood et al. 2013)	135	Sheffield (2008, 2009)	ns	ns	ns	0.203	ns	Yr. 1 M
			ns	ns	ns	0.179	ns	Yr. 1 M
McAndrew et al, 2017 (McAndrew, Ellis et al. 2017)	164	Cardiff, Newcastle					ns	Yr. 1 M

Foley at al drew on admissions data from 2010 to 2014 but did not report how changes to mean average UKCAT scores across those years (e.g. mean average 2010 = 2489, 2013 = 2642) had been accounted for. Researchers reported a significant but small relationship between UKCAT percentile and assessments score (r=0.118).

Lala et al (Lala, Wood et al. 2013) only found significant correlations between Decision Analysis and assessment outcomes.

The relationships observed by Lambe et al (Lambe, Kay et al. 2016) were moderate (between total UKCAT score and year 1 outcomes), although no significant relationships were found between subtest scores and assessment outcomes. The outcome measure here was described in the paper as 'academic knowledge of dental practice' but without further detailed information regarding the assessment it was not possible to explore this relationship further.

McAndrew et al (McAndrew, Ellis et al. 2017) found no significant correlations between UKCAT and examination performance in year 1 at both Newcastle and Cardiff Dental schools. The study did however identify associations between UKCAT score and poor performance (determined by grade boundaries).

Similarly to Cunningham et al (Cunningham, Patterson et al. 2019), it was difficult to draw conclusions from the findings reported for dentistry given the small number of both studies and significant results.

5.2.5 Medicine Outcomes

Of the remaining 18 studies, two papers reported exclusively on how the UKCAT Situational Judgement Test (SJT) predicted performance in medical school (Lambe Paul 2016, Patterson, Cousans et al. 2017). 16 studies reported correlations between UKCAT (cognitive subtests) and knowledge, skills and mixed assessments (coded proxy of exam type). Of these, 15 studies reported Pearson correlations or (within regression analyses) standardised regression coefficients (beta). These outcomes can be interpreted similarly to a correlation coefficient r (Bowman 2012).

These 15 studies are presented in *Table 6* which records correlations between UKCAT (cognitive subtests and total score) and medical school assessment outcomes.

Table 6 Characteristics of the Studies relating UKCAT with Assessment Outcome Measures – Knowledge (K), Skills (S), Mixed (M)

Study Source	N	University (Year Admission)	UKCAT Predictor (r)					Outcome Predictor Year K, S, M
			VR	QR	AR	DA	Total Score	
Adam et al, 2015 (Adam, Bore et al. 2015)	146	HYMS (2007)	ns	ns	ns	ns	0.181	Yr. 4 K
			ns	ns	0.231	ns	0.175	Yr. 5 K
			0.244	ns	0.25	ns	0.204	Yr. 5 S
Adam et al, 2012 (Adam, Bore et al. 2012)	146	HYMS (2007)	0.363	0.233	0.234	0.181, 0.18	0.39, 0.214	Yr. 1 K
			ns	ns	ns	0.197	ns	Yr. 1 S
			ns	ns	0.255	0.209	0.277	Yr. 1 M
			0.253	0.267	0.202	0.258, 0.204	0.212	Yr. 2 K
			0.331	0.275	ns	0.281	0.377	Yr. 2 S
			0.241	0.2	ns	0.291	0.323	Yr. 2 M
Hanlon et al 2011 (Hanlon, Prescott et al. 2011)	341	Aberdeen (2007, 2009)	ns	ns	ns	ns	0.167	Yr. 1 K
Husbands et al 2013 (Husbands and Dowell 2013)	147	Dundee (2009)					0.25	Yr. 1 K
							0.18	Yr. 1 S
Husbands et al 2014 (Husbands, Mathieson et al. 2014)	341	Aberdeen, Dundee (2007)					0.34, 0.24	Yr. 4 K
							0.36	Yr. 4 S
							0.29	Yr. 5 S
Lynch et al, 2009 (Lynch, MacKenzie et al. 2009)	341	Aberdeen, Dundee (2007)	ns	ns	ns	ns	ns	Yr. 1 K
			ns	ns	ns	ns	ns	Yr. 1 M
MacKenzie et al 2016 (MacKenzie, Cleland et al. 2016)	6294	All (2007, 2008, 2009)	0.216	0.1	0.111	0.131	0.208	Yr. 5 S
			0.242, 0.167, 0.148	0.102, 0.079, 0.061	0.164, 0.148, 0.096	0.167, 0.133, 0.094	0.253, 0.196, 0.155	Yr. 5 M
	4811	12 universities (2007, 2008, 2009)	0.177	0.079	0.052	0.077	0.16	Yr. 1 K
			ns	0.044	0.053	0.056	0.075	Yr. 1 S

Study Source	N	University (Year Admission)	UKCAT Predictor (r)					Outcome Predictor Year K, S, M
			VR	QR	AR	DA	Total Score	
McManus et al 2013 (McManus, Dewberry et al. 2013)			0.115	0.076	0.08	0.09	0.148	Yr. 1 M
Sartania et al 2014 (Sartania, McClure et al. 2014)	189	Glasgow (2007)	0.174	0.197	ns	0.172	0.252, 0.149	Yr. 1 K
			0.145	0.155	ns	ns	0.187	Yr. 5 K
			0.213, 0.201	0.219, 0.216	ns	0.174	0.216, 0.251	Yr. 5 M
Tiffin et al 2016 (Tiffin, Mwandigha et al. 2016)	6425	18 universities (2007, 2008)	0.153	0.072	0.098	0.086	0.172	Yr. 1 K
			0.065	0.039	0.06	0.065	0.1	Yr. 1 S
			0.163	0.081	0.065	0.089	0.167	Yr. 2 K
			0.072	0.021	0.09	0.067	0.113	Yr. 2 S
			0.209	0.116	0.064	0.11	0.207	Yr. 3 K
			0.111	0.045	0.052	0.072	0.12	Yr. 3 S
			0.194	0.099	0.096	0.079	0.196	Yr. 4 K
			0.13	0.062	0.097	0.075	0.154	Yr. 4 S
			0.188	0.108	0.11	0.11	0.217	Yr. 5 K
			0.131	0.072	0.096	0.082	0.161	Yr. 5 S
Yates et al 2010 (Yates and James 2010)	204	Nottingham (2007)	0.319, 0.189	0.24, 0.152	ns	ns	0.232, 0.211	Yr. 2 K
			ns	ns	ns	-0.155	ns	Yr. 2 S
Yates et al 2013 (Yates and James 2013)	193	Nottingham (2007)	0.215	0.173	ns	ns	0.192	Yr. 3 K
			0.188	ns	ns	ns	ns	Yr. 3 S
			0.237	ns	ns	ns	0.173	Yr. 3 M
			0.266	ns	ns	ns	0.176	Yr. 4 K
			0.224	ns	ns	ns	0.259	Yr. 4 S
			0.275	ns	ns	ns	0.242	Yr. 4 M

Study Source	N	University (Year Admission)	UKCAT Predictor (r)					Outcome Predictor Year K, S, M
			VR	QR	AR	DA	Total Score	
			0.255	0.203	ns	ns	0.205	Yr. 5 K
			ns	ns	ns	ns	ns	Yr. 5 S
			0.237	0.183	ns	ns	0.193	Yr. 5 M
Mwandigha et al 2018 (Mwandigha, Tiffin et al. 2018)	2107	18 universities (2008)					0.11	Yr. 1 K
							0.07	Yr. 1 S
							0.11	Yr. 2 K
							0.06	Yr. 2 S
							0.15	Yr. 3 K
							0.06	Yr. 3 S
							0.11	Yr. 4 K
							0.07	Yr. 4 S
							0.16	Yr. 5 K
							0.11	Yr. 5 S
Srikathirkamanathan 2018 (Srikathirkamanathan K 2018)	183	Southampton (2007, 2008)	0.278, 0.176	0.26, 0.146	ns	0.162	0.337, 0.25	Yr. 5 K
			0.204, 0.157	ns	ns	ns	0.241	Yr. 5 S
			0.285	0.195	ns	0.155	0.329	Yr. 5 M
Tiffin 2017 (Tiffin PA 2017)	1400	8 universities (2013)	0.14	0.16	ns	0.17	0.17	Yr. 1 K
			ns	ns	ns	0.15	ns	Yr. 1 S

ns indicates where relationships were explored but results were not significant; blank cells indicate that a relationship was not explored.

Sample sizes ranged from 44 to 6,294 with an approximate total of 23,000 candidates/applicants included in these studies. Twelve of the studies were single centre. Five studies included a larger number of universities, where authors drew on national datasets (UKCAT and UKMED[3]). Twelve of the studies utilised data from the first years of UKCAT delivery (entry to medical school in 2007 and 2008) with the most recent study drawing on data for 2014 entry.

Most studies looked at more than one programme year. 12 of the studies used year one outcome data with 7 studies using year two, 6 year three, 8 year four and 9 year five. The highest number of relationships were observed in years 5 and 1. The number of significant correlations identified were lowest for Abstract Reasoning and highest for UKCAT Total Score. The number of correlations observed for knowledge-based assessments were higher than for skills or mixed assessments.

The study reporting no significant relationships (Lynch, MacKenzie et al. 2009) investigated relationships between UKCAT and year 1 medicine outcomes at Aberdeen for the 2007 entrants. This finding was also reported in a later study (Hanlon, Prescott et al. 2011) looking at the same cohort in which only one weak relationship between UKCAT Total Score and assessment outcomes was found. Further analysis of this cohort in later years of medical school showed a moderate relationship between UKCAT Total Score and some year four and five assessment outcomes (Husbands, Mathieson et al. 2014).

5.2.5.1 Relationships between UKCAT and Medical School Assessments

Effect sizes and confidence intervals were calculated for each relationship identified in *Table 6*. Effect sizes and confidence intervals were then aggregated by UKCAT subtest (and total score) and programme year. This allowed forest plots to be generated to illustrate the relationships between each of the UKCAT subtests (and total score) and assessment measures over different years of study.

[3] The UK Medical Education Database (UKMED) is a partnership between data providers from across education and health sectors supporting the creation of a database to analyse issues relating to selection, medical education and training and impact on career pathways. https://www.ukmed.ac.uk/

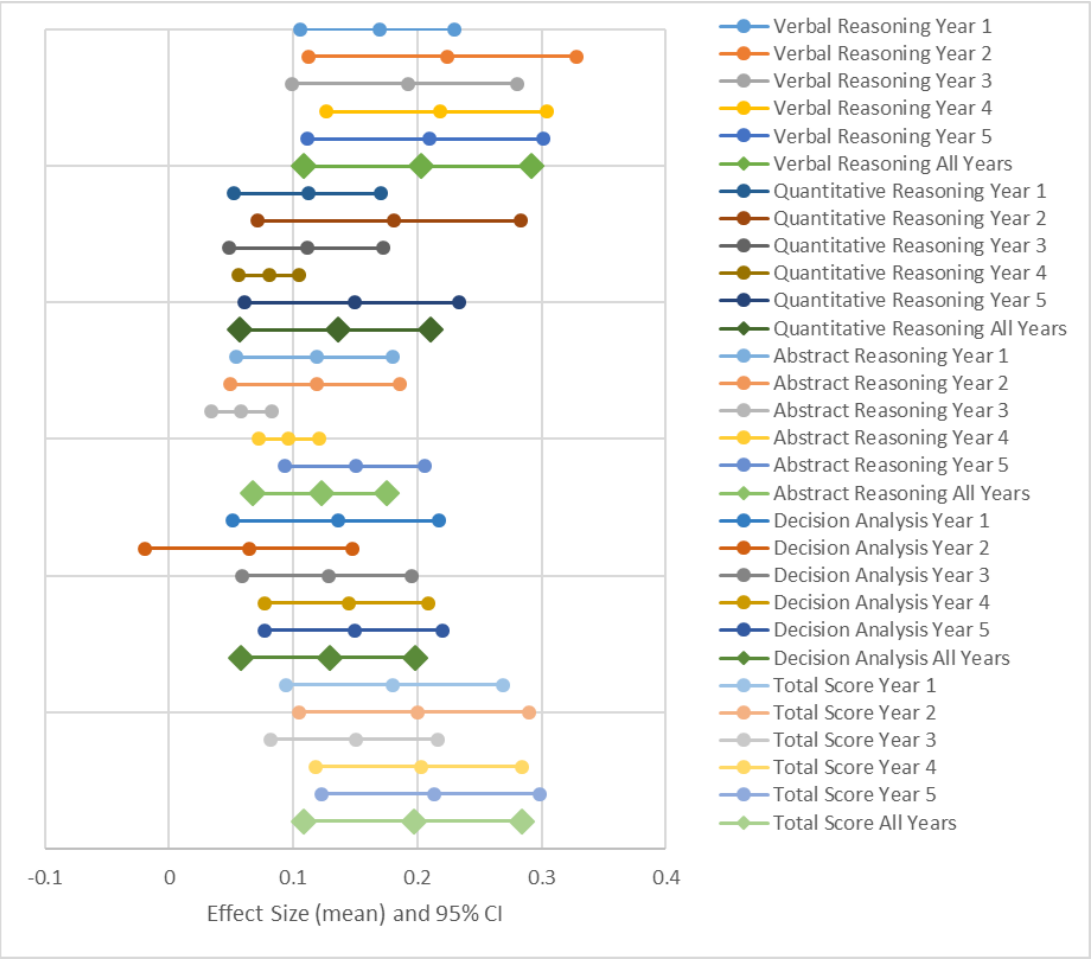


*Figure 24 Relationships between UKCAT Cognitive Subtests and **ALL ASSESSMENTS***

Figure 24 illustrates the aggregated relationships between each UKCAT subtest (and total score) with **all assessments** included in the studies for each programme year.

The strongest relationships with **all assessment outcomes** were observed for UKCAT Total Score and Verbal Reasoning although all relationships were weak. There was a very small upwards trend in relationships over the five years, with slightly larger trends observed for UKCAT Total Score and Verbal Reasoning.

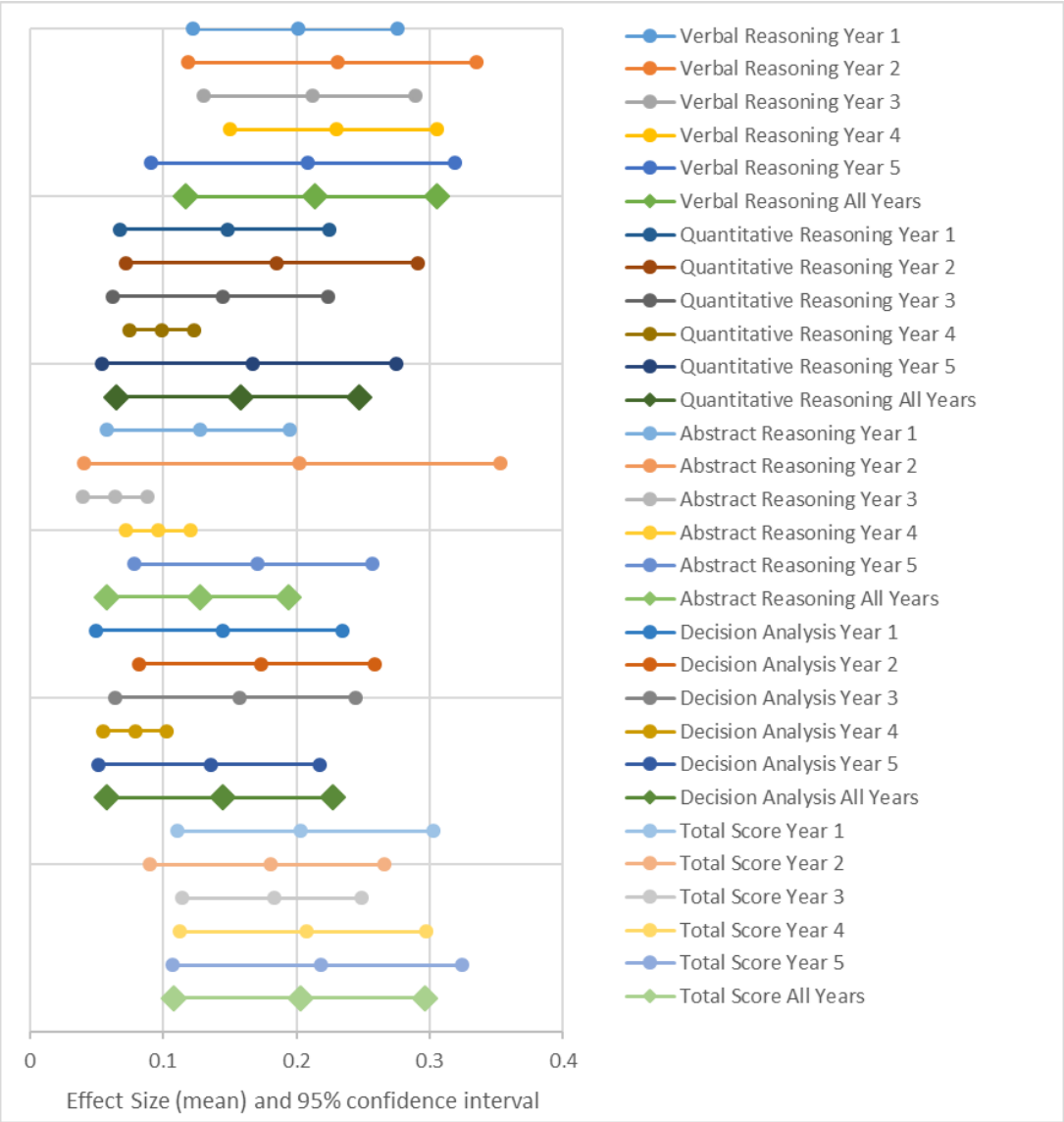


*Figure 25 Relationships between UKCAT Cognitive Subtests and **Knowledge Assessments***

Figure 25 reports relationships between UKCAT subtests and **assessments of knowledge**. The strongest relationships with **knowledge assessment outcomes** were observed for UKCAT Total Score and Verbal Reasoning, although all relationships were weak. Relationships remained fairly constant over the five years. The relationships with knowledge assessments were generally higher than those observed with all assessments.

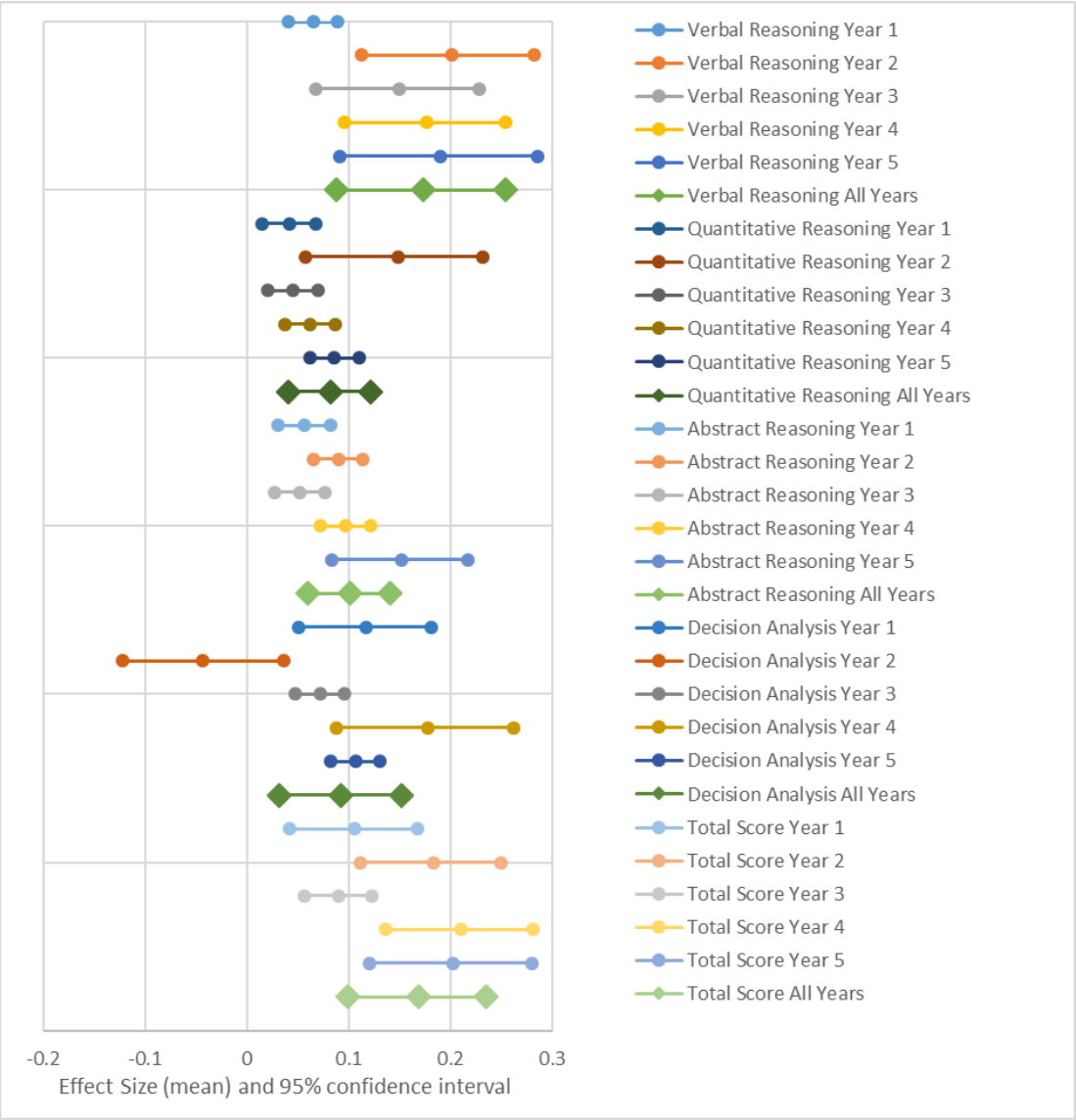


*Figure 26 Relationships between UKCAT Cognitive Subtests and **Skills Assessments***

Figure 26 reports relationships between UKCAT subtests and **skills assessments**. The strongest relationships with **skills assessment outcomes** were observed for UKCAT Total Score and Verbal Reasoning, although all relationships were weak. Relationships with quantitative reasoning, abstract reasoning and decision analysis were low. There was a slight upwards trend in relationships over the five years observed in for UKCAT Total Score, Verbal Reasoning and Abstract Reasoning. Relationships were lower than those observed for knowledge-based assessments although the upwards trend was more noticeable.

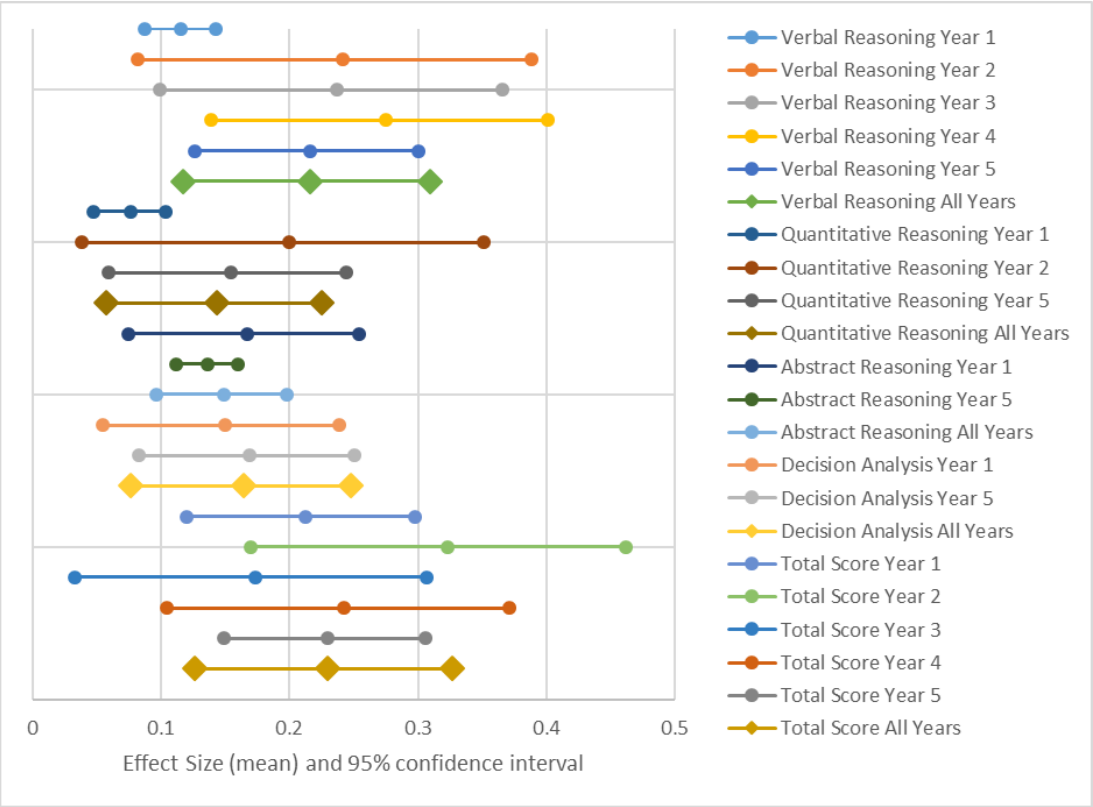


*Figure 27 Relationships between UKCAT Cognitive Subtests and **Mixed Assessments***

Figure 27 illustrates the relationships between UKCAT subtests and **mixed assessments** (those involving assessments of both knowledge and skills). The strongest relationships with **mixed assessments** were observed for UKCAT Total Score and Verbal Reasoning, although all relationships were weak.

Figures 24 to 27 suggest that the strongest relationships with assessment outcomes were observed for Verbal Reasoning and Total Score although all relationships were weak. Relationships with skills assessments were weaker than for other assessment outcomes. There was some evidence of an upwards trend in relationships over programme years but again this was small and varied across subtests.

5.2.5.2 Single/Double Centre Studies vs Multi-Centre Studies

Each study was identified as either small (single or dual centre) or large (multi-centre studies). Effect sizes were then aggregated (for UKCAT Total Score only) by programme year and outcome measure (knowledge, skills). This allowed forest plots to be generated to illustrate differences between relationships for studies of different sizes.

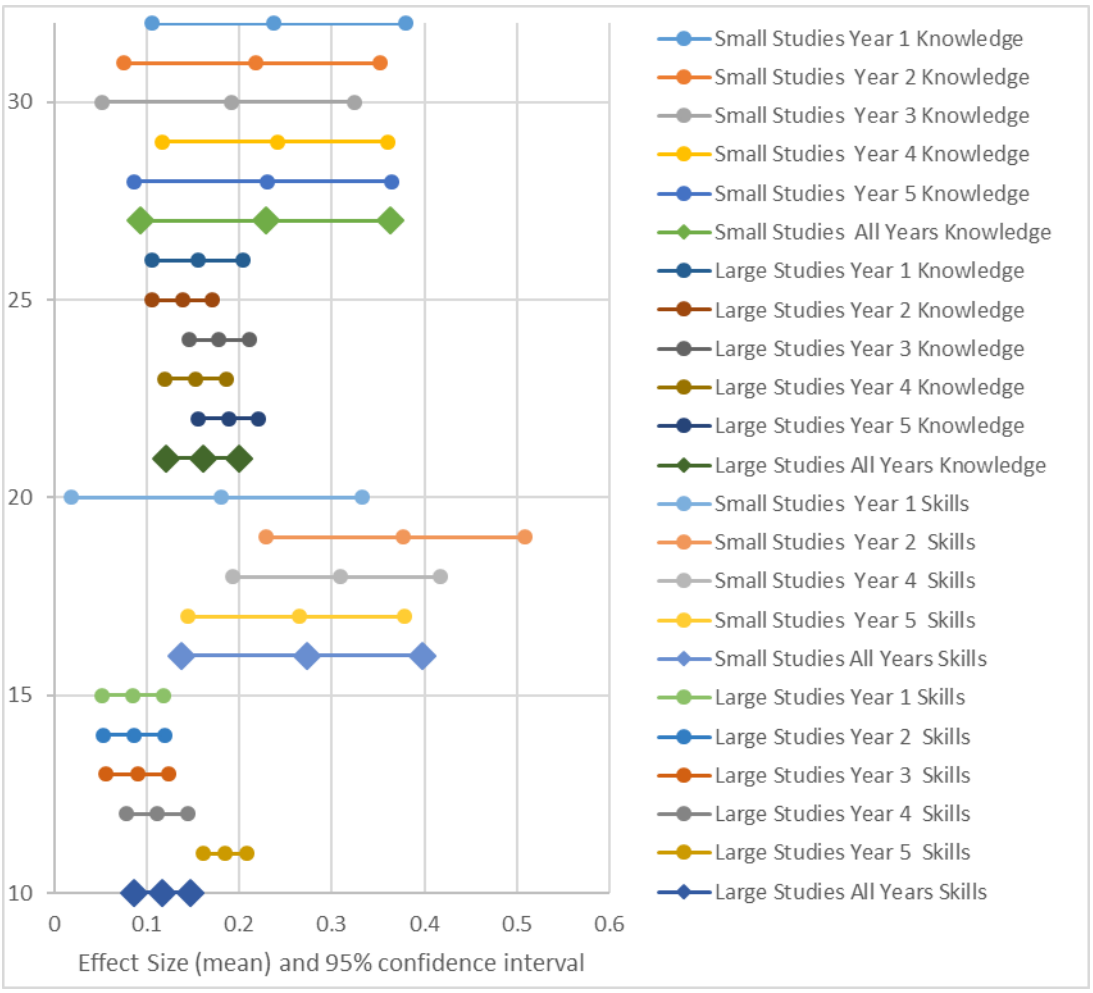


*Figure 28 Relationships between UKCAT Total Score and Assessments Outcomes by **Study Size***

The relationships observed for both knowledge and skills assessments were stronger for studies involving a smaller number of universities. Differences between outcomes for small and large scale studies were more noticeable for skills rather than knowledge based assessments. There was an increasing relationship overtime with skills outcomes in larger scale studies.

5.2.5.3 Regression Analyses

Table 7 summarises outcomes from regression analyses reported in the studies. Regression analyses examine whether a set of predictor variables predict the outcome variable. It is then possible to determine which variables are significant predictors of the outcome variable and the extent to which they impact the outcome variable. These analyses took a number of forms and, because of the diversity in both input and output variables, it was not possible to directly compare results.

Table 7 Regression Analyses

Study	Summary of Findings
Adam et al, 2015 (Adam, Bore et al. 2015)	Prior ability (including UKCAT) predicts some year 4 and year 5 outcomes although 'other academic achievement' predicts stronger than UKCAT; demographic variables (gender, age, domicile) reduce the effect of prior achievement.
Husbands et al, 2013 (Husbands and Dowell 2013)	UKCAT scores explain 6% of the variance in Dundee Year 1 assessments.
Husbands et al, 2014 (Husbands, Mathieson et al. 2014)	UKCAT Total Score explains 6-13% of variance in Aberdeen Year 4 and 5 exams and 11% of variance in Dundee year 4 examinations.
Hanlon et al, 2011 (Hanlon, Prescott et al. 2011)	No meaningful relationships reported between UKCAT scores and assessments outcomes.
MacKenzie et al, 2016 (MacKenzie, Cleland et al. 2016)	UKCAT Total Score and all subtest scores were significantly and positively associated with all four outcome measure in UKFPO applicants.
McManus et al, 2013 (McManus, Dewberry et al. 2013)	The incremental validity of the UKCAT over current educational attainment was small but significant.
Sartania et al, 2014 (Sartania, McClure et al. 2014)	UKCAT Total Score was independently associated with course performance before and after adjustment for gender, age, ethnicity and deprivation. UKCAT scores predict knowledge outcomes although in most cases, effects reduce upon adjustment for the effect of advanced qualifications.
Tiffin et al, 2017 (Tiffin PA 2017)	SJT scores remained significant predictors of theory performance even after adjustment of cognitive ability.
Yates et al, 2010 (Yates and James 2010)	UKCAT Total Score has a weak relationship with two curriculum themes. Quantitative Reasoning predicts outcomes in one theme and Verbal Reasoning in two themes.
Yates et al, 2013 (Yates and James 2013)	UKCAT Total Score, Quantitative Reasoning and Verbal Reasoning showed significant effects for a knowledge based exam.
Wright et al, 2010 (Wright and Bradley 2010)	UKCAT a significant predictor of performance in almost all exams.

The variety of models used by researchers in regression analyses creates a challenge regarding the presentation of these data. However, these analyses further support findings outlined above that UKCAT has a low but significant relationship with performance outcomes. Effects tended to reduce once other prior achievement was considered.

5.2.5.4 *Incremental Validity*

Possibly because of the lack of available data and/or the complexity of the analysis, few studies explored the extent to which the UKCAT predicts performance over and above conventional measures of academic attainment such as A-levels. McManus et al concluded that the incremental validity of the UKCAT after taking educational attainment into account was 'small but significant' (McManus, Dewberry et al. 2013). This finding was further supported by Tiffin et al (Tiffin, Mwandigha et al. 2016).

5.2.5.5 *Predicting other Progression Outcomes*

Tiffin et al (Tiffin PA 2017) analysed the odds of students passing at first sitting (compared with other academic outcomes such as fail or resit). The UKCAT SJT score predicted the odds of passing at first attempt (odds ratio = 1.28). The authors interpret this finding: 'for every one standard deviation above the mean for applicants scored on the UKCAT SJT, the odds of passing first time will increase by around 28%.'

Adam et al (Adam, Bore et al. 2012, Adam, Bore et al. 2015) looked at predictors of appearing in the top 20% and bottom 20% of performers at medical school. Being in the top 20% of achievers in year 5 written examinations was associated with having a higher UKCAT Total Score (higher Quantitative Reasoning, Verbal Reasoning and Abstract Reasoning scores in particular).

Average UKCAT total scores were reported by McManus et al (McManus, Dewberry et al. 2013) for students with different end of year 1 progression markers. The mean average score for students that passed all assessments first time was 2544, for those who passed after resits this was 2486 and those required to repeat the year 2457.

The increasing ability of the UKCAT to predict performance as students progressed through their courses (over and above prior attainment) was reported by Tiffin et al (Tiffin, Mwandigha et al. 2016). Similarly, Sartania et al (Sartania, McClure et al. 2014) concluded that whilst both UKCAT and a locally derived science score predicted year 1 performance, UKCAT was the only pre-admission measure to independently predict final course performance ranking.

MacKenzie et al (MacKenzie, Cleland et al. 2016) reported an unexpected relationship between the UKCAT and the UKFPO SJT (r=0.208), a finding replicated in a later study (Garrud and McManus 2018).

5.2.6 Discussion of Section 1

The systematic review of the predictive validity of the UKCAT, for the first time synthesises outcomes in the literature. Relationships between the UKCAT and medical school outcomes were observed in the vast majority of the 22 identified studies with researchers reporting upon these relationships using a range of analyses.

Verbal Reasoning scores appear to predict rather more as a stand-alone subtest than the other subtests. It has also been reported that this subtest has the lowest correlation with education attainment (school leaver) of all the subtests (McManus, Dewberry et al. 2013) suggesting incremental value over (mainly) science school leaver qualifications. This finding ought to be of interest to selectors – if Verbal Reasoning predicts outcomes rather better than other aspects of the test and measures something different to other performance measures, there may be a case for it being treated differently to the other subtests in selection. Verbal Reasoning is also relatively underweighted as part of the total score (having a lower mean average); scaling adjustments could be made to address this if this subtest is deemed to be of greater value.

McManus et al reported that the correlation with assessment outcomes for mature applicants (r=0.252) was higher than non-mature (r=0.137). The mature applicant population (21+) is a diverse one. If the UKCAT offers more utility for this group, it allows selectors to compare such applicants who may be offering a diverse range of qualifications (potentially non-standard). There is little sub-group analysis included in these studies – it would be reasonable to speculate that the UKCAT might also predict better for other diverse sub-groups of applicants such as international and widening access applicants.

Two of the larger and more detailed studies (McManus, Dewberry et al. 2013, Tiffin, Mwandigha et al. 2016) confirm that UKCAT has small incremental validity over and above other measures of academic attainment available at the point of selection. Tiffin et al's findings were across all years of medical programmes which is of particular interest as the paper also reports the declining ability of prior academic achievement to predict outcomes over the course of medical programmes.

Stronger relationships were observed in smaller single and dual centre studies identifying a tension between the power of large cohort studies (with large sample sizes and access to a potentially greater range of consistent demographic markers) against the ability of more local studies to utilise more fine-grained (and arguably more meaningful) outcome markers. The large cohort studies included in this studied were only able to utilise high level end of year assessments without any direct knowledge of exam content. Understanding the differences between the studies and potential magnitude of differences in outcomes will be helpful when interpreting future studies.

The ability of the UKCAT to predict final course outcomes (Sartania, McClure et al. 2014, Tiffin, Mwandigha et al. 2016) might be explained if the impact of innate cognitive performance continues whilst the impact of prior education declines. The relationship between the UKCAT SJT and UKFPO results is more difficult to explain, suggesting the need for a greater understanding of the traits being measured in this subtest.

5.2.6.1 Limitations in Identified Studies

- *Outcome Markers:* Authors comment on the limitations in predictive validity studies created by a lack of relevant outcome markers. Adam et al (Adam, Bore et al. 2015) sought to address this issue by looking at a wide range of fine grained outcome markers. Husbands et al (Husbands and Dowell 2013) also noted the need to further investigate how selection criteria predicted 'specific cognitive and non-cognitive attributes for which they were designed'. The identification of relevant outcome markers has created additional challenges in attempts to validate the UKCAT SJT, with researchers utilising bespoke outcome measures in studies (Patterson, Cousans et al. 2017) due to the difficulties of identifying existing relevant assessments within medical and dental programmes.

- *Interpreting Results:* The audience for research into admissions is a diverse one. Medical educators are interested in the outcomes alongside others involved in medical selection, applicants and their advisors. In light of this, additional interpretation of findings in studies is desirable. In the studies reviewed, some authors provided additional interpretations to demonstrate the strength of reported relationships (Wright and Bradley 2010, MacKenzie, Cleland et al. 2016).

- *Range restriction:* Range restriction in selection is an important issue to understand when interpreting such studies. It is inevitable for example that the range of scores in successful students is smaller (and significantly so at some) than the range of scores of the total candidate population. A PhD funded by UKCAT evaluated methods for correcting for bias due to range restriction when reporting on predictive validity (Mwandigha 2017).

  Range restriction creates challenges with these studies because outcomes can only be observed for successful applicants, who are likely as a group to have scored higher in the UKCAT than unsuccessful applicants. Some studies adjusted correlations for range restriction, providing these findings alongside unadjusted outcomes. The adjustments observed by Husbands et al (Husbands and Dowell 2013) result in stronger correlations between UKCAT and outcome markers (r=0.34 after adjustment vs r=0.25; r=0.24 after adjustment vs r=0.18). Tiffin et al (Tiffin, Mwandigha et al. 2016) included figures which show the magnitude of increases in coefficients following correction for range restriction. The relationships were stronger in all cases once this adjustment had taken place, with the largest increases observed for Verbal Reasoning and Total Score.

  Wright at al (Wright and Bradley 2010) comment on the limitations of restricted range in their work but preferred to report the more conservative approach given the recognised limitations of adjusting.

There has been perhaps a reluctance on the part of researchers to utilise such methods in case they might be regarded as 'artificially' increasing relationships. It has however been argued that corrected correlations are less biased than those reported without correction (Zimmermann, Klusmann et al. 2017).

5.2.6.2 Implications for the Future

UKCAT should encourage stakeholders to continue to undertake predictive validity studies to further inform the development of the test and selection processes more generally. The creation of the United Kingdom Medical Education Database (UKMED) will facilitate researchers undertaking full cohort studies (across an entire or multiple medical student intake(s)) with a consistent range of demographic, academic achievement and progression markers. At the same time however, more local studies able to investigate the ability to predict individual assessment outcomes will continue to have utility. UKCAT should consider supporting a large cohort study every five years to ensure that validity evidence remains up to date.

5.2.6.3 Recommendations for Future Studies

Those undertaking future studies should:

- Provide consistent detail regarding assessment outcomes being investigated.
- Interpret findings clearly for the benefit of selectors (and test takers).
- Consider performance in the middle years of medical school as fewer studies have looked at years three and four.
- Focus on the lack of evidence regarding dentistry.
- Undertake analysis of relevant sub-group differences (e.g. age, gender, international, widening access) with regard to prediction of outcomes.
- Include analysis which adjusts for range restriction or at least comment explicitly on the limiting impact of not undertaking such analysis.

UKMED opens up the opportunity to explore the extent to which all selection criteria (including other aptitude tests) predict performance in medical school and the interaction between such criteria. Future studies will increasingly be able to investigate outcomes beyond medical school into postgraduate training and beyond. This will create opportunities to investigate how factors used in selection predict career progression and choices.

The work undertaken by Tiffin et al (Tiffin, Mwandigha et al. 2016) in particular would lead naturally to further studies investigating the extent to which the UKCAT might compensate for lower A-level achievement and the impact this might have on opening up routes to widen access.

The UKCAT Consortium should consider the findings from this systematic review in relation to the future development and use of the UKCAT itself. Verbal Reasoning

clearly plays a great part in the relationships observed. If (as McManus suggests) this subtest also correlates least with prior attainment (A-levels and equivalent) then there is a case to be made for this subtest having a higher weighting or at least being treated differently in selection. A radical approach along these lines would logically increase the ability of the test to predict assessment outcomes in the future.

5.2.6.4 Strengths and Limitations

A particular challenge of longitudinal studies in selection is that during the time required to observe relevant output measures nothing else stands still. This systematic review allows conclusions to be drawn from studies over a significant time period. Even so, the UKCAT test itself has gone through significant change as have university curricula; the expansion in student numbers may also impact on applicant demographics and test performance. Whilst it can be assumed that outcomes reported here might be generalizable to an extent, there is an ongoing need to undertake further studies to reassure future selectors and test takers that the UKCAT remains fit for purpose.

Variability between studies makes generalising across them challenging. Studies took place over a number of years and so year 1 in one school was not the same year 1 (by calendar year) in another. Outcome markers varied in nature and number between schools.

To provide greater interpretation of complex data, some results are presented as aggregated outcomes by year group and ought as such be treated with caution.

5.3 Section 2: A Review of the Wider Literature

5.3.1 Use of the Test in Selection

How the test was used in selection was described in an early paper (Adam, Dowell et al. 2011) looking at use of the test in medical schools up to 2010. The categorisation of use of the test within that paper (e.g. threshold, factor, borderline) has since then been used routinely within the Consortium and by researchers (Tiffin, Dowell et al. 2012, Tiffin, Webster et al. 2013). A more descriptive paper subsequently looked at how the test had been used in dental selection (McAndrew and Greatrix 2014).

The approach adopted in both these papers informed reporting on how the use of the UKCAT has changed since 2006 (Greatrix and Dowell 2020, Greatrix and McAndrew 2021). These findings can be found in Chapter 6.

5.3.2 Predictive Validity

The ability of the test to predict performance in medical and dental education is discussed above. The UKMED database has allowed for more ambitious projects to take place such as an investigation into predictors of fitness to practice declarations (Paton, Tiffin et al. 2018) where UKCAT was included as one of the predictors. Findings from this study were mixed, with some non-cognitive outcomes (trialled within the UKCAT from 2007 to 2010) associated with increased risk of a conduct

issue (higher self-esteem) and of a health declaration (low confidence). Researchers noted that the small effect sizes observed in these relationships did not support the further use of these particular tests of non-cognitive traits within selection processes.

A later study (not included in the systematic review) used UKMED data to investigate the impact of programme type on outcomes for graduate entrants to medicine and included UKCAT scores (and other aptitude test scores) in its predictor variables (Garrud and McManus 2018). The authors noted the finding in the UKCAT-12 study (McManus, Dewberry et al. 2013) which suggested that the UKCAT had greater predictivity for mature (age >21) students than school leavers. Perhaps reflecting the competition for such programmes, UKCAT (and GAMSAT) scores for those on accelerated 4-year programmes had higher scores than those entering standard programmes. Higher UKCAT scores was associated with successful completion of programmes and with the FPAS Educational Performance measure (EPM) and SJT. Correlations between UKCAT and GAMSAT scores (for those candidates who had taken both tests) was high. GAMSAT (with significant science content) had a higher correlation with FPAS EPM (r=.240-0.36 vs r=0.170-.202) whilst, more difficult to explain, UKCAT had the higher correlation with the FPAS SJT (r=.304-0.358 vs r=0.235-0.339).

UKMED data was also used to investigate relationships between performance at medical school collection and in Royal College of Surgeons (MRCS) examinations (Ellis, Brennan et al. 2021). UCAT, GAMSAT and BMAT independently predicted performance having adjusted for prior academic attainment.

In one further interesting study researchers created a 15 minute mock UKCAT test (drawing on UKCAT's own practice materials) which was then taken by 167 doctors and 26 lay people (Blackmur, Lone et al. 2016). The authors concluded that 68% of the doctors 'passed' the mock test (achieving a score they deemed to be equivalent to that required to be selected for interview). Differential scores in the cognitive tests were observed between specialities (anaesthetists and intensive care physicians scoring the highest and GPs scoring lowest). Whilst it is difficult to draw many conclusions from this study given the sample size and methodology, differential attainment between specialities is certainly an area for exploration in the future.

5.3.3 Testing Non-Cognitive Traits

Several studies have sought to investigate the utility of tests of non-cognitive traits drawing on evidence from the UKCAT research database. Between 2007 and 2010 the UKCAT trialled different approaches to measuring non-cognitive abilities such as empathy, resilience and confidence. These are described in more detail in Chapter 4. In 2012 the UKCAT Situational Judgement Test was trialled within the test and has been used operationally since 2013.

One test of non-cognitive ability trialled within the UKCAT was the Personal Qualities Assessment (PQA). A study across medical schools in Scotland explored the predictive validity of this particular test noting its potential for use in selection through the UKCAT (Dowell, Lumsden et al. 2011). The authors found no significant relationships between separate elements of the PQA and performance in medical school. A similar study at Hull York Medical School considered how the different tests of non-cognitive ability trialled in the UKCAT in 2007 predicted performance in early years at Hull York Medical School (Adam, Bore et al. 2012). Relationships were observed, although they were generally weak, leading the authors to conclude that such measures, if developed with care, might improve medical student selection. These findings were further supported by a follow up study looking at performance in the later years of the programme (Adam, Bore et al. 2015).

Two studies were able to analyse the ability of these tests of non-cognitive attributes to predict undergraduate performance using foundation year application data (educational performance measure and foundation SJT) (MacKenzie, Dowell et al. 2017) and on-course assessments (Finn, Mwandigha et al. 2018) as outcomes. Findings were limited and it was not felt that these studies supported the further use of these measures in selection.

The first study to analyse how the UKCAT SJT might predict on-course performance looked at a dental programme (Lambe, Kay et al. 2016). The paper also modelled the impact of the use of the SJT within existing selection processes. SJT bandings correlated with interview score but not with first year study assessments. Use of the SJT to identify successful applicants would have resulted in offers being made to some candidates rejected after interview (including red-flagged candidates). In addition, the authors modelled the impact of identifying candidates for interview using a combination of UKCAT cognitive scores and SJT bandings. Again, this would have resulted in some invitations to interview to undesirable candidates (as determined by existing processes). The authors concluded that the study did not provide evidence to support the use of the SJT in that particular admissions process. The lack of predictive validity might well lead to this conclusion although it is possible that you would not expect the SJT to predict the outcomes outlined in this study. Equally, no selection system is perfect, and it seems flawed to compare one unproven selection criteria with another and conclude that one is less valid. The developers of the SJT have argued that the test ought to be used in combination with other factors and would not themselves support this subtest being used as a single determiner of offer for candidates.

In an attempt to address the difficulty in identifying appropriate outcome markers to measure the SJT against, a further study used a bespoke supervisor rating as an outcome marker (Patterson, Cousans et al. 2017). Correlations existed between SJT scores and supervisor ratings which though not large, supported the potential of the test to add value to selection processes. Further analysis of these data in an unpublished study demonstrated differences in how different versions of the test

were performing (Paton 2017). 2013 had been the first year the SJT had been used operationally and the authors urged additional work to create greater uniformity within the test.

Researchers investigating the impact of demographics on outcomes in the UKCAT SJT reported that effect sizes for socio economic class were lower for the SJT than for the cognitive tests (2012 and 2013 test takers) (Lievens, Patterson et al. 2016). Females outperformed males in this subtest. Whilst noting that the SJT was early in development the authors concluded that the SJT showed some potential to widen access if introduced to selection processes.

An unpublished study on the UKCAT website reports on relationships between the UKCAT SJT and MMI outcomes from 7 medical and dental schools (Adrian Husbands 2018). A small correlation was observed between MMI and the SJT (r=0.12). Significant relationships reported for individual schools varied from 0.14 to 0.3. The variations at play here would support further investigation of these relationship, including gaining a better understanding of the content of MMIs were relationships were observed.

A broader look at the use of situational judgement tests in healthcare selection (a systematic review) concluded that their use was supported with greater evidence available during postgraduate assessment/recruitment rather than undergraduate (Webster, Paton et al. 2020). The authors noted the difficulty in designing appropriate outcome measures to test the performance of SJTs used in undergraduate selection and highlighted the need to understand how traits may change over time. There is also an issue (relevant to how the UKCAT SJT is used in selection) that if the relatively small number of lower performing candidates are excluded at selection, then restriction of range in those in training will be significant.

Between 2013 and 2016 UKCAT trialled a measure of confidence within the decision analysis subtest. Having responded to an item, candidates were asked to rank how confident they were that their answer was correct. The performance of this measure was investigated in 2019 (Tiffin and Paton 2019). Overconfidence was inversely related to cognitive performance; under confidence was associated with increased odds of receiving an offer to study medicine. There was a modest link between confidence and academic performance (defined as passing first year examinations). Whilst these findings were of interest, the authors highlighted significant challenges in the utility of these measures in high stake testing. For example, a proportion of candidates did not vary their answers and it was unclear how candidates could be prevented from this behaviour. There would always be a potential for candidates to 'game' their responses if it was understood that under confidence was a more desirable outcome.

5.3.4 UKCAT and Selection Processes

A single centre study at the University of Aberdeen set out to compare UKCAT performance from 2006 with medical student selection outcomes (Fernando, Prescott et al. 2009). UKCAT correlated with local selection scores although correlations were generally weak. UKCAT scores amongst academic rejects were significantly lower than other candidates. The authors concluded that these weak correlations suggested the UKCAT was testing other traits to those currently assessed within the local selection process and that further long-term studies were needed to determine its predictive value.

A similar study at another university also set out to determine the extent to which the UKCAT might select suitable candidates for interview (Turner and Nicholson 2011). Analysis of the UKCAT scores of candidates selected for interview (using academic scores and a review of the UCAS form) demonstrated that the rejection rate of candidates with lower UKCAT scores was 2.7 times that of candidates with higher scores. No relationship was found between UKCAT scores and interview performance.

In order to evaluate the introduction of MMIs, St George's Medical school compared performance in standard and multiple mini interviews. UKCAT results were used as a variable in the study (which included only 47 candidates) but there were no relationships found between UKCAT scores and interview performance.

5.3.5 Widening Access

Given the increased focus on widening access since 2006 it is unsurprising that studies have focussed on the role UKCAT may have had in widening access.

Early analysis of UKCAT candidate demographics (from 2006), test results and A-level outcomes confirmed that independent predictors of high UKCAT scores were male sex, white ethnicity, being from a professional or managerial background and independent/grammar schooling (James, Yates et al. 2010). The impact of schooling on test performance was not as significant as that for A-level outcomes. A later study supporting these findings concluding that test scores may be more influenced by gender though less sensitive to school type than A-levels (Tiffin, McLachlan et al. 2014). A much smaller study at Queen's Belfast reported that only 3% of the variance in UKCAT scores could be attributed to social class (McKinley, Stevenson et al. 2017), perhaps suggesting differences in applicant cohorts between the countries of the United Kingdom.

A study including 8459 applicants to medicine in 2009, analysed the impact on candidate demographics of different uses of the UKCAT in selection (Tiffin, Dowell et al. 2012). The authors concluded that candidates from lower socio-economic groups were more likely to receive an offer from schools using the test strongly (defined as applying a threshold) in selection. It is possible that applicants from lower socio-economic backgrounds with high UKCAT scores would have been selected for interview at schools applying a threshold when they might not have

made the cut at schools scoring personal statements. A follow up to this study set out to forecast the impact on demographics if medical schools moved to a stronger use of the test (Tiffin, Webster et al. 2013). The authors modelled a move from factor approach to threshold approach across schools and suggested this would result in modest increases in admission of male applicants (approximately 5-7%) and a more significant increase in admission of state school candidates (approximately 10%).

The UKCAT research database allowed researchers to undertake a study in 2016 exploring the socio-economic status of successful and unsuccessful applicants to medicine (Steven, Dowell et al. 2016). Drawing on data across three application cycles, researchers had access to data from 32,964 applicants to medicine. Analysis raised many questions regarding contradictions between different measures of socio-economic class. However, it was clear that those from less affluent backgrounds were less likely to gain an offer to study medicine. Variation in outcomes between the countries of the UK and individual medical schools were also highlighted.

A further study using the UKCAT research database investigated whether the introduction and increase in graduate entry programmes had achieved the aims of widening access to medicine (Kumwenda, Cleland et al. 2018). Whilst some small differences in candidate populations were observed, the vast majority of applicants to these programmes remained from the highest socio-economic groups.

A later study drawing on UCAS data for *all* applicants to medicine contradicted Tiffin's earlier findings (Mathers, Sitch et al. 2016). The study, using UCAS data from 2004 to 2011, investigated what impact the introduction of UKCAT had on candidate demographics. The authors found that apart from a reduction in advantage to female candidates, the impact on other demographics (ethnicity, socio-economic class, school type) was minimal.

In an opinion piece Nicholson reflected on progress to widen participation in medical education (Nicholson 2018). The paper noted limited improvements and a significant lack of evidence as to what works in selection to widen access. The paper did note an increasing evidence base to support the use of contextual admissions.

An article focussing on the challenges of widening access at the Glasgow Medical School commented on actions taken to allow candidates meeting access criteria to achieve lower UKCAT scores (alongside lower grades required for school leaver qualifications) (Sartania, Haddock et al. 2018). The authors commented on the potential utility of including UKCAT Bursary entitlement amongst criteria to identify applicants from lower socio-economic backgrounds.

A large study drawing on UKCAT data (medical students from 18 UK medical schools who tested between 2007 and 2014) sought to identify any impact on widening access from changes to medical school selection periods over this time (Fielding,

Tiffin et al. 2018). The study combined UKCAT candidate data with information regarding how the test had been used at different medical schools. These data were used to identify where change in selection processes had taken place in individual schools. Data was than analysed to see whether such change impacted on candidate demographics. Change in selection processes did not appear to lead to observable differences in candidate populations, leading the authors to conclude that more radical interventions were required if progress was to be made around widening access.

5.3.6 The Candidate Experience

A feature in the BMA in the early years of UKCAT's development reported that the students would be calling on the Medical Schools Council to abolish the UKCAT at a conference taking place in 2008 (Samuel 2008, Lambe, Greatrix et al. 2016). The article criticised the test for lack of predictive validity and noted the error during 2007 testing which had led to the withdrawal of one subtest. Students were also concerned about the impact on widening access of an additional barrier to applicants.

A further feature in 2010 summarised progress in how the test was developing and research findings to date (mostly fairly contradictory findings regarding predictive validity) (Jaques 2010). The author concluded that whilst accepting selection was not an exact science, the UKCAT Consortium was trying to make the process fairer.

The student view was explored further in 2011 when a mixed methods study (questionnaire and focus groups) explored student views of the UKCAT (Cleland, French et al. 2011). Only 20% of respondents agreed UKCAT was useful. Students remained concerned about the test's lack of face validity, fairness, cost, use by medical schools and the impact of preparation on scores.

Fairness in terms of access to support, advice and preparation materials was explored in a single centre study in 2011 (Lambe, Waters et al. 2012). Differentials in access to support and advice were found to exist, linking directly to school type. The authors noted the need for school to be more informed about the test and how candidates ought best to prepare and hoped that the survey findings would inform future advice provided.

Building upon the findings in this study, during 2012 testing all UKCAT candidates were asked to complete a survey reporting how they had prepared to take the UKCAT (Lambe, Greatrix et al. 2016). An unpublished study (Lambe, Waters et al. 2013) reported that a 'support index' (derived from the outcomes) was positively associated with test performance. 82 UKCAT scale points separated those candidates with the highest support index from those with the lowest. Whilst this difference was relatively small, the authors noted that differentials in support almost certainly existed in relation to other aspects of the admission process emphasising the need to improve information available to candidates and their advisors.

A further unpublished paper (Lambe, Greatrix et al. 2016) from this study reported on candidates' views of the UKCAT. Candidates called for more information on how the test was used in selection (Lambe, Greatrix et al. 2016). Over half of candidates did not consider the test to be fair or a good way of measuring the attributes required to be a good doctor or dentist. Unsurprisingly, candidates who had scored better in the test were more likely to consider the test to be a fair way of discriminating between candidates. The majority of candidates (78%) agreed that the test itself was not too hard, but that timing created a greater challenge to candidates. There was greater acceptance amongst candidates that the SJT was relevant to the selection process.

An unpublished paper used survey data (linked to test results) to explore the use and effect of preparation materials. 16.1% of responders reported attendance at a school-based preparation course, 81.5% reported using the official UKCAT practice tests and 56%) reported using paid commercial materials.

Use of the freely available official UKCAT practice tests was associated with greatest impact on test performance (67.77 scale scores points). Attendance at a school-based preparation course (36.66 scale points) or using paid commercial materials (37.65 scale points) also significantly improved test performance though to a lesser degree. Performance in the test increased in proportion to time spent preparing. Score differences (compared with those not using these materials) were most observed in abstract reasoning and quantitative reasoning, suggesting perhaps these sections have the higher preparation effects.

Sub-group differences were observed both in test performance and in accessing preparation materials. Those reporting Asian or Black ethnicity were more likely to have attended a school-based preparation course, used free commercial materials or spent more than 20 hours preparing for the test (compared to White ethnicity). Again, compared to White students, those reporting Asian ethnicity were more likely to have accessed paid commercial resources. Those candidates in the most socio-economically deprived quintile were significantly less likely to use the official tests, paid commercial materials or prepare more than 20 hours for the test compared to the least deprived (compared to those in SEC 1). Similarly, candidates in receipt of a bursary were significantly less likely to report using paid commercial materials or to have prepared for more than 20 hours compared to those not in receipt of a bursary.

5.4 Summary of Findings

The systematic review of factors associated with success in medical school, commissioned by the Committee of Deans and Heads of Medical Schools (now Medical Schools Council) in 2000 (Ferguson, James et al. 2002), provides an insight into the volume of research regarding medical selection at this point. The search identified 117 articles of relevance. Of these, 51 articles dated from 1990 onwards, but only eight of these emanated from the United Kingdom. Whilst some of these

were really substantive studies, this does indicate a paucity of evidence underpinning selection at the time.

Undertaking this exercise not only introduced me to the rigour of the systemic review process but also gave me an opportunity to explore more rigorously the findings of this set of papers. This has given me a greater confidence in presenting the case around using the UKCAT in selection.

This chapter reports a significant increase in research into medical and dental selection since 2006 and a growing robustness in the evidence base upon which selectors can make decisions. In reviewing the evidence for this chapter, approximately 30 articles were identified since 2006 where the main focus of investigation is the UKCAT. A by-product of some of these studies has been a deeper dive into issues such as predictive validity (McManus, Dewberry et al. 2013, Mwandigha 2017) and range restriction (McManus, Dewberry et al. 2013) which have subsequently informed later studies. There are many other drivers around the creation of a more solid evidence base for selection and of course UKCAT cannot claim sole credit for the increase in activity.

Following the publication of the Schwartz report, an increased focus in the sector on widening access was driven by the creation of the Medical School Council's Selection Alliance and subsequent establishment in International Network for Researchers in Selection into Healthcare (INRESH). This shift in research focus moved the debate away from one reviewing the utility of specific selection tools to understanding and impacting on inequalities in selection. UKCAT's contribution to this debate has been an important one - a number of these critical studies were funded by UKCAT as part of a strategic plan to increase relevant research and research capacity. UKCAT also made available for the first time data to explore some of these emerging research questions.

This systematic review reported in Section 1 of this chapter supports the use of the UKCAT in selection; the test predicts performance in medical school. However, the relationship is small, and selectors ought not to use the test in isolation but alongside other selection criteria. The UKCAT Consortium should reflect on findings regarding the verbal reasoning subtest, which perhaps support consideration of this subtest separate to the total test score used most frequently by universities in selection. Further studies are required so that relationships between a changing/developing UKCAT and changing/ developing medical school curricula continue to be understood.

Today a good understanding exists of how selection to medicine and dentistry works, the differences in selection processes across institutions and the robustness of different selection criteria. At the same time, there is a much better understanding of who applicants (successful and unsuccessful) are, where they come from how they go on to perform in medical and dental school. The difficulties in identifying widening access candidates, attracting such applicants, supporting

them through selection processes and within training is much better understood. This growing understanding has led to a greater confidence in the use of contextual factors in admissions with an evidence base to support their use.

There remains a lack of strong evidence regarding preparation and coaching effects on test performance. The evidence reported above suggest that the strongest association with performance relates to use of the UKCAT freely available practice materials, with much smaller association with paid for materials. Whilst this is to some extent reassuring, the group differences in accessing resources and in test performance remain a challenge. More work is needed to better understand these differentials and to see to what extent better communications and advice can impact on these differences.

UKCAT was a key driver in the creation of UKMED which provides an opportunity for more significant and ambitious studies over time to further inform the development of the test and inform selection in medicine. There appear to be fewer opportunities to research dental selection which perhaps ought to be a future focus of the UKCAT Research Group.

Chapter 6: How has Selection to Medicine and Dentistry changed since 2006?

6.1 Introduction

This chapter presents information as to how selection to medicine and dentistry has changed since UKCAT was first delivered in 2006 and explores some of the external drivers around this change.

The first section in this chapter draws on data presented in two papers (Greatrix and Dowell 2020, Greatrix and McAndrew 2021) which used information that I had collected during annual interviews with admission tutors to document changes to selection methods at individual universities over time.

In the second section, an update on national policy relating to selection, looking at the role of organisations such as the Department of Health, General Medical/Dental Councils and Medical/Dental Schools Council have had in driving change in selection since 2006. This section also includes reference to some of the more relevant discussions at Medical Schools Council Admissions Deans meetings, providing a flavour of discussions which may have impacted on UKCAT development.

6.2 How has use of UKCAT in selection changed since 2006?

Traditionally there were three stages in the selection of students for UK medical and dental schools: an initial assessment of academic qualifications alongside a further assessment of qualities obtained from the Universities and Colleges Admissions Service (UCAS) application form (personal statements and references). The outcome of this stage would have usually identified applicants to be invited to interview. Whilst a small number of universities did not interview applicants in the past, all medical and dental schools now use an interview as part of selection. Selection processes are at the discretion of individual universities. Whilst the core approach to selection may be largely similar, differences exist between medical schools (Cleland, Dowell et al. 2012). Selection approaches remain fairly similar across dental schools (Cunningham, Patterson et al. 2019).

The UKCAT (www.ucat.ac.uk) was created in 2005 and first delivered in 2006. In 2021, 32 universities used the UKCAT as part of their selection processes. This included new medical schools created following the expansion of medicine student numbers (Higher Education Funding Council for England 2018).

The UKCAT Consortium set out to provide medical and dental schools with an additional selection tool, to assist in the challenge of discriminating between the large number of academically high achieving applicants. At the same time universities were looking for measures (over and above academic achievement and cognitive ability) to identify the traits necessary in applicants to make them good doctors and dentists. From the outset, the Consortium was interested in the extent to which the test predicted performance in medical and dental schools.

The UKCAT originally comprised four cognitive subtests (verbal reasoning, quantitative reasoning, abstract reasoning and decision analysis), providing four subtest scores (each with a scale score range of 300-900) which when totalled produced an overall score (range 1200 – 3600) for each candidate. The UKCAT Situational Judgement Test (SJT) (targeting the non-academic attributes of integrity, perspective taking, team involvement, resilience and adaptability) was introduced in 2013 with candidates allocated to one of four bands with Band 1 being the highest performing candidates. The Decision Analysis subtest was replaced by Decision Making in 2017. Further information regarding UKCAT test content is available on the UKCAT website (www.ucat.ac.uk) and in annually produced technical reports e.g. (Pearson VUE 2019).

UKCAT test scores provide standardised measures that medical and dental schools may use in selection but there is no explicit policy or recommendation made to those schools regarding usage. UKCAT Consortium members are informed on an annual basis about content, scoring and the statistical performance of the test to enable them to decide how best to use the test in their selection processes e.g. (Pearson VUE 2019).

Changes and trends in UKCAT's contribution to medical and dental student selection between 2007 and 2018 (inclusive) are described below. Information was obtained from admission tutors/officers during annual telephone interviews. Data obtained through these interviews represents a unique source of information regarding selection processes for medical and dental programmes more broadly.

6.2.1 Methods

Building upon a previous paper (Adam, Dowell et al. 2011), telephone interviews (n=23-26) have been undertaken with UKCAT Consortium medical and dental schools on an annual basis. Results are reported below for medicine (from 2007 entry to 2018 entry) and for dentistry (from 2011 to 2019) separately. Interviews used a standard questionnaire (Appendix E) that I had conducted each summer term. The interviewees were either admission tutors or administrators familiar with local selection processes. Interviews focussed on selection to the core programme for each school, which in most cases was a five-year undergraduate programme, although the study also includes schools that only delivered a graduate-entry programme. Results are not reported here for how schools used the test for other than their core programme such as gateway, accelerated and graduate entry programmes. Key points of each interview were noted, and a summary sent to each interviewee giving them an opportunity to make corrections. In most cases interviewees confirmed the document as a correct summary of the interview. On occasion, interviewees made minor changes to the document.

The Faculty of Medicine & Health Sciences Research Ethics Committee (University of Nottingham) confirmed that the nature of this research and its methodology falls under the category of service evaluation/systematic review and as such did not require formal research ethics approval. The committee confirmed that implied

consent by those individuals taking part in the annual interviews was sufficient on the basis that this is information routinely collected to inform the evaluation of the UKCAT (attached as Appendix F and G).

6.2.1.1 Categorisation of Use of the UKCAT in Selection

Use of the UKCAT in selection has previously been categorised as Borderline, Factor, Threshold and Rescue Methods (Adam, Dowell et al. 2011). This categorisation of use of the test has been utilised subsequently by others (Tiffin, Webster et al. 2013, Mathers, Sitch et al. 2016) to describe trends in use of the UKCAT over time.

Table 8 Categorising Use of the UKCAT in Selection

Method	Description
Borderline Method	Universities using this method used the UKCAT as an objective measure to discriminate between applicants lying at a decision borderline for either interview or offer. Some schools included this method in their selection 'toolkit' but did not always use it depending on outcomes at other selection stages. This method was generally only applied to a small number of applicants and as such defined as 'light touch'.
Factor Method	Where the Factor method was utilised, universities used weighted criteria to create a unique algorithm determining a score for applicants which could then be compared. The Factor Method was used most frequently to determine invite for interview and, on occasion, to make offers. Weighted criteria used to identify applicants for interview included academic scoring, UKCAT scoring (usually the UKCAT total score), personal statement scoring and university own questionnaires. Following interview, some universities weighted the interview score alongside academic, personal statement and other scores. An example of how the Factor Method is used in selection is provided below. Weighting and the range of scores for different criteria determined impact on selection outcomes. If, for example, academic score range was limited, then regardless of how low the UKCAT weighting, UKCAT may still have had significant impact on outcomes.
Threshold Method	Universities using a threshold required applicants to achieve a minimum UKCAT score to progress to the next stage of a selection process. Thresholds were most commonly used to identify those to invite for interview, often applied following an assessment of academic qualifications and/or other criteria. 'Actual' thresholds have been pre-determined and often published for the information of applicants. Actual thresholds may have been used to reduce the number of applicants for consideration at a further stage (e.g. to reduce the number of UCAS forms for scoring). 'Convenience' thresholds ranked applicants by UKCAT total score. A cut off score was then chosen which provided the N applicants required for interview. Applicants would not have been certain that

	their score would meet this requirement, although schools have on occasion published indicative scores to guide applicant choice. This method has been regarded as giving UKCAT a higher impact on outcomes than other measures. In some cases however, where cut off scores were low, the impact was less significant, screening out small numbers of applicants.
Rescue Method	The rescue method allowed universities to use high UKCAT scores to 'compensate' for a lower score in another part of the selection process, 'rescuing' applicants who might otherwise have been rejected. Overall impact of this use was light touch, affecting small numbers of applicants.

The following example describes how the use of the UKCAT in selection might have changed over time for a hypothetical university.

University X – Selection Processes for Main Medical Programme

In 2006, University X initially rejected those applicants not meeting a minimum academic threshold. The remaining applicants had their academic record (achieved and predicted) and their UCAS Form (personal statement and reference) scored. The scores were combined, contributing 50% each to a total first stage score. Applicants with the highest scores were invited to a traditional interview. Interviewees were then ranked and applicants with the highest scores made offers.

---

*The UKCAT was used for the first time in selection in 2007 and only to discriminate between applicants at a border line (for offer) after interview (**borderline use**), affecting decisions for four applicants.*

---

*This use of the UKCAT continued until 2011 when UKCAT was used as an additional criterion in selecting applicants for interview. Academic record contributed 50% towards the first stage score and UKCAT and the UCAS Form both contributed 25% (**factor use, 25% weighting**). The UKCAT continued to be used for borderline applicants after interview.*

---

*In 2016, the University ceased scoring the UCAS Form. The first stage score now comprised 60% academic record and 40% UKCAT (**factor use, 40% weighting**). In addition to the academic screen, applicants who had achieved Band 4 in the UKCAT Situational Judgement Test (SJT) were rejected without further consideration (**Threshold Method**).*

In the same year the University moved from traditional to multiple mini interviews (MMIs) and discontinued using the Borderline Method.

6.2.1.2 Data Extraction

Data from each questionnaire was extracted into an excel spreadsheet. There were some minor variations (e.g. some universities did not interview throughout this period) but selection processes were usually split across three stages: pre-screening, selection for interview, selection for offer.

6.2.2 Results - Medicine

The following figures show trends in the use of the UKCAT in selection to medical training since the first year of testing. The number of medical schools using the test has changed over this period with some schools having left the Consortium and others joining.

6.2.2.1 Invitation for Interview – Medical School Selection

This section describes how the first stage of medical school selection processes (to select applicants for interview) has changed since 2007. Each of the main uses of the UKCAT in selection (as described in Table 7) are discussed separately.

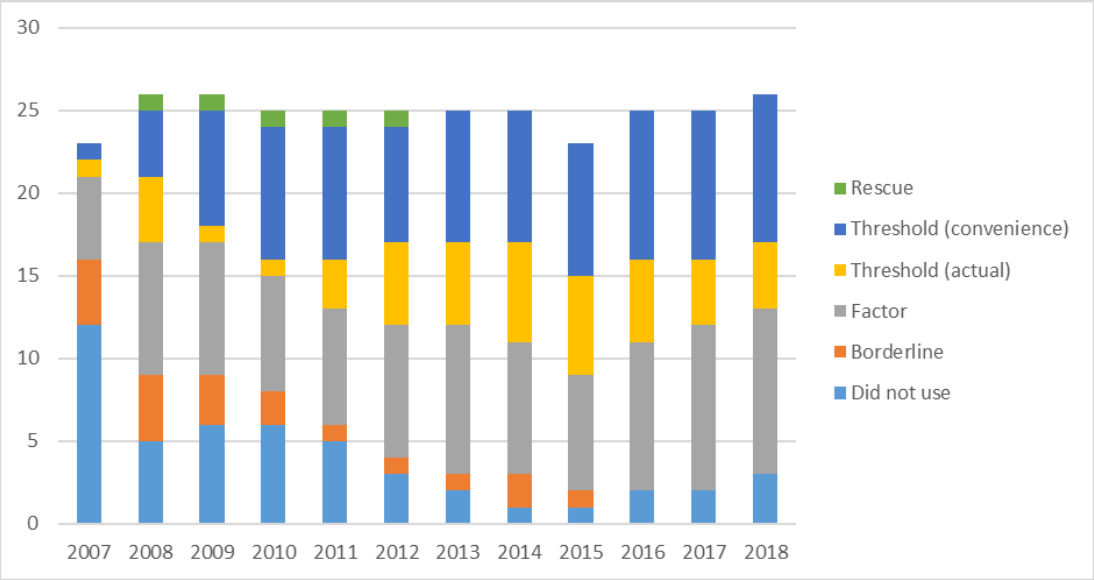


Figure 29 High Level Summary, Use of the UKCAT by Medical Schools: Invitation to interview

Figure 29 provides a high level summary of test use to select for interview by medical schools. In the first year of testing the majority of medical schools either made no use of the test at this stage (n=12) or only used the test to discriminate between borderline applicants (n=4). Five medical schools used the Factor Method. By 2018, only three medical schools were not using the test at this stage; 10 utilised the Factor Method; five applied an actual threshold and eight a convenience threshold. The growth in use over time is greater for schools using the Threshold Method. For 2018 entry no schools used the Borderline Method at this stage of selection.

The most notable trend has been the decline in the number of medical schools not using the test or solely using the test for borderline applicants; at the same time there has been an increase in schools (2008, n=7; 2018 n=13) applying a threshold.

The introduction of the UKCAT SJT in 2013 provided medical schools with a further criterion for use in selection. Most schools using the SJT at this stage excluded the lowest performers in this subtest (i.e. the approximate 10% of candidates that achieved a Band 4).

Some medical schools used the test in more than one way at this stage and these data are presented in Appendix H.

Threshold Method (Invitation for Interview)

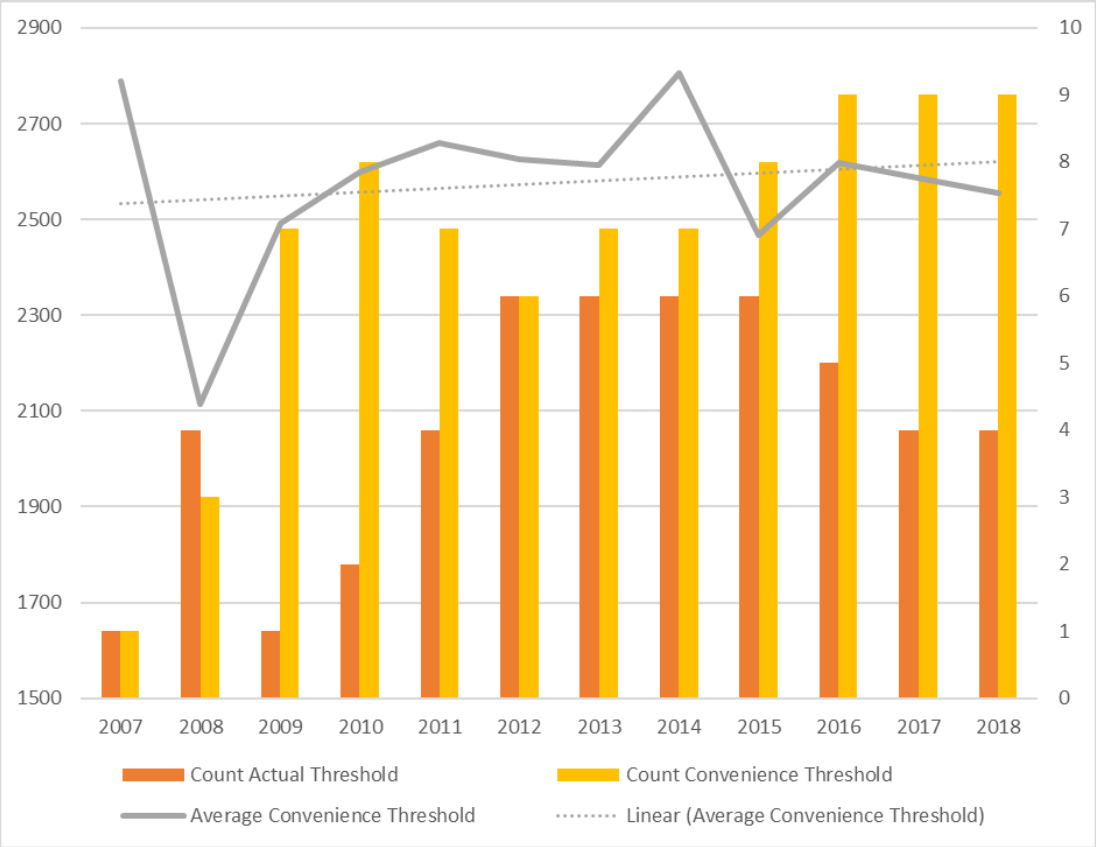


Figure 30 Threshold Method, Medical Schools (Invitation for Interview)

Figure 30 illustrates how the use of threshold scores changed over time and how these thresholds compare with mean average test scores. Additional data recording mean thresholds (actual and convenience) are provided in Appendix I.

Four medical schools used the test to pre-screen applicants using a pre-defined 'actual threshold' score in 2018. Whilst this number is higher than in 2007 (n=1), the maximum number of medical schools using this method was six (2012-2015). The small increase in average actual threshold scores used by medical schools in this manner can largely be accounted for by a rise in mean average total scores over this period. Most medical schools using the test in this way screened out lower

performing applicants although there were medical schools which applied high actual threshold score in some years (2700+).

In 2018, nine medical schools used a 'convenience threshold' UKCAT score to select applicants for interview. This number rose steadily over time from one medical school in 2007. The average convenience threshold has also increased over time. In 2018 this figure (2544) was only slightly higher than the mean average UKCAT score (2540); in previous years (with one exception) the convenience average score was higher than the overall mean average.

Convenience thresholds applied by individual medical schools varied over time because decisions made each year by schools relate to each university's 'gathered field'. That is, the number of applicants a university may want to interview depends on total application numbers and the quality of those applicants. This is further informed by previous experiences in relation to how many interviews needed to be undertaken to arrive at the correct number of offers. This resulted in some schools applying relatively high thresholds at this stage (e.g. three medical schools had a threshold score 2640+).

Factor Method (Invitation for Interview)

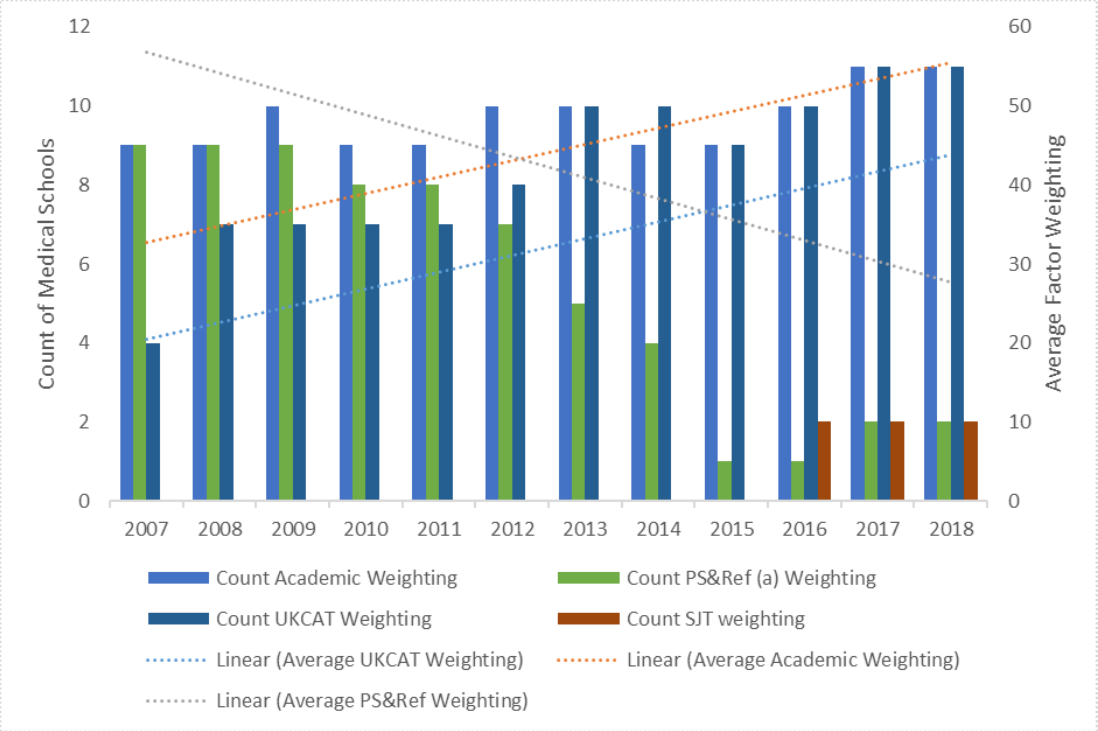


Figure 31 Factor Method, Medical Schools (Invitation for Interview)

Figure 31 illustrates the number of medical schools using the Factor Method to identify applicants for interview and the average weighting applied for each criterion in each year. Additional data recording mean average criteria weightings are provided in Appendix I.

There has been a significant shift in how this method has been used to select applicants for interview. The number of schools weighting academic scores has increased marginally. In 2007, nine medical schools used the personal statement and reference as a weighted criteria in selection for interview. However, by 2018, this number declined to only two medical schools. In 2007, the personal statement and reference score accounted for (on average) 58% of the weighting for selection for interview and was the highest % weighting used until 2012. By 2018, this figure had declined to 36%. From 2013, the highest % weighting was for academic scoring. In 2007, seven medical schools used the UKCAT as a weighted criterion at this stage; in 2018 this increased to 11. During this same period, the weighting applied to the UKCAT increased on average from 26% to 39%. In 2018, a number of schools applied a high weighting to the UKCAT, with one medical school using the test as 66% of the weighting to determine whether to invite applicants to interview.

Of the 11 universities using the UKCAT in this way, there has been a clear increase in weighting for the majority (n=7). For most of these universities this step change occurs at the point where reliance on the use of personal statements/references reduces or disappears. In four cases the weighting for the UKCAT has remained steady or reduced over time.

In 2018, two medical schools used the SJT subtest as a weighted criterion at this stage in selection.

6.2.2.2 Making an Offer – Medical School Selection

This section describes how the second stage of medical school selection processes (to select applicants for offer) has changed since 2007. Each of the main uses of the UKCAT in selection (as described in Table 7) are discussed separately.

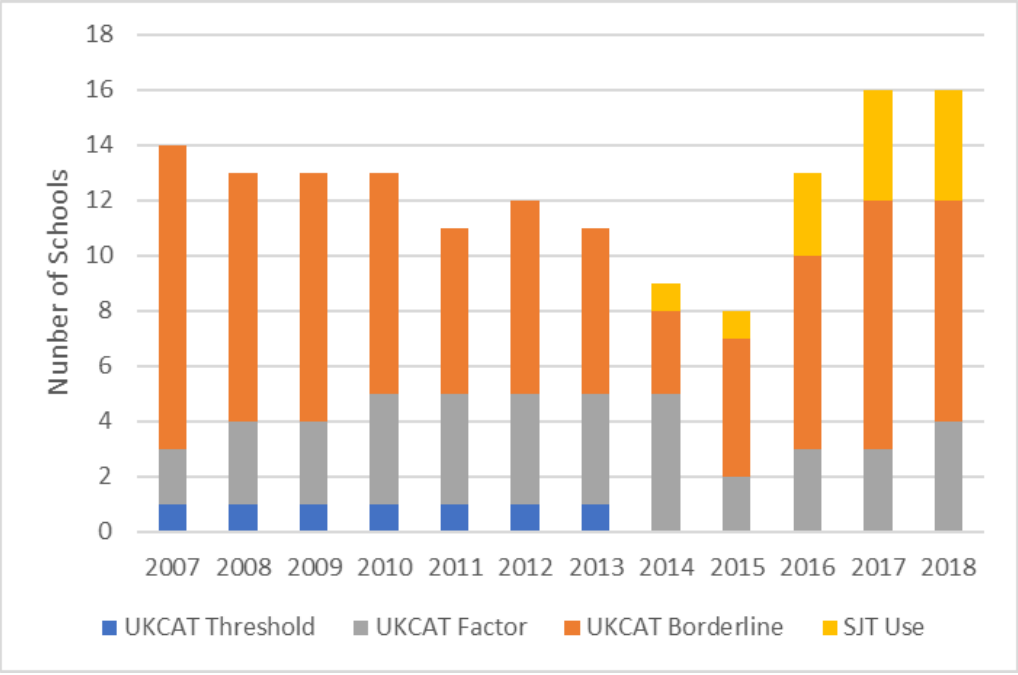


Figure 32 Use of the UKCAT: Making an offer, Medical Schools

In the first year of testing, 11 medical schools used the test to discriminate between borderline applicants at offer stage. A further two medical schools used the Factor Method at this stage whilst one school applied a UKCAT threshold.

By 2018, use of the UKCAT at offer stage remained relatively limited; eight medical schools used the UKCAT within a Factor Method (four medical schools using the cognitive test total score and four schools using the SJT). A further eight medical schools used the test to select between applicants at a borderline after interview.

There has been some recent growth in use of the test at the offer stage with medical schools identifying this as an appropriate place to include the SJT within selection. In addition to the four schools using the SJT as a weighted criterion (alongside interview outcomes), a further two medical schools used the SJT to discriminate between borderline applicants. One school used SJT Band 4 as a potential red flag in MMIs and one applied an SJT threshold for offer.

Factor Method (Making an Offer)

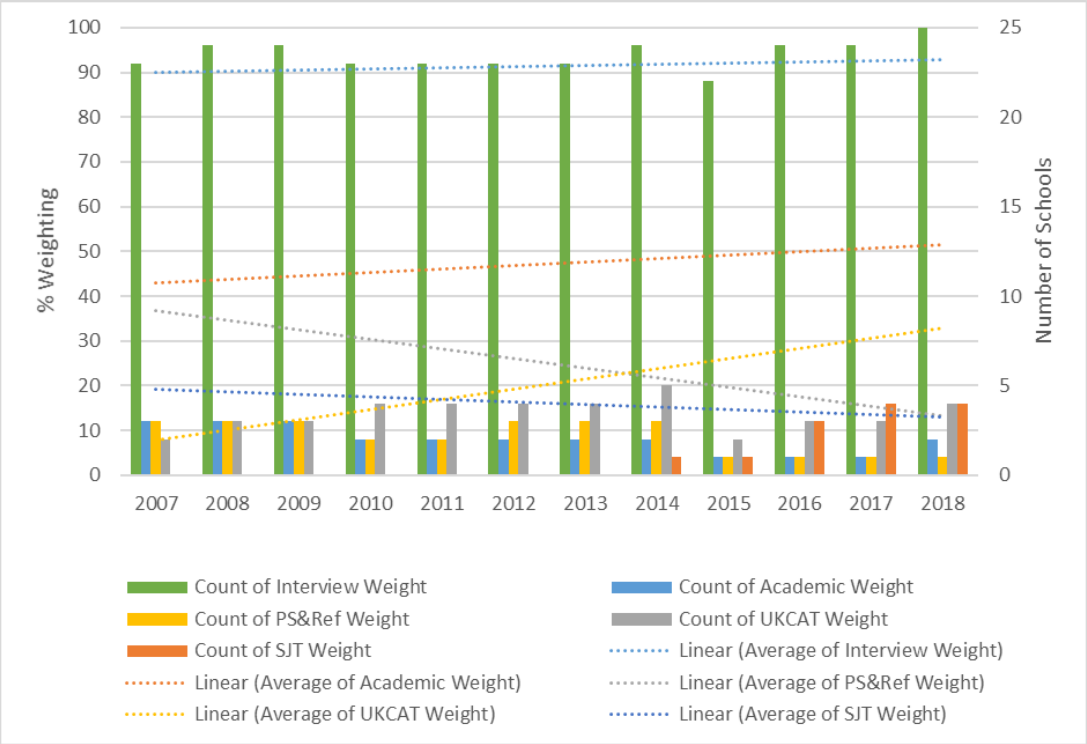


Figure 33 Factor Method (Making an Offer), Medical Schools

Figure 33 illustrates the extent to which, as would be expected, the interview continues to dominate decision making regarding offers to medical applicants.

In 2018, only four medical schools used the UKCAT cognitive test scores as a criterion for making an offer to applicants and this number changed little since 2007. The average weighting of the UKCAT cognitive tests during this period increased from 8% to 27%. Only one medical school has used the personal statement and reference at this stage since 2015.

In 2018, three medical schools used the UKCAT SJT as a criterion at this stage.

Borderline Method (Making an Offer)

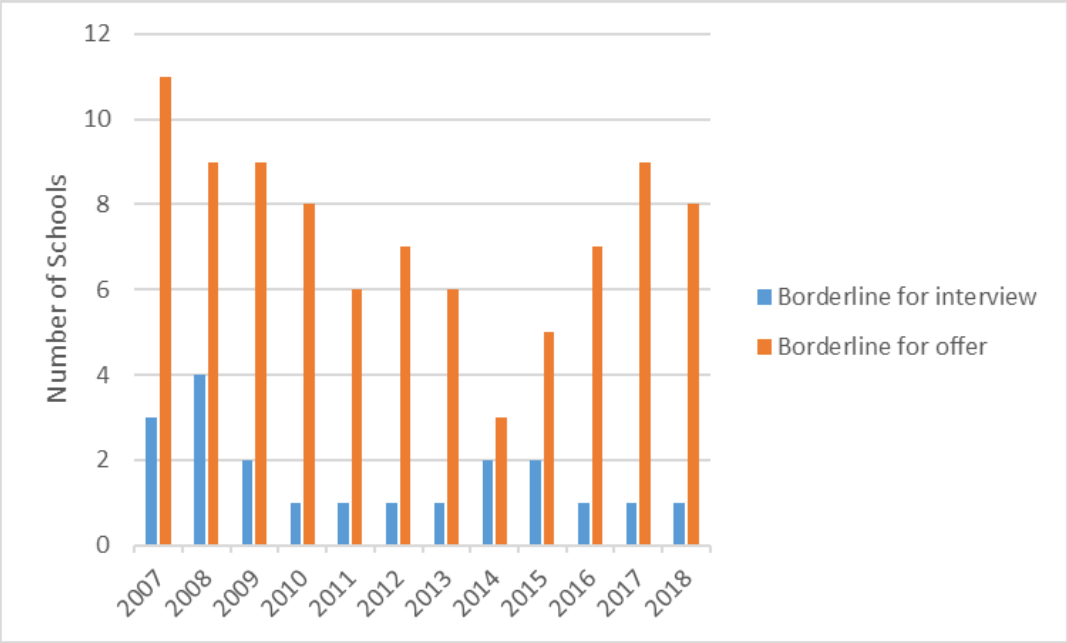


Figure 34 Borderline Method Use of the UKCAT, Medical Schools

In 2018 nine medical schools used the UKCAT to discriminate between applicants at a borderline often in conjunction with other uses. In UKCAT's early years there was greater use of the test in this way, reflecting a more cautious approach to its use in selection. Two of the medical schools using the test for borderline applicants in 2018 used the SJT to discriminate at this stage.

Rescue Method

It is only possible to identify two medical schools definitively using the Rescue Method in the early years of UKCAT. One medical school (in one year) invited applicants to interview on the basis that a high UKCAT score might mitigate for a low scored personal statement and reference.

The Rescue Method use, depending on how and when it is applied feels very similar to both the Factor Method (if weighted) or Borderline Method (if introduced after the majority of applicants had been identified for interview).

6.2.2.3 Use of the Situational Judgment Test in Medical School Selection

The UKCAT SJT was introduced operationally in 2013 testing. Medical schools have been cautious about using the subtest in selection. In 2018 four medical schools excluded applicants who had achieved a Band 4 SJT. Two schools applied a weighting to the SJT in selecting applicants for interview. Eight schools used the SJT at offer stage, four as a weighted criterion (some within an MMI), two for borderline applicants, one as a threshold and one as a marker of concern within MMI processes.

6.2.2.4 Multiple use of UKCAT scores in Medical School Selection

UKCAT scores were used by some medical schools in more than one way and at different stages in selection processes. The following uses of the UKCAT were observed in the 2018 admissions cycle (note number of medical schools = 26).

Table 9 Use of the UKCAT by Medical Schools, 2018

Use of the UKCAT (to select for interview)	N	Use of the UKCAT (at offer stage)	N
Cognitive Subtests Weighted (Factor)	12	Cognitive Subtests Borderline	6
Cognitive Subtests Threshold (Convenience)	9	Cognitive Subtests Factor	5
Cognitive Subtests Threshold (Actual)	5	SJT Weighted (Factor)	4
SJT Threshold	4	SJT Borderline	2
SJT Weighted (Factor)	2	Other	2
Cognitive Subtests Borderline	1		

In 2018, eight universities used the test at only one point in their selection process; 12 universities used the test in two different ways; four universities used the test in three different ways and two universities used the test at four points in their selection process. The most common 'secondary' use of the UKCAT was to distinguish between applicants at borderlines for offer (n=8) and the application of an additional SJT threshold to select for interview (n=4).

6.2.2.5 Other Findings Medical School Selection

The interviews provide a rich source of data regarding medical school selection. In this section various additional findings are commented on.

Widening Access

All universities in the UK are expected to have in place strategies to widen access to higher education by increasing participation of students from under-represented groups with a particular focus on those from lower socio-economic backgrounds. Some medical schools offer programmes specifically aimed at these applicants. Schools may also flag (based on demographic data provided) widening access applicants at the point of application, allowing those involved in selection to consider this group separately.

Information was not routinely collected in detail regarding use of the test for widening access programmes and/or applicants. However, respondents were asked to comment on the extent to which use of the test varied between applicant groups. At the same time, some schools provided additional information regarding programmes other than their main programme.

In the first years of testing very little information was reported by medical schools regarding adjustments to processes for widening access applicants. In 2008,

information was only reported by five schools. In 2018, all but seven schools referred to a specific aspect of their selection processes adjusted for widening access applicants. There were distinctions made by some between applicants coming through recognised widening access programmes and other widening access applicants. The detail as to how these two different applicant cohorts were treated had not been provided in enough detail for distinctions to be made here.

Some medical schools did not require identified widening access applicants to take the UKCAT or did not use the test for widening access applicants (flagged at application) (2018, n=5). 12 schools adjusted criterion scoring for widening access applicants. Of these, eight adjusted academic scores and six adjust UKCAT scores. In most cases this involved adjusting an academic and/or UKCAT score for identified widening access applicants to increase their chances of being invited to interview.

Types of Interview

In 2007 only two (out of 25) medical schools used MMIs as the final stage of their admission processes. By 2018, 21 (out of 26) medical schools used MMIs. The significant shift to the use of MMIs took place from 2013 onwards.

Applicant Numbers

When completing the questionnaires, admission tutors provided information regarding the number of places on their programme(s), applications, interviews and offers made. These data were collected at a global level (and therefore some institutions may have included more than one programme). There are missing data in the early years when this was not collected systematically. However, the data reveal some interesting trends:

- Between 2010 and 2018, the number of reported places available at UKCAT Consortium universities increased from 5,208 (universities n=25) to 6,226 (n=26) (+20%).
- There were fluctuations in reported applications (and changes in Consortium membership); the increase in reported applications between 2010 and 2018 was only 1,367 (+2.4%). In 2018 there were 57,543 applications reported to UKCAT medical programmes (this figure includes applicants making multiple applications).
- The ratio of applications to places fell (from 10.8 to 9.2).
- The number of reported interviews increased significantly from 14795 to 23057 (+56%) (this will include applicants being interviewed by more than one university).
- The ratio of interviews to places increased (from 2.8 to 3.7).
- The number of offers made increased from 8,952 to 14,146 (+58%).
- The ratio of offers to places increased (from 1.7 to 2.3).

6.2.2.6 Summary of Medicine Results

Over time there has been a significant shift in medical schools to use the UKCAT in selection with greater emphasis. In 2018, 23 out of 26 medical schools used the Threshold or Factor Method to select applicants for interview (2008, 16 out of 26). In 2018, the mean average weighting of the UKCAT to select applicants for interview was 39% (2008, 22%). By 2018, four medical schools used an actual UKCAT threshold to screen applicants at an early stage in their selection processes. A further nine medical schools used a convenience threshold score to select applicants for interview. Whilst threshold scores used by medical schools increased little over time, average threshold scores remained around the mean average of the test. In 2018, 18 universities used the test at more than one stage in their selection processes.

The mean average weighting for personal statements and references declined over this same period to 36% (2008, 58%) with only three universities still using this criterion at this stage (2008, n=9). It is of further interest to note that in 2008, 16 schools either weighted or scored personal statements as part of their selection processes. In 2018, this figure was just three.

The mean average weighting for the academic record to select for interview has increased to 51% (2008, 39%).

6.2.3 Results – Dentistry

In 2011, 10 of the 14 UK dental schools used in the UKCAT in selection (University of Aberdeen, Queen Mary University of London, Cardiff University, University of Dundee, University of Glasgow, Kings College London, University of Manchester, University of Newcastle, University of Sheffield, Queens University, Belfast). In 2013, Plymouth University joined the UKCAT Consortium, the Universities of Birmingham and Liverpool joining in 2016 and the University of Bristol in 2017.

6.2.3.1 Invitation for interview – Dental School Selection

This section describes how the first stage of dental school selection processes (to select applicants for interview) has changed since 2011. Each of the main uses of the UKCAT in selection (as described in Table 7) are discussed separately.

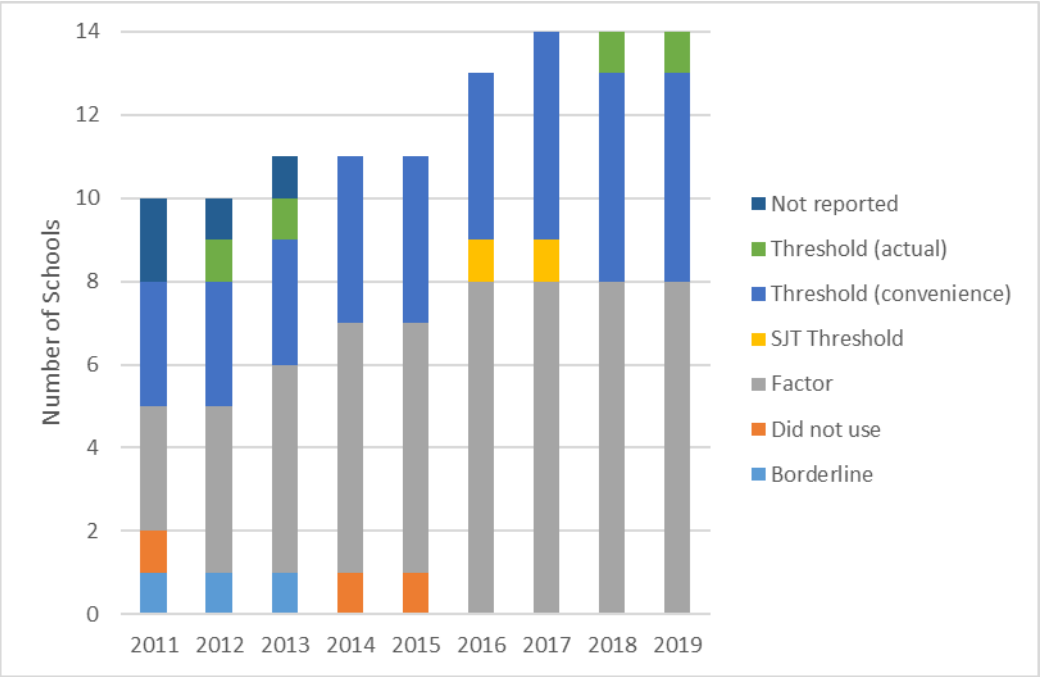


Figure 35 Use of the UKCAT: Invitation for Interview, Dental Schools

Figure 35 reports the main use of the UKCAT to select dental applicants for interview since 2011. Some dental schools used the test in more than one way at this stage and this is described further in Table 10. Whilst the number of dental schools using the test has grown, the balance in main use across factor and threshold methods has not changed significantly.

Factor Method (Invitation for Interview)

The most popular application of the test in 2019 is factor use (n=8).

Table 10 Factor Use of the UKCAT (Selection for Interview): Number of Dental Schools and mean average weighting

	Factor							
	Academic Score		**Personal Statements**		**UKCAT (Cognitive)**		**UKCAT (SJT)**	
	Schools	Factor Weighting	Schools	Factor Weighting	Schools	Factor Weighting	Schools	Factor Weighting
	N	%	N	%	N	%	N	%
2011 (N=3)	1	45%	1	36%	1	19%		
2012 (N=4)	3	59%	2	31%	3	20%		
2013 (N=5)	4	65%	2	33%	4	19%		
2014 (N=6)	5	66%	1	25%	5	29%		
2015 (N=6)	5	69%			5	31%		
2016 (N=7)	5	67%			5	31%	1	9%
2017 (N=8)	6	61%	1	50%	7	27%	1	11%
2018 (N=8)	6	61%	1	50%	7	27%	1	11%
2019 (N=8)	6	61%	1	50%	7	30%	1	11%

Table 9 presents results from dental schools using a factor approach to select applicants for interview and the mean average weighting applied for each factor in each year. Where percentage weightings were not provided by schools, these could not be included in the mean averages.

The use of academic scores remains the highest weighted factor used to select applicants for interview. Weighting of academic scores has increased over time but has been relatively consistent since 2013. A maximum of two dental schools used personal statements as a factor in selection during this period. The mean average factor use of the UKCAT increased from 20% to 30%. The range of UKCAT factor weighting was relatively small (19% - 31%). Over this period three universities shifted their weighted use of the UKCAT upwards whilst in the remaining four dental schools there has been little change.

Threshold Method (Invitation for Interview)

In 2019 one dental school used a pre-defined 'actual threshold' as a pre-screening tool prior to using academic scores to select applicants for interview.

A further five dental schools used a 'convenience threshold'. It should be noted here that convenience thresholds applied by individual dental schools varied over time because decisions made each year by dental schools related to the number and quality of applicants in any application cycle. A university will identify the desirable number of interviews based on previous experience and conversion rates (rates of acceptances to offers) in particular.

Convenience thresholds 'scores' used by dental schools since 2011 have not varied greatly with the mean average ranging from 2407 to 2545 (mean 2465, SD 53). Apart from one year (2013), the mean is below the mean average candidate score; schools are essentially screening out lower performing applicants.

Other Use (Invitation for Interview)

In 2011-2013 one dental school used the test at this stage to consider borderline applicants. No dental schools used the UKCAT rescue method.

6.2.3.2 Making an Offer – Dental School Selection

A small number of dental schools used the UKCAT at offer making stage. In 2011, five out of ten dental schools used UKCAT at this stage with three dental schools using the test to discriminate between borderline applicants and two dental schools using the overall test score as a determining factor. In 2019, five (out of 14) dental schools used the test at this stage; four for 'borderline' applicants and one using SJT scores.

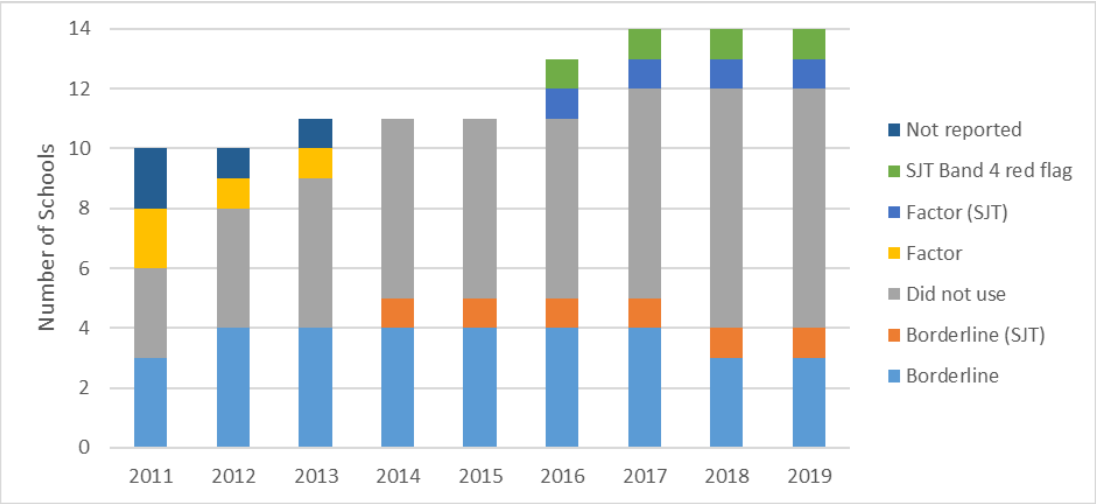


Figure 36 Use of the UKCAT: Making an offer, Dental Schools

6.2.3.3 Increased Use of the Situational Judgment Test in Selection

The UKCAT SJT was introduced operationally in 2013. Dental Schools have been cautious in the use of this subtest during selection and use of the SJT remains relatively light touch. In 2019, five dental schools used the SJT at some point in their selection processes. Two dental schools utilised an SJT threshold (rejecting SJT Band 4 applicants), one dental school applied a factor weighting to select applicants for interview. In two dental schools, SJT results contributed to interview outcomes.

6.2.3.4 Multiple use of UKCAT scores

Since 2011 some Dental Schools have used the test in more than one way during their selection processes and this is illustrated for 2019 entry in Table 10. During this cycle, five Dental Schools (D, E, I, M, N) used the test at only one point in their selection processes; 7 Dental Schools (A, B, C, F, G, H, K) used the test at two points; 2 Dental Schools (J, L) used the test at three points.

Table 11 Use of the Test by Dental Schools (anonymised), 2019

	Use for Selection for Interview						UKCAT Use for Offer		
School	UKCAT Total Score			SJT		UKCAT	SJT	UKCAT	SJT
	Threshold (Actual)	Factor	Threshold (Convenience)	Factor	Threshold	Borderline	Factor	Borderline	Borderline
A	X	X		X				X	
B			X		X				
C		X						X	
D			X						
E		X							
F		X				X			
G	X	X					X		
H	X				X				
I		X							
J			X					X	
K		X							
L			X						X
M		X							
N			X						

6.2.3.5 Other Findings Dental School Selection

These annual interviews have provided a rich source of data regarding dental selection. In this section various additional findings are commented on.

Widening Access

In the first few years of testing very little information was reported by dental schools regarding adjustments to processes for widening access applicants. In 2019, nine of the 14 dental schools referred to a specific aspect of their selection processes which had been adjusted for widening access applicants. Five dental schools adjusted required UKCAT scores for widening access applicants (alongside, in some cases, adjustments for A-level or equivalent qualifications).

Types of Interview

There has been a gradual shift towards greater use of Multiple Mini Interviews (MMIs) across dental schools with 10 out of 14 dental schools using MMIs in 2019.

6.2.3.6 Summary of Dentistry Results

Since 2011, there has been a shift to use the test more robustly in dental student selection, most notably as a factor method to identify those suitable for interview. In 2019, eight dental schools used the test in this way; on average UKCAT was weighted to contribute a value of around 30% in selection decisions.

The selection landscape for dentistry remained relatively stable. Whilst during the period of the research there had been significant change in medical school student selection where there have been drivers for change particularly around a shift away from the use of personal statements due to the lack of a credible evidence base (Ferguson, Sanders et al. 2000, Husbands and Dowell 2013, Tiffin, Mwandigha et al. 2016). Very few dental schools used personal statements during this period as a significant element of selection over this period.

6.2.4 Strengths and Limitations

This comprehensive summary of entry requirements to UKCAT Consortium medical and dental schools is a unique and valuable dataset. The completion rate is excellent, but it is limited to those schools that use the UKCAT, which has in itself shifted over time. Unfortunately, there is no information for schools that use or have used alternative admission tests or no admission test. The study focussed generally on UK applicants to the main programme offered by each medical or dental school which in most cases will be an undergraduate five-year course. There will be further variation in requirements for some subgroups of applicants (e.g. widening access, international) and other programmes (e.g. access, graduate entry); this would need to be investigated separately to determine influences or trends.

Information reported regarding medical applicant numbers provides a snapshot of fluctuations over time but ought to be treated with caution. Data was not always reported consistently and may have been interpreted differently by schools.

6.3 Drivers for change: UK selection to medicine and dentistry since 2006

This section seeks to describe some external drivers in existence since 2006 which may have contributed to the changes in medical and dental selection described in the previous section.

6.3.1 National Reports and Discussions

Drawing upon evidence from the Schwartz report (Schwartz 2004), the Medical Schools Council updated the **guiding principles for the admission of medical students** in 2006. The changes had the support of the BMA and Department of Health. The document highlighted the need for selection criteria to include core academic and non-academic qualities. A significant number of potential desirable traits were highlighted including honesty, integrity, communication skills, the ability to make decisions under pressure, ability to cope with stress, teamwork, curiosity, creativity, initiative, flexibility, and leadership. The report went on to stress the need for transparency in selection processes and the need to welcome diversity amongst applicants.

Meeting of the Medical Schools Council Admission Deans May 2007

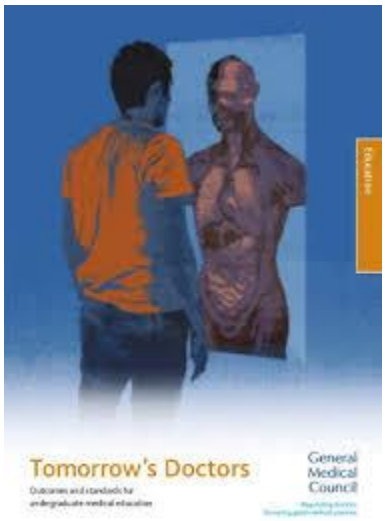
Topics for discussion included impact of UKCAT on applications to UK Medical Schools; Schwartz and the Supporting Professionalism in Admissions programme.

Meeting of the Medical Schools Council Admission Deans May 2008

Professor Chris McManus (UCL) presented 'What can A levels and aptitude tests tell us?' focussing on findings from a number of studies. He concluded that past behaviour was the biggest predictor of future behaviour, with GCSEs and A Levels the most effective predictors of performance in Higher Education. Aptitude tests (including BMAT and GAMSAT) had not yet demonstrated predictive validity. Accepting the need to select between high performing A level students, Professor McManus believed that this should be done by testing job content, job related performance or subject knowledge.

Professor David James (University of Nottingham) presented 'Uses and abuses of the UCAS application' describing outcomes from a study aiming to identify aspects of the personal statement which might predict performance at medical school. He concluded that the correlation between university performance and conscientiousness was higher than academic performance. Members noted that external influences on personal statement writing were leading some schools to cease to use this within their selection procedures.

A further presentation from the Australian Medical Schools Outcome Database described a project to collect data linking medical education outcomes, postgraduate outcomes and demographics.

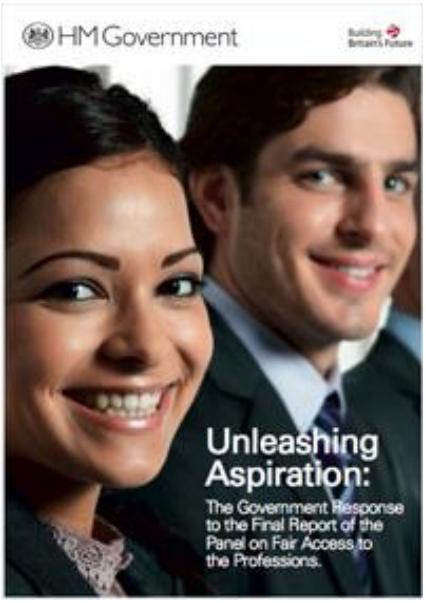


A further review of outcomes and standards in medical education reported in 2009 (General Medical Council 2009) building upon previous publications.

The themes within the document relating to selection strengthened messages in previous publications. Selection processes ought to be open, objective and fair and to achieve this medical schools needed to publish clear information regarding selection processes.

Selection criteria needed to include both the personal and academic qualities needed in a doctor. Selectors ought to be trained and promote equality and diversity. Processes should recognise best practice, be valid and reliable and based on relevant research.

In 2009**, Unleashing Aspiration (Milburn 2009)**, a report from the Panel on Fair Access to the Professions was published.



The panel, led by Alan Milburn MP had been established to advise on how professional careers could be made more open and in doing so contribute to increased social mobility in the UK. The report was extensive, containing numerous recommendations which included a significant focus on improved school outreach from universities.

Individual professions were called upon to focus on solutions in their particular areas and to review current practice.

The report commended work in medicine to recruit from ethnic minorities. However, whilst noting some successful schemes to widen access, it also reported that medicine ranked alongside law as the most socially exclusive profession.

The study commented on the potential of admission tests to provide additional information regarding applicant potential.

A report, published by the Medical Schools Council (MSC) provided guidance to medical schools in relation to professional behaviour expected of medical students and expectations regarding fitness to practice mechanisms (Medical Schools Council 2009). Whilst admissions criteria were not considered within the document, the

explicit focus on behaviours may have encouraged medical schools to review the traits being measured within selection procedures.

Meeting of the Medical Schools Council Admission Deans April 2009
Dr Sandra Nicholson (QMUL) presented on the role of UKCAT and efforts being made to measure non-academic attributes within the test.
Janet Graham updated members on work being undertaken by Supporting Professionalism in Admissions (SPA).

The 2006 guiding principles for selection were revisited in a further report from the MSC (**2010 Guiding Principles for admitting medical students (Medical Schools Council 2010)**), with the authors drawing upon criteria drawn from Tomorrow's Doctors 2009, the Schwartz Report and Unleashing Aspiration. Tomorrow's Doctors 2009 required medical schools to 'demonstrate open, fair and objective processes for student selection'. The principles reinforced the need to select those with appropriate attributes as outlined in the Role of the Doctor Consensus Statement and to address fitness to practice issues. Transparency of admission processes was further emphasised as was the need to engage with activities that would widen participation in the profession.

Meeting of the Medical Schools Council Admissions Deans April 2010
Topics for discussion included an update from SPA and widening participation.
Dr Sandra Nicholson (QMUL) updated delegates on developments within UKCAT, highlighting current research priorities.

Meeting of the Medical Schools Council Admissions Deans April 2011
SPA had conducted a review of admission tests and issued guidance on what makes a good admission test.
Sandra Nicholson (QMUL) updated delegates on progress with UKCAT noting plans for a longitudinal study and the decision to no longer offer subtest 5 (non-cognitive) until further analysis had been done on the data.
The GMC attended to report on a review of the GMC's role in selection.

The report on **Fair access to professional careers – a progress report by the Independent Reviewer on Social Mobility and Child Poverty (Cabinet Office, 2012)** noted that medicine had lagged behind other professions in responding to the original report and called for a step change in approach. The report cited particular issues with applicant access to work experience and noted being unable to see any systematic effort to address this inherent unfairness. Using HESA data, the report

noted that between 2002/3 and 2010/11, the proportion of medical and dental undergraduates from private schools and those from lower socio-economic groups had barely changed. The authors also noted that whilst a number of universities were active in schemes to improve access, little evaluation of these schemes had taken place and there was no agreed definition of 'success'. The report went on to urge selectors to make the use of contextual data as the norm.

The report referred specifically to the use of the UKCAT by 26 of 31 medical schools. Whilst accepting that such a test might theoretically help applicants from lower socio-economic groups, the authors were fairly sceptical. They noted the Sutton Trust claim that 'there is some evidence that these tests and interviews can place bright non-privileged pupils at a disadvantage because their schools do not have the capacity to prepare them for the test and may be unfamiliar with the new test format.'

In 2012, the General Medical Council commissioned a literature review to examine the available evidence on the effectiveness of different methods used by medical schools to select students and to examine the available evidence on the effectiveness of widening access initiatives used by medical schools to promote fair access **(Cleland, Dowell et al. 2012)**. The paper went on to be used to inform the **Selecting for Excellence** work undertaken later by the MSC. The study included a specific section regarding aptitude testing. In their conclusions, the authors noted that whilst validity evidence was mixed, aptitude tests appeared to have small but real incremental validity in predicting performance in medical school. The authors drew attention to preliminary evidence that aptitude tests could impact positively on widening access (Wright and Bradley 2010, Tiffin, Dowell et al. 2012) and noted the small impact of preparation on performance in UKCAT.

This study was wider ranging, focussing on all aspects of selection for medical education. Other key findings included:

- That whilst the 'academic record' provided the strongest correlations with on-course academic outcomes, there was clear evidence of a significant socio-economic bias.
- Research evidence supporting the use of personal statements and other autobiographical submissions was limited with greater threat of such criteria being significantly influenced by third parties such as parents and school advisors, thus creating again socio-economic biases.
- Well constructed Situational Judgement Tests were among the best and most valid methods of selection, less susceptible to coaching with lower adverse impacts on minority groups.
- Consensus appeared to be emerging regarding the 'reliability, validity and acceptability of Multiple Mini Interviews' and that further studies were needed to explore the potential of Selection Centres.

A significant part of the review focussed on evidence regarding the effectiveness of widening access initiatives in medical schools. The authors noted the lack of thorough evaluations of pre-entry programmes. Whilst it is probable that these programmes did impact on individual applications, the number of widening access applicants remained small with little change in socio-economic profiles. Further work was required to investigate the impact of different selection procedures on widening access and on the use of contextual data.

Meeting of the Medical Schools Council Admission Deans April 2012

Representatives from Cambridge Assessments (BMAT), GAMSAT and UKCAT outlined how aptitude tests were being used. In discussion members noted that no single test had emerged as being more valid than others and that the incremental validity of the tests was not well understood. The tests allowed a more granular approach in selection and that tests of personal qualities were desirable.

Professor Sir Bruce Keogh (medical Director NHS England) presented to delegates regarding value based selection in the NHS.

A discussion took place regarding diversity in selection processes and the potential for greater convergence. It was felt that a reliable test of values would help achieve greater consistency.

Meeting of the Medical Schools Council Admissions Deans April 2013

A summary of the GMC study 'Identifying best practice in the selection of medical students' was presented, focussing in particular on widening access.

A further presentation took place regarding value based selection.

Professor Fiona Patterson (Work Psychology Group) gave an overview of situational judgement tests.

In July 2013, the Medical Schools Council launched the **Selecting for Excellence project** in response to criticisms in the Fair Access Reports and the outcomes from a GMC literature review into best practice in selection. This early report (**2013 Selecting for Excellence end of year report (Medical Schools Council 2013))** outlined the scale of the challenge facing medicine. The project identified four work streams:

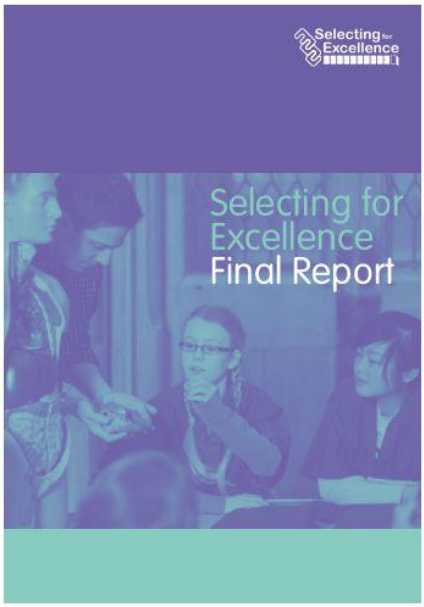
- Widening Participation
- The Role of the Doctor
- Selection Methods
- Evidence Base

Using data, including some provided by UKCAT (Cochrane 2013), the report outlined the demographic background of medical students, doctors and applicants. Whilst

the data supported claims that applicants to medical schools were drawn from higher socio-economic groups, the report noted inconsistencies within the data with particular conflicts noted regarding the use of NS-SEC, IMD and POLAR markers. The report noted the development of the UK Medical Education Database (UKMED) and the potential of linked datasets to help evaluate selection methods in the future.

Meeting of the Medical Schools Council Admission Deans May 2014

Delegates received an overview of the Selecting for Excellence report. A discussion regarding value based recruitment took place.



The final report from SEEG in 2014 included a significant number of recommendations in each work stream (Medical Schools Council 2014). The report concluded that whilst insufficient evidence existed at the time to create a national framework for selection, there was enough evidence to advise medical schools to move towards processes combining academic attainment with performance in aptitude tests and MMIs. By 2018, all schools in the UK were using an admission test as part of their selection processes and as reported above, most used MMIs. Despite this report providing advice to medical schools, dental schools too have moved towards processes combining academic attainment with performance in aptitude tests and multiple mini interviews (MMIs).

The Medical Schools Council went on to establish the Selection Alliance charged with implementing these recommendations. Recommendations of particular relevance to UKCAT were:

- The need to undertake research to look at the impact of different weightings within admission processes on widening participation;
- That Medical Schools move towards selection processes that include evaluation of academic attainment alongside use of an aptitude test and multiple mini interviews (MMIs).
- Encouragement for the further development of UKMED and the need for all stakeholders to support this.

6.3.2 Key Research Evidence Driving Change

6.3.2.1 Change in Selection

Since 2006, there has been a shift away from the use of personal statements and references in selection. At the same time, universities moved to greater reliance on

a combination of academic achievement and potential, aptitude tests and multiple mini interviews. Whilst some of the shift has been a pragmatic one, this has been informed and influenced by external drivers (described above) and an emerging research evidence base (described in greater detail in Chapter 5).

Many schools had wanted to move away from a reliance on personal statements, given the evidence that both these and school/college references had limited predictive validity (Ferguson, Sanders et al. 2000, Husbands and Dowell 2013, Tiffin, Mwandigha et al. 2016). At the same time concern increased regarding the authenticity of personal statements given the growing coaching industry around medical student selection and the variation in advisor support for applicants (Stephenson and Stephenson 2010).

Whilst the relationship is weak, papers had been published showing a significant positive relationship between the UKCAT and performance in medical school (Yates and James 2010, Adam, Bore et al. 2012, Husbands and Dowell 2013, Yates and James 2013, Husbands, Mathieson et al. 2014, Sartania, McClure et al. 2014, Adam, Bore et al. 2015, MacKenzie, Cleland et al. 2016, Tiffin, Mwandigha et al. 2016, Greatrix, Nicholson et al. 2021) including evidence that the test has incremental validity over and above academic attainment (McManus, Dewberry et al. 2013). Evidence for dental selection however has been more limited and conflicting (Cunningham, Patterson et al. 2019, Greatrix, Nicholson et al. 2021).

A systematic review of selection methods in medicine, building upon work commissioned by the Medical Schools Council, provided additional guidance for medical schools reviewing selection methods (Patterson, Knight et al. 2016). The review looked at eight different selection methods including aptitude tests, academic records, personal statements, situational judgement tests (SJTs) and interviews. The relative strengths of selection methods were discussed using four evaluation criteria: effectiveness (reliability and validity); procedural issues; acceptability, and cost-effectiveness. The authors concluded that academic records, MMIs, aptitude tests, SJTs and selection centres were more effective and generally fairer than traditional interviews, references and personal statements. However, the paper also reported ongoing challenges to the use of aptitude tests in selection highlighting mixed evidence regarding predictive validity and fairness (noting some groups performed better in such tests), and the potential impact of coaching on performance.

Similar work at a national level regarding dental selection does not exist, perhaps reflecting the fact that government focus on widening access is often on medicine (rather than dentistry).

6.3.2.2 Widening Access

There is limited and conflicting evidence regarding the effect use of the UKCAT might have on widening access to medicine. Tiffin concluded that strong use of the UKCAT might lead to more equitable distribution of offers across some under-

represented groups (Tiffin, Dowell et al. 2012) and that the UKCAT might be less sensitive to school type than A-levels (Tiffin, McLachlan et al. 2014). However, a longitudinal review of the impact of the UKCAT found no evidence of the UKCAT reducing disadvantage (Mathers, Sitch et al. 2016). Indeed, there would appear to be little evidence that changes to selection processes impact significantly on the demographic of applicants admitted to medical school (Fielding, Tiffin et al. 2018). More recently however, evidence that the UKCAT continues to predict undergraduate performance throughout medical schools (unlike school leaving qualifications) (Mwandigha, Tiffin et al. 2018) may lead schools to consider greater use of the UKCAT as a contextual measure for applicants from low performing schools.

Whilst widening access has been a key driver of change in medical selection, such strong drivers have not significantly influenced dental student selection, possibly because competition has been traditionally lower than for medicine. In addition, political attention around widening access though purportedly relating to all the professions has tended to focus on medicine. Whilst dentistry (anecdotally) can claim to have greater ethnic diversity in applications, widening access in terms of socio economic status remains a challenge (Gallagher, Calvert et al. 2017). In the event of there being a greater focus in the future on widening access in dental selection, this might lead to further changes in selection processes. Whilst the test may be less sensitive to some measures of socio-economics than school leaver qualifications (Tiffin, McLachlan et al. 2014), there is no evidence to date of it reducing disadvantage in medical selection (Mathers, Sitch et al. 2016). This is perhaps acknowledged by those medical and dental schools adjusting UKCAT requirements (usually alongside school leaver requirements) for these applicants.

6.4 Implications for the Future

Whilst medicine applicant numbers reported above ought to be treated with caution, they do however reveal some interesting trends. Whilst the number of places available at medical schools continues to rise it would appear that this is not necessarily matched (up to 2018) by increasing applicant numbers. At the same time, Universities appear to be interviewing significantly more applicants. It is perhaps the case that the increased places and the convergence of core selection criteria towards a focus on more objective measures (academic scores and UKCAT) and away from more subjective measures (personal statements) is impacting on conversion rates. That is, a reliance on these criteria, results in more applicants being invited to multiple interviews than observed previously. Whilst a move away from a significant use of personal statements in selection is to be welcomed, this was one area where diversity in selection methods existed, with schools assessing forms in different ways. Whether universities will wish to act around declining conversion rates (the proportion of applicants who accept an offer from an individual institution) remains to be seen.

Tables 8 and 10 not only summarise use of the UKCAT in selection but illustrate the potential complexity which applicants are faced with when making their UCAS choices. It remains unclear what impact on outcomes this diversity in selection processes has, but it is possible that the impact is fairly marginal (at a population level). Perhaps the sector ought to reflect on whether simpler approaches might create greater transparency for applicants. One advantage of the UKCAT is that applicants are aware of their scores prior to making their university choices. There ought to be sufficient information for applicants to avoid wasted applications. Greater exploration of features influencing applicant decision making such as that being undertaken in the UK Medical Applicant Cohort Study (UKMACS) may influence future decisions regarding selection processes.

The creation of new medical schools, expansion of student numbers and a continued focus on widening access inevitably requires schools to continue to review selection processes. Researchers seeking a better understanding of the impact of the test on applicant demographics and the predictive validity of the test would benefit from an understanding of how schools use the UKCAT and how that use has changed since UKCAT's inception. This will be of interest to researchers investigating broader issues in medical school selection.

This chapter does not seek to explore the rationales behind medical and dental school decisions as to the nature of their selection processes. The growing evidence base around selection perhaps allows medical and dental schools to reflect further on how different selection criteria align with the aims of their curricula and institutional values.

This work provides a unique overview of dental selection in the UK and perhaps provides a backdrop for further exploration of the factors influencing selection decisions for dental programmes, where the evidence base is fairly limited. There is a pressing need to explore how different selection tools predict performance in dental undergraduate and postgraduate education. The level of detail included in this chapter should help researchers planning such studies understand the complexity involved.

The results outlined above demonstrate the benefit of understanding in detail how selection processes operate using information not systematically collected elsewhere. This level of detail ought to allow for more nuanced analysis of the impact of processes on applicant demographics and outcomes. If the UK Medical Education Database (Dowell, Cleland et al. 2018) is to be used to analyse the impact of selection on outcomes in medical school and beyond, then a greater understanding of different selection processes will be necessary. There would be particular benefit from collecting information regarding admissions to widening access programmes (and use of contextual data) from schools in order to be able to properly evaluate the success of such initiatives.

6.5 Summary of Findings

In 2006, the UKCAT Consortium set out to provide medical and dental schools with a tool to assist in the challenge of discriminating between increasingly large numbers of academically high achieving applicants. At the same time universities were looking for measures (over and above academic achievement and cognitive ability) to identify the traits in applicants felt desirable to make them good doctors and dentists.

The UKCAT is now firmly established as part of the admissions landscape for medicine and dentistry in the UK, used by 30 out of 38 medical schools and all but two universities for entry to undergraduate dental training. A greater focus on the evidence base around selection has led to the UKCAT largely replacing the use of personal statements and references at most universities.

In the first years of testing, an unfamiliarity with the test led to an apparent reluctance on the part of schools to shift appreciably away from tried and tested selection methods. As a result, a large group of schools either made no use of the UKCAT or used the test in a light touch way, mainly to discriminate between applicants at a borderline for interview or offer. The use of the test by medical and dental schools has grown, diversified and strengthened since 2006. In most schools UKCAT contributes in a substantial way to selection outcomes. Even in cases where the UKCAT usage might appear relatively weak, UKCAT scores are likely to have an impact given the ceiling effect of high academic grades required by universities.

There is a tentative but growing use of the UKCAT SJT in selection, with some schools keen to include a measure of something other than academic achievement or potential. Other schools are awaiting further evidence regarding the predictive validity of this test.

Given the size of the UKCAT Consortium, the fate of most medical and dental applicants will be in at least part determined by their performance in the test.

National policy drivers and key research evidence is summarised. On a more practical level, the UKCAT has provided medical and dental schools with a simple tool for use in selection. Selection is a time consuming and costly activity and schools may have been attracted by the use of the test to narrow down applications requiring closer scrutiny. The UKCAT offers a simple, objective tool to discriminate between applicants. Whether used to reduce the number of UCAS forms to be reviewed, replacing the review of UCAS forms or used as a primary method of selection for interview, this undoubtedly streamlines and speeds up processing of applications.

Chapter 7: Admission Tutors' Views of the Impact of UKCAT on Medical and Dental Selection

7.1 Introduction

Evidence suggests a rapid and significant change in admission processes over the period this research focusses on, driven in part by political demands (Medical Schools Council 2014), a desire to use more evidence based criteria (Patterson, Knight et al. 2016) and pragmatic responses to deal with significant numbers of highly qualified applicants.

Elsewhere in this thesis, UKCAT's impact on selection had been explored quantitatively (in Chapter 6) providing information about how the use of the test has changed since 2006. The review of the literature directly relating to the test (Chapter 5) provides an insight into the contribution the test may have had in improving the evidence base in selection.

This chapter focusses on the impact UKCAT has had on selection practices and what has driven change from the perspective of Medical and Dental Admission Tutors. Admission Tutors at Consortium medical and dental schools are best place to comment on what has driven change in selection locally (Cleland, Nicholson et al. 2015).

The methods used to identify interview participants, undertake and analyse the data are described below. Themes arising from the analysis are then presented. The chapter concludes with a discussion focussed on how these themes help us answer the research questions and the practical implications for UKCAT and more broadly in selection to medicine and dentistry.

As outlined in chapter 2, a decision was made to explore this area of focus through the use of semi structured qualitative interviews. Alternatives were considered (e.g. a survey of all tutors) but given the quantitative analysis undertaken in Chapter 6 it was felt that qualitative analysis would allow for greater depth in analysis and the possibility of triangulating research findings. Given that the intention of the interviews was to answer specific research questions, semi-structured interviews underpinned by an interview guide was deemed most appropriate. At the same time whilst focussed on relevant topics, questions were relatively open, allowing interviewees a significant degree of flexibility in their responses. This approach was intended to help ensure that the research questions were addressed and at the same time to allow for new and additional information and considerations to be introduced.

7.2 Aims

The aims of a series of interviews were to understand:

- How Admission Tutors view the impact UKCAT has had on selection locally.

- What the current challenges are in selection and how tutors think UKCAT might develop to better support selection practices in the future.
- To what extent admission tutors believe UKCAT has met its original organisational objectives.

The intention of this phase of work was to explore what impact feels like on the ground, to better understand what has driven change in selection, and to consider the extent to which the test and UKCAT as an organisation can claim to have influenced this.

Looking to the future, understanding how change comes about in selection, considering the current challenges faced by admission tutors and understanding their aspirations regarding UKCAT, may help frame the future direction of the organisation.

Interviews took place by video call in summer 2020 at a point where selection was impacted by the COVID19 pandemic (Cleland, McKimm et al. 2020) and plans were being put in place for a very different round of UKCAT testing. It is inevitable therefore that responses to some extent were impacted by this experience, in particular when admission tutors were considering future challenge and change.

7.3 The Role of the Admission Tutor

Admission Tutor roles vary between universities. They tend to operate at a relatively senior level, appointed by the Head of Department/Dean or equivalent. Often working with a wider admissions team (including academic and administrative staff), the admission tutor is the lead academic responsible for developing and implementing selection processes locally. The extent to which they have autonomy in their roles may be influenced by local structures, administrative systems and the extent to which selection is managed centrally within institutions.

In a systematic review of stakeholder views of selection methods for medicine (Kelly, Patterson et al. 2018), selectors, defined here more broadly than admission tutors, viewed the interview as the most valid selection method, with the MMI being regarded as most fair interview method. Whilst noting the usefulness of aptitude tests, when used alongside the academic record to identify applicants for interview, some selectors noted concerns regarding lack of validity evidence and the potential for such tests to act as a barrier to certain applicant groups.

Surveys and interviews with those responsible for undergraduate medical selection have in the past included single centre studies and/or looked at specific aspects of admissions such as MMIs e.g. (Marrin, McIntosh et al. 2004, Kumar, Roberts et al. 2009, Razack, Faremo et al. 2009, Turner and Nicholson 2011, Dowell, Lynch et al. 2012) or broader surveys to explore changes over time (Monroe, Quinn et al. 2013).

A study at McMaster University (Marrin, McIntosh et al. 2004) sought to investigate the importance of key qualities in the selection process for medicine amongst

stakeholders. Academic staff identified validity, fairness and accessibility as being the most important qualities.

The admission tutor view in relation to the impact UKCAT has had on selection is an important one. Tutors play a pivotal role in delivering and developing selection processes. They will often also be the university's UKCAT representative and as such have a greater understanding of the test and the organisation than other colleagues involved in selection.

7.4 Methods

Ethics approval for this study was obtained from the University of Nottingham (Reference 458-2001 attached as Appendix N).

7.4.1 Selection and Withdrawal of Participants

Each UKCAT Consortium university identifies two Consortium Representatives who act as a first line of communication between UKCAT and the university. These Representatives tend to be either an Admission Tutor (academic role) or Admission officer (senior administrative role). Interviewees were drawn from representatives in academic roles as they would normally, whilst working with colleagues, be aware of and responsible for setting and developing local selection policies.

Sampling was used to ensure a balance of interviewees bearing in mind the following factors:

- Medical/dental school
- Those new to the Consortium
- Admission tutors who could comment on changing selection processes (i.e. been in role for some years)

It was anticipated that 10-15 interviews would be required; ultimately 14 interviews were undertaken.

At the time of organising the interviews, there were 30 UKCAT Consortium universities and consequently, 60 representatives. At universities with both a medical and a dental school, one representative normally represents the medical school and one the dental school. At universities without a dental school, both representatives would be medical. The following steps were taken to achieve an appropriate sample to interview:

- Identify the maximum number of representatives (n=60);
- Omit administrative staff where there is an academic contact at the same medical school (n=15);
- Omit academic staff where there are two academic contacts at the same medical school (n=3), the representative longest in role was selected.

Step 1: An invite for expressions of interest to take part in these interviews (Appendix I) was sent out to the remaining 42 representatives; the invite outlined the purpose of the study, how the interviews would be conducted and next steps.

Step 2: Expressions of interest were received from 19 representatives. These were reviewed to achieve a balance of interviews, reflecting actual distribution, across medical and dental representatives and to include a majority of representatives who had been in place for some time and therefore able to reflect on change and impact. UKCAT Board members were excluded from the sample to avoid issues of bias. This sample was independently reviewed before moving onto the next step.

Step 3: Those identified to take part were provided with a Participation Information Sheet (Appendix J) and consent form (Appendix K). They were asked to confirm a convenient time for the interviews to take place. Consent forms were obtained from participants in advance of interviews going ahead.

Step 6: A week in advance of each interview, arrangements were confirmed with participants, and they were provided with a brief overview of areas to be explored.

7.4.2 Interviews

At the start of each interview, the purpose of the study was outlined, and receipt of the consent form confirmed. Participants were reminded that the conversation was being recorded and offered the opportunity to raise any issues.

An interview guide (Appendix L) was used to create a consistent structure for interviews. The guide allowed for additional follow up to questions and for interviewees to explore alternative areas. The content of the interview guide was informed by work to date including that outlined in Chapter 6 looking at how use of the test has changed since 2006, in Chapter 4 where the aims of the organisation are explored alongside a review of archived materials describing the development of the organisation and in Chapter 5 the review of the literature directly relating to UKCAT published since 2006.

An early draft of the interview guide was trialled with an Admissions Officer which led to the refinement of several questions. It was anticipated that modifications to the interview guide might be required as the interviews progressed. In the end the only modification was the addition of a question (following the second interview) regarding feedback they may receive from applicants regarding UKCAT.

The interview guide intentionally focussed on the research questions being addressed in this chapter. Whilst this had the potential to constrain the discussion, questions were reasonably open ended, allowing interviewees a significant degree of flexibility in responding with options to widen the discussion.

7.4.3 Analysis of Interviews

Interviews were audio-recorded and transcribed using on-line software (Otter). Transcriptions were further checked against the original recordings to ensure accuracy. Once anonymised, transcribed interviews were entered into NVivo

software. A discussion with a research collaborator took place regarding the extent to which data could be analysed quantitatively. Given the number of interviews that had taken place and the focus elsewhere in the thesis on quantitative analysis of how the test was used, a qualitative approach remained the preferred option.

A thematic framework was created using the software, in discussion with collaborators, to allow the coding of each interview using template analysis (King, Cassell et al. 2004).

As discussed in more detail in chapter 2, thematic analysis provides researchers with a method to describe and analyse patterns in, usually, qualitative data by identifying themes within data collected to answer the research question. Braun and Clarke proposed approaching analysis through a series of six steps (Braun and Clarke 2006); their method becoming widely adopted.

These six steps as applied to the analysis of interview data are described below.

Step 1: Familiarisation with the data

Interviews were audio recorded (by me) and transcribed using transcription software. Initial transcriptions were checked in detail against recordings, ensuring my familiarity with interview content at the point of data analysis.

Step 2: Generating initial codes

Now familiar with the data available, I started to develop a coding framework linked to the interview guide. After discussion with collaborators it was agreed that this approach might be too restrictive. In light of that, I revisited the coding. A new initial coding template was agreed with the input of the research collaborator. After being applied to the first interviews, the initial coding template was then applied across all the data and revisited regularly as each interview response was reviewed. As coding progressed additional codes (autonomy and confidence around research) were added drawing out nuance which had not been initially identified in the original analysis. A revised coding template was shared with the research collaborator at a mid-point of the analysis.

The final coding framework is attached as Appendix N.

Steps 3 - 5: Search for and refining themes

The interview questionnaire and interview prompts were designed around the research questions. Given the semi-structured nature of the interviews, discussion with interviewees intentionally focussed on these areas. Interviewees were encouraged to talk about change in selection over time, drivers for that change, impact of UKCAT on selection, current challenges and future development of the test and the organisation. They were also asked to comment specifically on the extent to which UKCAT had met its original objectives around increasing fairness in selection, widening participating and improving the research evidence base.

Originally, 6 themes had been identified (earlier change, UKCAT impact, fairness, widening participation, evidence base, future change). These themes were revisited after discussion. 'Motivation for change' was felt to better describe the many drivers around change which tutors referred to during the interviews. Widening Participation was referred to as both a driver for change and in the context of discussions around fairness. It was felt appropriate to integrate widening participation into these two themes. Having relooked at the data it was felt that 'defending selection decisions' was discussed both in terms of change, as an aspect of fairness and with regarding to the research evidence base. his was therefore identified as a standalone theme. The following themes were agreed:

- Motivation for Change
- UKCAT's Impact on Selection
- Fairness in Selection
- Defending Selection Decisions

Thematic analysis helped identify where these themes arose more broadly across the interviews, allowing exploration of them in different contexts, in greater depth and to triangulate responses.

After identifying the themes, coding was revisited to provide reassurance that sufficient data was available for each theme and that overlaps where they existed made sense.

Step 6: Producing the manuscript

Results are presented having worked through codes associated with each theme.

7.5 Reflexivity

The importance of reflexivity in this thesis in general and specific to qualitative interviews is discussed in Chapter 2. Throughout this study there has been a need to consider the active involvement of the researcher and the potential impact on outcomes. My role in the organisation is outlined in Chapter 1. I knew most of the interviewees relatively well, having worked with them in the organisation, in some cases over a period of years. The interviewees and I almost certainly have a level of loyalty to the UCAT and their own Universities which needed to be recognised when analysing the data and reflecting on key messages.

The approach to inviting admission tutors to take part in interviews was discussed with supervisors. The potential of tutors volunteering to take part being dominated by those more involved in the Consortium and interested in its development was noted. Tutors were given more than one opportunity to volunteer to take part. This risk was mitigated to an extent by the fact that sufficient numbers came

forward to ensure that Board members could be excluded from the sample. The sample was independently reviewed before being finalised.

The interview guide was trialled before implementation. During the design of the interview guide I reflected on my own presumptions regarding outcomes. It was anticipated that interviewees might make assumptions about my level of knowledge regarding local selection processes which could create issues when coding the data. A number of opportunities to prompt interviewees and/or seek clarity were built into the interview guide. It was expected that some interviewees, with less experience, would not have sufficient knowledge of historical selection processes to comment on change. Whilst recruiting interviewees, care was given to achieve a balance across experienced admission tutors and those that were relatively new. The interview guide included a question regarding the 'biggest challenge in selection'. Interviews took place during the COVID pandemic which was impacting significantly on selection processes. Interviewees were asked to put this to one side when reflecting on 'challenge'.

Interviewees were asked to reflect on whether UKCAT has achieved its original aims. This was potentially the area of the interview where the relationship between the interviewer and interviewees and loyalty to the organisation had the potential to impact most. Questions here were as open ended as they could be, and prompts avoided so as not to influence responses. Following the identification of the themes and drafting of the discussion, the original transcripts were revisited to check that content in this chapter was a fair reflection of the interviews.

Reviewing taped interviews allowed me to reflect on my own performance as an interviewer.

Involvement of an independent reviewer at critical stages of the process allowed challenge to decisions made regarding sampling and data analysis. As described above, the coding framework was developed with the input of collaborators and revised in the light of their input. Themes emerging from the coding framework were again discussed and agreed with collaborators and further reflected on as the chapter progressed.

When writing the discussion, I reflected at regular intervals as to whether my position in UKCAT and prior knowledge of the organisation influenced conclusions being drawn. This was achieved both by input from collaborators and regularly reflecting on whether an alternative view might be taken by an individual independent of the organisation.

7.6 Results

In this section the interview participants are described, and the themes identified above explored.

7.6.1 Participants

Admission tutors from 19 UKCAT Consortium medical and dental schools volunteered to take part in the study. Of these, 14 were interviewed which included 10 medical and 4 dental admission tutors. Tutors came from a mixture of new and old schools including some medical schools very new to the Consortium. Those interviewed included individuals who had been in the role for only a few years, whilst others had been active in medical/dental selection for over 10 years.

The mean length of interview was 31 minutes and 11 seconds (range 11-45 minutes). In total there was approximately 7 hours of interview time.

7.6.2 Themes

7.6.2.1 Motivation for Change

During the interviews, tutors reflected on how and why selection had changed since the creation of UKCAT. Examples of motivation for change tended to reference significant process change such as a shift away from the scoring of personal statements or a movement to multiple mini interviews. For some, change had been driven by a desire for greater fairness and transparency, with some current and previous elements of selection being deemed more or less fair than others.

> *…we want a selection process, so it's within our own ethos, but it's also within the ethos of the institution to have a selection process that is open and transparent. (Tutor 1)*

Local factors such as applicant conversion rates, time and resources and local analysis of data had motivated change. Widening participation, regarded both as a local and national driver, influenced selection practices significantly, with the newer medical schools in particular commenting on this being central to decision making.

> *…it is our, our requirements from the government in terms of local WP, making sure we get close to what we're supposed to be doing, or achieving better than that… (Tutor 13)*

The research evidence base around selection had driven change at a number of universities, with the Selecting for Excellence Report (Medical Schools Council 2013) in particular being regarded as providing guidance on good practice. At the same time, admission tutors referred to the research evidence base specific to UKCAT directly driving change.

> *And as the evidence around UKCAT increased and improved, we felt able to discard personal statements, so we don't read them at all now. And the reason we wanted to move away from that was partly because we were finding that most personal statements are very similar to each other, so it's not a very fine selection tool. The evidence is against them in the literature as a selection tool and the guidance from Medical Schools Council is against using them. And...as the evidence around UKCAT improved and it was shown in*

large studies, that it is predictive of how students will perform at medical school and how students will perform in their UKFPO application. So it has predictive validity and outcome validity. We felt it was a suitable test to be able to, to substitute for reading personal statements. (Tutor 1)

7.6.2.2 UKCAT's Impact on Selection

UKCAT's impact on selection was explored during the interviews in a number of ways. Tutors described the use of the test in selection and were asked explicitly to quantify impact on selection decisions. Tutors were also asked to comment on feedback from applicants and any observed impact of the use of the test on applicant demographics.

In terms of the impact the test itself had on selection, tutors commented on its use as an additional selection tool:

I think just it gives us another tool for selection. It may not be a big impact, but we just kind of feel it does give us another tool. And it gives an opportunity for maybe those candidates who have maybe not been at the best school for their GCSEs to prove themselves in another way…. (Tutor 6)

…we're looking at things such as the school you know, so we've got contextualised things coming in and so the performance of the school, GCSE scores, and like. ….. it's an important part of the formula……. (Tutor 11)

For some schools, the test helped screen large numbers of applicants.

…we get more than 2000 applications per year. And based on the grades, many of them would meet the cut off. However, we just need somehow to differentiate. So operationally that is useful. And that's how I would say the impact mostly to just helping to manage the numbers. (Tutor 10)

And also it helps us manage our numbers…it comes so early, it's almost the first thing we look at, is that it's actually a screening tool. It's an initial stage of the whole applications process. (Tutor 3)

Associated with this was a perception that the test was a more efficient and less time-consuming selection tool, particularly when compared with scoring of personal statements.

The biggest challenge in selection I think it's, I think it's a simple one of time, really. It is a hugely time consuming process. (Tutor 1)

...the amount of time it saves as compared to using personal statements is enormous. So there's a huge benefit to us there. (Tutor 1)

...it seems just a more manageable thing to do to drop the personal statements and to do a pure academic and UKCAT scoring system. (Tutor 6)

However, this did not necessarily impact positively on the quality of selection decisions.

…one of the big drivers has been to try to keep things efficient. So, I mean, it doesn't kind of produce better results necessarily, but just with the numbers applying and the numbers that we're having to interview, we have had to look at keeping things as efficient as possible and looking at ways that we can select applicants that is not going to overburden our systems. (Tutor 6)

Tutors were most likely to comment on UKCAT having a significant impact on selection outcomes. Impact was perhaps more quantifiable for schools applying a score threshold.

…UKCAT is the second hurdle and then everyone is invited to interview and at that stage, everyone on the level playing field. Got quite an impact here. (Tutor 10)

However, even where schools used the test in a less direct way, some tutors commented that the impact on individual applicants remained significant:

…it's maybe more at borderline decisions later on… (Tutor 12)

I guess for the applicants that we review, the sort of more borderline applicants if you like, then a good UKCAT score definitely lifted them. So they might, they can afford to lose more points on their academic score. So I think it makes an immense significant difference. (Tutor 4)

But looking at the SJT portion if you had a very, very low UKCAT, particularly if they were in band four…they will be very unlikely to be called to interview. (Tutor 8)

In a couple of schools, tutors commented that with increasing numbers of interviews taking place, the UKCAT score threshold was now relatively low. Tutors regarded the UKCAT as therefore having a lower impact on selection.

Probably not as big as some other places. ...because we interview so many, they can get an interview with us with a relatively low UKCAT score. (Tutor 6)

Given that widening participation was one of the initial aims of the organisation, tutors were asked to comment on the impact UKCAT had had in this area.

Those that considered UKCAT as having a positive impact on widening access talked about UKCAT's support for test candidates from low-income backgrounds.

> *…you very openly support widening participation and you make, you make the test available to people who can't afford it. You put your resources out there to support. So in that respect you do… (Tutor 12)*

Some tutors pointed out that their local data did not highlight any disadvantage for this group of applicants.

> *…our view is that it probably doesn't particularly disadvantage candidates from any particular background. (Tutor 1)*

> *…it looks like UKCAT is actually acting as a good factor in selection. It isn't discriminating against any particular group of people. (Tutor 14)*

However, it was more common for tutors to comment on the poorer test performance of this group of applicants and to reference impact where local adjustments that had been made to selection processes.

> *it does worry me to be honest why people who can do very, very well from a WP scheme but with lower UKCAT scores that they...if we just use UKCAT the way, you know, without considering it, then that would be something that wouldn't be right for us, because we would not consider them basically. So I don't know why, I don't know why that those groups don't do so well considering all the information that is there for them. (Tutor 12)*

> *I still worry that the students, although there's all these free tests and everything out there, I still worry that the students from the WP background are not performing as well on UKCAT. (Tutor 6)*

> *And we exempt applicants who have applied through our widening participation scheme…. (Tutor 1)*

The perception existed for some, that UKCAT was an actual barrier for this group of applicants. Where tutors reported UKCAT as having a negative impact on selection this was largely in relation to the impact on widening access and other applicant subgroups.

> *There has been a concern expressed in some quarters at the university, mainly by those who run our widening participation*

programme, I suppose, because they see UKCAT as being the major barrier. (Tutor 1)

I think it's students from some of the widening participation really have had a tough time and I think putting that extra pressure on them, to do an extra exam to prepare, to study for, is a lot on top of their A levels. So I think it may have had a negative effect potentially on them. Because of the extra workload needed and the extra pressure. (Tutor 5)

…there's also a kind of difficulty with the international students doing situational judgement test. They also perform a lot worse in the situational judgement test. So I don't know whether that's an English language thing. (Tutor 6)

Other negative comments related to the potential impact on applicant conversion rates in the event of universities selecting the highest scoring applicants.

Whilst tutors on the whole were familiar with their own selection data, they did not feel that their use of UKCAT impacted on the demographics of their selected applicants.

Generally tutors had little feedback from applicants regarding the test, commenting that most applicants now regarded the test as an established part of medical/dental selection:

What we see is people just accept it as part of the landscape. And they say, well, if the medical schools use the test, they must know what they're doing. Therefore, it must be something that impacts upon me being a doctor later. (Tutor 2)

Applicants tended to be most concerned about how universities used the test rather than the test itself.

…the UKCAT is probably the single most often asked question about the whole admissions process. It's not about grades, often it's not even subjects. It's about how do you use the UKCAT because everyone seems to use it differently. (Tutor 3)

However, some tutors commented on applicants failing to see the relevance of some UKCAT questions:

I think it tends to be the Abstract Reasoning they don't see the benefit of it, or why it's there. I'm not sure I do either… (Tutor 12)

…I think applicants feel they do like SJT part of it but they definitely don't feel the cognitive part, especially the abstract reasoning part really they hate. Some people have issues also with the decision making. So it's all about trying to trick you, they said… (Tutor 10)

7.6.2.3 Fairness in Selection

Most tutors talked about wanting to achieve fairness and transparency within selection.

Transparency also related to information applicants received about selection processes more broadly and specifically in relation to UKCAT.

> *…we do publish the previous year's cut off scores; for as long as we've been doing this, they're all on our admissions web pages, so that applicants can see that covered with caveats that of course, you are competing with this year's cohort, not last year's, and the cut points can be different. And indeed, from the tabulated scores that we've put on our on our web pages, you can see that they're different every year. So we're driven by a desire to be transparent and a desire to be equitable… (Tutor 1)*

Many tutors made explicit links between fairness and widening participation.

> *It's being fair, isn't it? You know, and being mindful of it, again, it's the whole widening participation side of things, you know, the different context in which people are, you know, coming here different backgrounds and so on… (Tutor 11)*

Most tutors agreed that the UKCAT had increased fairness in selection although this was discussed in different ways.

The very existence of an additional selection tool for use in selection was considered to increase fairness.

> *…the more the more aspects you've got to look at when you're looking at particular group of individuals or the whole cohort, then in a sense, the fairer you can make it. (Tutor 4)*

> *…..because the more you spread your net, then different people are going to shine at different things and it will become more fair. So in terms of adding another string to our bow, yes. (Tutor 9)*

Some tutors commented on UKCAT's support for candidates contributing to fairness.

> *It's produced a transparent way of selecting. And I think that candidates perceive that then as being fair, because they have a kind of equal opportunity to practice for it. And to then, you know, take the test. And I think the work that UKCAT's done in kind of getting all the free tests out there has helped with that fairness. (Tutor 6)*

> *And I think the work that you have done in relation to widening participation means that, you know, the bursary schemes, the*

UKCATSEN 50%, 25%, all these extra things you do, I think, yes, it is helping fairness. I do believe that. (Tutor 12)

Those tutors less convinced about the contribution UKCAT had made to increasing fairness in selection commented on the impact of coaching on performance in the test.

I think again, it's, it's about the opportunities that are available for those who have got support be it financial, just practical support to do to get as many practice papers to do the courses or whatever it is. (Tutor 4)

7.6.2.4 Defending Selection Decisions

Linked to fairness tended to be a desire on the part of tutors to defend the decisions they were making (at an applicant level) and across the wider selection process.

The challenge for admission tutors to select the right candidates appeared to be little changed since UKCAT was established.

Our biggest selection recently has been that we've had a huge number of very, very high-quality candidates. And differentiating between them has been very difficult actually. Certainly, since moving to MMIs we're interviewing more people. And we've just found the, the academic standards have been extremely high and standard at interview also extremely high. (Tutor 8)

I think more and more what we're trying to get is a predictor of how people will behave, sort of their attitude and how they'll behave professionally and how, you know, throughout the entire five years. So we seem to get quite a nice bunch of students at the moment. Biased but you know…. But we still have people who, you know, we still have, we all have problem students don't we. (Tutor 9)

…the biggest challenge nationally and probably internationally is whether we are selecting the right group of doctors and many things feed into that whether in terms of personal qualities, these are the right people, that in terms of academic achievement, and learning for medicine, these are the right people and whether in terms of social justice, these are the right people, or whether they are the right people, but there could be better people. And I think that's always the perennial challenge of selection. (Tutor 2)

Some tutors commented on the role of the UCAT in improving their ability to defend selection decisions.

…it's really being challenged on decisions that we make. …I want to make sure that I can defend any of the decisions that we're submitting to paper, so that when we do get challenged, I've got something robust and clear that I can feed back to the students. (Tutor 7)

…it allows us to demonstrate that we are fair, which is important. And it allows us to defend our decisions, which is increasingly important…(Tutor 1)

Other tutors commented that UKCAT was more objective than some other selection tools, commenting about the difficulty in defending use of other selection criteria (personal statements in particular).

So it's a concern for me to grade or mark or attach something like that to a statement which could have been carefully crafted by an applicant or their friends, family or colleagues at the college or school. And I'm not trying to be negative about it. I'm just saying these are doubts that I have. (Tutor 3)

….it's very much an objective type of test as opposed to a lot of the other things people are going to be using such as interviews or scoring personal statements or what other methods they are using. So, so I think that's got to result in a fairer process. (Tutor 11)

…for the reasons I outlined earlier regarding transparent…. lack of transparency and other metrics that can't be validated. So I think undoubtedly, undoubtedly, UKCAT contributed significantly to our knowledge and fairness and validation. (Tutor 3)

Defending the use of selection criteria had clear links to the research evidence base around selection and the test in particular

Most tutors understood the organisation to have contributed to research around selection and in relation to the test in particular.

I would say the research group was active and creating all those study groups together to produce the evidence base. I think that was useful. (Tutor 10)

And I know there's loads of stuff out there as well. There's lots of information out there from, from you guys and from the universities, and there's lots of published information… (Tutor 13)

Some comments were made about how the research base itself had changed and developed over time.

I think the initial evidence from sort of 5/10 years ago was a bit shaky really, because there were lots of single centre studies meaning 'Well, yeah'; with the advent of multi centre studies, and hopefully perhaps in the future multinational studies, then the evidence base is increasing all the time. (Tutor 1)

Tutors also commented on the role of UKCAT as an organisation, as a forum for discussion and dissemination of research.

it's a very academic environment, you know, it's not guesswork or let's, you know, why don't you do this, why don't.....it's a very academic, evidence based environment. (Tutor 14)

But I like that aspect of the meetings when things are presented to us as well. (Tutor 6)

Some comments were made regarding the lack of evidence to support claims in relation to the test's ability to widen participation.

I would like to see more on these types of applicants, how they perform within individual medical schools, what their barriers are. I think that's what I would like to see… (Tutor 12)

Whilst tutors felt it was desirable for the Situational Judgement Test to become a more significant element of selection and that more evidence was perhaps needed to justify its increased use.

I'd like the SJT to be more robust so that we could put it front and centre. (Tutor 1)

I think the SJT is probably a valuable addition, although I am not sure. I think, yeah, well, obviously there's a research base that needs to be established there. (Tutor 2)

Research and data analysis undertaken locally did not always support the claims made in other studies.

I will say that the small amount of sort of analysis of our cohort performance in years one and two doesn't indicate that there's a huge effect of UKCAT. (Tutor 6)

The lack of evidence around dental selection was commented on as a particular shortcoming for dental selectors defending the decisions they were making.

I'm not aware of any recent dental research. So I know it's out there for medicine. But again, seems to be seems to be correlating with the same sorts of you know, the academic performance and dentistry is very different from medicine. (Tutor 9)

Whilst tutors understood the importance of an evidence base to underpin their selection processes, the confidence with which tutors talked about the literature varied. Some tutors appeared to be familiar with the literature and gave examples of applying this knowledge to selection processes.

> *So we were driven by the evidence and where possible, we would like to be evidence based in our selection process being mindful that at the moment, although it is improving, the evidence base around admissions is still relatively light. (Tutor 1)*

Some tutors however, were less confident about the evidence base although aware of its existence.

> *I have not really engaged with it. (Tutor 7)*

> *…keeping on top of literature etc, is actually quite challenging. So I do like it when the meetings help us to look at those kinds of things. (Tutor 6)*

Looking to the future, some tutors were interested in exploring their own data to inform future development of selection processes. On occasion, time and resource appeared to be a limiting factor here.

> *I believe it's really important to look at our own information and look at the cohort where we're sort of accessing, look at how they perform, and so on… (Tutor 14)*

> *We don't have an awful lot of free time to do some, some of the things that we'd like to like sort of going over and analysing some of our data and our performance, etc (Tutor 6)*

7.7 Discussion

7.7.1 How do Admission Tutors view the impact UKCAT has had on selection?

7.7.1.1 Change in Selection

The extent to which tutors commented on change in selection depended to some extent on their time in the role. Some were able to reflect on several admission cycles whereas for other tutors their involvement in selection was relatively recent and on occasion they were unaware of prior change that had occurred or the reasons for it. For some tutors, their ability to comment in detail on some aspects of selection processes was limited by the fact that elements of the process were undertaken elsewhere within institutions. I was surprised at the limited insight a few tutors had into their own selection processes due to the centralisation of some functions. If some tutors, due to the centralisation of admission processes, are less autonomous in their roles than others it is perhaps more difficult for them to influence selection processes to the same extent. Whilst conflict is not inevitable in such situations, there is a potential tension here if what motivates change and the values underpinning decision making vary between the central department and the

medical/dental school. For organisations such as UKCAT this creates a potential disconnect as decision making and communication structures rely on relationships with medical and dental schools. At times of significant change, exploring wider relationships where centralised decision making exists may be of value.

There was consensus that the admission landscape had changed significantly since 2006, most notably in a move away from a reliance on personal statements and towards a wide use of multiple mini interviews. These changes in particular were considered to have been motivated by both the research evidence base with a particular focus by interviewees on the Selecting for Excellence Report (Medical Schools Council 2014) and national policy drivers. Change was also driven by a range of local factors, alongside a desire to widen participation in medical and dental education. Whilst this is reassuring, as discussed below, not all admission tutors were confident about the research evidence base. National organisations involved in medical and dental selection, including UKCAT, ought to support admission tutors in accessing and understanding the research evidence around the test and potentially more broadly in selection.

7.7.1.2 UKCAT Impact

Tutors accepted UKCAT as being embedded in the selection landscape and clearly impacting on selection. How tutors quantified UKCAT's impact on local selection processes varied. For some schools, applying a UKCAT threshold was described as having a significant impact. However, one school noted that given the large numbers of applicants they were interviewing, the impact on selection decisions was relatively low. A couple of schools noted that whilst they used the test in a relatively light touch way, the impact of its use on a small number of applicants (e.g. those at a borderline) was significant. This tension between the global impact of test scores on selection outcomes and the outcome for individual applicants is one I have reflected on before and it was interesting to hear this come through in the interviews. Applying a score threshold, however low, impacts on applicant choices of medical and dental schools. Schools using the factor method in selection might be *perceived* as having a lower UKCAT impact than those using thresholds, but this may not always be the case.

This highlights the fact that quantifying the impact of the test on selection outcomes locally or globally is difficult and can be misleading. I had not focussed previously on the potential disconnect between perceived and actual impact. If this tension is present in admission tutor descriptions of processes, then it is not surprising that applicants perhaps over emphasise the effect the test has on selection outcomes. Most universities provide information to applicants about their use of the test in selection, and many provide information regarding not only cut-off scores but the score ranges for interviewed and successful candidates. All Universities should provide this information and take significant care when describing the use of the test in selection. Impact was also highlighted more globally in terms of selection processes, particularly around the ability of the test to

help efficiently screen large numbers of applicants. It is possible that efficiency and defensibility have been as strong in driving change as the available evidence base.

7.7.1.3 Fairness and Widening Access

Fairness in selection is not a straightforward concept. Whilst I had reflected on this previously this came through really strongly in discussions. Barriers exist to many applicant groups particularly those from lower socio-economic backgrounds (Medical Schools Council 2014) reflecting social and educational disparities impacting on educational outcomes and individual aspirations. Selectors are clearly interested in these factors, but many are largely beyond their control. The discussion here focusses therefore on fairness in a system already flawed in this context and on the relative fairness of selection tools, none of which are perfect.

Fairness is discussed by tutors in a variety of contexts. UKCAT provides tutors with an additional selection tool and this in itself was regarded as making selection fairer. The provision of free preparation materials was deemed to help level the playing field, although countered to an extent by the now firmly established coaching industry. The provision of bursaries removed one barrier for candidates from low income backgrounds in taking the test. UKCAT is deemed to be more reliable and objective than the previous use of personal statements, perhaps allowing medical and dental schools to be seen to be fairer and, in particular, objectively defend selection decisions. Tutors also talked about the research evidence base providing some reassurance about how fair the use of UKCAT is in selection.

At the same time tutors were concerned about the impact of the test on some subgroups, with applicants from a widening access background thought to perform less well in the test. A number of schools had put in place adjustments to their processes, such as accepting lower scores, from these applicants. I was interested to see the extent to which the use of such adjustments had grown. Associated with this group of applicants was also a perception that performance in the test was influenced by both support for the applicant and specific test preparation coaching, whether from school/college or commercial companies. Evidence would suggest that the actual impact of coaching is relatively low (Lambe, Waters et al. 2012, Griffin, Carless et al. 2013, Griffin, Horton et al. 2021). However, again, the perception of such impacts, acting as they are understood to do against candidates from lower socio-economic backgrounds, is very damaging to admission test providers. There is a need to present the literature around coaching in a meaningful and accessible way to stakeholders.

Differential performance in candidate subgroups is widely reported (Tiffin, McLachlan et al. 2014) and described in Chapter 4. UKCAT publishes information regarding candidate performance on an annual basis e.g.(Pearson VUE 2020). Performance in the test is regarded as being less impacted by candidate demographics than school leaver qualifications. However, why these performance differences exist is not well understood although the reasons behind this are likely

to be complex and additional research is required (Patterson, Knight et al. 2016). Qualitative research would be particularly helpful in exploring differences in how candidates prepare for the test and inform future candidate advice and possible adjustments to test content. In addition, a better understanding of subgroup performance data across the test and within subtests could lead to identifying changes to test content which might reduce observed differences.

7.7.2 To what extent do Admission Tutors believe UKCAT has met its original aims?

7.7.2.1 Improving Fairness in Selection

Despite some of the concerns outlined above, tutors were generally of the view that UKCAT had achieved its aim of improving fairness in selection. This possibility reflected a belief that the UKCAT, whilst far from perfect, was fairer than the widespread scoring, or equivalent, of personal statements in the past. As mentioned above, fairness is discussed by tutors in a number of contexts during these interviews. Tutors were of the view that UKCAT was a fairer selection tool than others used previously and allowed medical and dental schools to be seen to be fairer when discriminating between large numbers of highly qualified applicants. However, tutors remained concerned about differential performance in the test. Whilst there may be a consensus that UKCAT has improved fairness in selection there remain significant concerns about fairness in selection for some candidate subgroups.

7.7.2.2 Widening Participation

I was not surprised to note that tutors were less convinced that UKCAT had achieved its stated aim of widening access to medical and dental training as these comments have been made previously in discussion within the Consortium. Support for candidates, both financial support for the test fee and preparation materials, was acknowledged as positive. However, whilst some tutors did not feel that UKCAT acted as a barrier to these applicants, more tutors were able to cite negative impacts and adjustments put in place locally to compensate for these.

7.7.2.3 Improving the Evidence Base

Tutors were positive about UKCAT's contribution to improving the evidence base around selection, also commenting on its organisational role as a forum for disseminating research and analysis. Whilst some tutors spoke very confidently about the evidence base, others, though aware of existing research, were less confident about applying this to their practice. The evidence base around selection has increased significantly over this period. However, if practitioners are sometimes struggling to keep on top of and interpret this evidence (Patterson, Knight et al. 2016), there is a need for organisations such as UKCAT and the Medical Schools Council to facilitate dissemination and help interpret findings.

Kelly et al (Kelly, Patterson et al. 2018) comment on the preponderance of quantitative studies in their systematic review of stakeholder views and go onto comment 'Qualitative research, on the other hand, is ideally suited to understanding the meaning of selection for the respective stakeholder groups and

can greatly add to our understanding of the views and attitudes of stakeholders'. It is perhaps the case that given the amount now known about the validity and reliability of most selection tools, a focus on understanding how these tools are experienced by stakeholders through a greater qualitative focus is now needed.

7.7.3 The Future

On a very practical level, and possibility reflecting the timing of interviews (during the COVID 19 pandemic), some tutors welcomed the ability for the Consortium to meet online given restrictions on their time. The potential of online testing, offered as an option to candidates for the first time in 2020, impacting on future test delivery was commented on.

I was surprised that tutors had little to say about potential change to the test itself. The test needs to remain fit for purpose and without selector and educator engagement in discussions about and indeed driving test development this could drift. Given the impact the test has on selection it is desirable for admission tutors to be actively engaged in discussions regarding test content and future options for development. The traits measured within the test remain little changed since 2006, except for the introduction of the Situational Judgement Test and the replacement of Decision Analysis with the Decision Making subtest. There are alternative traits which could be measured and different approaches to testing these traits. The Consortium ought to be directly involved in an ongoing review of test content to ensure the test remains fit for purpose and that its content is defensible.

The area of the test that admission tutors did comment on more consistently was the Situational Judgement Test. There was a general desire for a greater evidence base around its use to give selectors greater confidence to use this test of noncognitive traits more strongly in initial selection. I find it interesting that admission tutors feel less confident about using this subtest without a more significant evidence base. Making decisions regarding applicants' cognitive ability feels very different from measuring their understanding of, for example, empathy. There is now a significant body of evidence regarding the utility of SJTs in selection (Patterson, Zibarras et al. 2016). However evidence regarding the predictive validity of the UKCAT subtest is less extensive than that for the cognitive subtests (Greatrix, Nicholson et al. 2021). There are challenges around such studies largely because of the lack of appropriate educational outcomes in medical and dental education to measure the test against. However, given the desire of admission tutors to use this test more actively in selection then it is important that research in this area is prioritised.

7.8 Conclusions

The themes identified in these interviews help us understand the values underpinning selection to medicine and dentistry in the UK and how these impact in a very practical fashion on selection decisions. At the same time, findings help us understand the admission tutor role in a different light, reflecting on issues of autonomy for example. The impact of UKCAT on selection is discussed from

different perspectives, highlighting a tension between perceived and quantifiable impacts, with implications around how UKCAT and universities communicate these messages to potential applicants. For many admission tutors, the concepts of fairness and widening participating were linked, although their views on the extent to which UKCAT had met its original objectives in relation to these areas tended to be different. There were significant differences in the confidence with which some tutors talked about UKCAT's research evidence base. Admission tutors were keen to have greater confidence in increasing the use of the UKCAT SJT in selection processes.

Research into UK medical selection in particular has grown significantly in recent years. Given this fact and the importance of the admission tutor in developing and enacting local selection policies it is perhaps surprising that little is known about their views on selection (Kelly, Patterson et al. 2018). This study provides an insight of admission tutor views into one specific aspect of selection to medicine and dentistry. Admission Tutors believe that UKCAT now has a significant impact on selection outcomes. They believe that whilst this has increased fairness in selection, negative impacts on specific subgroups are less desirable.

Chapter 8: Evaluation and Conclusion

8.1 Introduction

This thesis investigates the impact UKCAT has had on selection to medicine and dentistry since its creation in 2006. At the same time the thesis aims to explore the extent to which UKCAT can be said to have met its original aims. This chapter attempts to answer the original research questions drawing on evidence provided in the thesis. The impact of the UKCAT on selection is summarised highlighting at the same time some of the unintended consequences of the development of the test. This chapter also focusses on the future of the organisation and, where findings include recommendations for development and change, identify these in order to inform and facilitate future discussions within the organisation.

8.2 What were the drivers behind the establishment of the UKCAT Consortium?

Chapter 4 presented the original ambitions of the organisation which can be summarised as seeking to improve fairness in selection; providing a selection tool that could discriminate between high achieving candidates; testing traits not necessarily being assessed in school leaver qualifications such as critical thinking, problem solving and those characteristics desirable in good doctors and dentists; and helping widen participation. The extent to which UKCAT has met these aims is discussed later in this chapter.

In the record of these early discussions, fairness is not defined and can be interpreted in different ways. In terms of the fairness of actual selection criteria, fairness might be best understood as a desire to move away from the use of less objective selection measures such as personal statements and references. There was no evidence of these selection criteria having predictive validity and there was uncertainty in particular about the value of personal statements given the significant influence of schools and parents on their contents.

Of course, the UKCAT was initially an unproven tool and early criticism of the use of the test in selection was that there was no evidence it predicted performance in medical and dental programmes. However, those involved in the creation of the test will have been aware (as outlined in Chapter 3) of an established evidence base showing that other selection tests (such as MCAT and DAT) predicted performance in medical and dental school over and above academic qualifications.

Fairness might also relate to candidate performance, and it would appear that those creating the UKCAT had this in mind given the focus from an early stage of widening participation. In this context fairness might be discussed in terms of applicant support both from schools/college and family background. To an extent, similarly linked to widening access, some applicant groups perform better than others in terms of school leaver examinations which raises issues about fairness of opportunity for some applicant groups.

Selectors would have been aware of the extent to which applicant subgroups performed differently at A-level (and other school leaver qualifications). There appeared to have been a desire at the time to create a selection tool where applicant performance would be less impacted by demographic differences; an aspiration to create a test which genuinely tested innate abilities in some of these desirable traits. It is unclear the extent to which this aspiration was informed by the literature with some evidence available that aptitude tests provided little predictive value (McManus, Powis et al. 2005). My interviews with admission tutors would suggest that selectors would be more sceptical about any high stakes test being able to level the playing field in this way. McManus et al did go on to suggest that medical schools might wish to create their own test emphasising the need for a research agenda to run alongside such a development. UKCAT accepted from the start that creating an evidence base around the test itself and improving the evidence base around selection in general was an important objective. The scarcity of UK evidence around selection at the time (which is commented on in Chapter 3) would have contributed to this ambition.

At the point that UKCAT was being created, widening participation was a particularly hot topic given the recent publication of the Schwartz report (Schwartz 2004). The report highlighted under-represented groups in higher education, again called for greater fairness in selection and highlighted the challenge in selecting to the most competitive programmes such as medicine and dentistry.

Whilst those involved in creating the test emphasised the need for the test to include innate abilities contributing to learning, there was also a desire from the outset to include measures of those traits deemed to be desirable in good doctors and dentists. This ambition had been noted in Tomorrows Doctors (General Medical Council 1993, General Medical Council 2003). The test as first delivered did not include any assessment of non-academic traits. These were trialled in the test from 2007 and 2010 and the UKCAT SJT not introduced operationally until 2013 testing.

The drivers to create the UKCAT were varied. The literature pointed to the limited value in some existing selection tools and national policy was providing a clear steer to universities in the direction of widening access. At the same time selectors were struggling with the difficulties of selection between the many (due to grade inflation) high performing applicants, objectively and efficiently. The appeal of a selection test that might reduce the reliance on time consuming examination of personal statements would have been significant. Given the potential candidate numbers which the Consortium would have anticipated taking the test even in the early days, a computer-based test would have been seem as desirable which would have immediately steered the Consortium down a certain route in terms of test content. Taken together these factors may explain why the test was so widely adopted from the outset.

8.3 How was the UKCAT Consortium established and how has it developed? How has the test changed over time?

A timeline of organisational development is presented in Chapter 4, drawing upon minutes of relevant meetings and available documents. From preliminary discussions amongst a small group of universities, a consortium of 21 universities was created in 2006. UKCAT now comprises 33 universities including a number of new universities which did not even deliver medical programmes at the outset. Candidate numbers have increased from 18,540 in 2006 to 37,230 in 2021.

Ongoing development, both organisational and within the test itself, feels iterative and at times tentative, perhaps reflecting the collaborative nature of the organisation and the need for change to embed. Once firmly established within medical selection, change becomes challenging in terms of preparing universities and candidates for the impact on selection. Iterative change in computer-based testing can be carefully planned and anticipated; more significant change can have unintended consequences. For example, whilst issues relating to the Decision Analysis subtest were apparent from relatively early on, attempts to fix this had a relatively low impact with the ultimate (significant) decision to replace the subtest not taken until the 2016 testing.

Evaluation and evidence inform decision making within the Consortium and the commitment to create a research evidence base around the test is evident from the outset. Creation of a research infrastructure takes time, but the commitment is eventually realised with publications, research and analysis regularly presented to the Consortium. More recently it would appear that the focus of research priorities is less well articulated. This may reflect the fact that the original research ambitions of the organisation have been achieved; it is important however, that the evidence base around the test continues to be strengthened.

Creating a research infrastructure and in particularly establishing the research database was not without challenges. A major achievement of the organisation was the spearheading of efforts to link selection data in medicine to undergraduate and postgraduate education performance, ultimately leading to the creation of the United Kingdom Medical Education Database (UKMED). UKMED now allows researchers access to a significantly richer dataset across all medical schools leading to a greater and broader focus on research into selection across medical training. Research relating to dental selection remains sparse; this is discussed below.

There is evidence of close attention paid throughout by the Consortium to the candidate experience and how this could be improved. Policies to support the candidate experience were regularly revisited and developed with input from Consortium universities. Particular attention has been paid to widening access candidates with efforts made to expand and simplify bursary criteria and an increasingly large offering of free preparation materials.

Given testing numbers and the complexity of test construction it is perhaps unsurprising that there have been some significant issues affecting test delivery over time. In the timeline in Chapter 4, the need to withdraw results from subtests on two occasions is noted and there is a further occasion requiring the rescoring of some candidate tests.

The most significant changes to the test itself have been the replacement of the Decision Analysis subtest with Decision Making and the introduction of the UKCAT SJT. Concerns regarding the reliability of Decision Making were noted from the first year of delivery and several attempts made to improve this subtest prior to withdrawal from the test in 2014.

A significant focus from the outset in terms of test content has been around the ability to assess non-cognitive traits reliably in a high stakes environment to a largely novice population. Early attempts to assess such traits struggled to produce meaningful results that could be reliably tested in a high stakes environment. In 2021 the Consortium started to work with Pearson VUE and the Work Psychology Group to develop the UKCAT Situational Judgement Test. The UKCAT SJT is now an established part of the test but not used operationally in selection by all universities (as discussed in Chapters 6 & 7), reflecting a need for further work on the subtest and its evidence base.

The review of UKCAT documents overtime shines light on some of the recurring concerns and challenges which the Consortium has grappled with. Concerns regarding commercial preparation companies, the quality and cost of their advice and the impact on test outcomes had been expressed from the start. Other common themes are around performance differences between candidate subgroups and, in particular, between standard test takers and those taking the extended test. Linked to this discussion have been efforts over time to better understand the extent to which the test is speeded (the ability of candidates to complete each subtest within allocated timings) and how this might be ameliorated.

There have been discussions with other professions regarding the use of the test in their selection processes which have not had positive outcomes. On occasion universities have used the test for additional programmes (e.g. dental hygiene, physician associates) but these have not led to more significant developments. A number of international universities have used the test as associate members of the consortium. In 2021 the associate members were the American University of the Caribbean and the University of Kurdistan Hewler.

In 2018 UKCAT entered into discussion with a consortium of universities from Australia and New Zealand who were at the time using UMAT within their selection processes. This partnership has evolved into an international collaboration with the test being used by 16 universities in Australia and New Zealand with over 14,000 candidates testing in 2021.

Admission tutors comment positively in interviews on the collaborative nature of the organisation, welcoming opportunities to work with colleagues to share best practice and learn from the literature. The development in Australia and New Zealand may well present further opportunities for international collaboration in future years.

In 2020 due to actual and potential impact of the COVID pandemic on candidates and test centre availability (in the UK and internationally), an option to sit the test online was offered for the first time. 32% of candidates took this opportunity. Whilst this development ensured that the test could be delivered in this very challenging year, delivery was not without challenges. Smaller numbers took the online test in 2021 and their experience has been carefully monitored.

In looking to the future, UKCAT is perhaps left to reflect on how it intends to develop beyond 2020 drawing on the experience of delivering the test during the COVID 19 pandemic. Delivering the test online with appropriate levels of customer services may well be the future of mass testing, having learnt the lessons from 2020.

8.4 What is the evidence base around the utility of the UKCAT itself?

There is now a considerable evidence base around the UKCAT. Chapter 5 provides an overview of the literature relating directly to UKCAT since 2006. This includes research commissioned and undertaken by the organisation alongside work undertaken independently of UKCAT.

8.4.1 Predicting Undergraduate Performance

The systematic review (Chapter 5) of the ability of the test to predict performance in medical and dental programmes was able to draw on data from 15 medical schools studies (Greatrix, Nicholson et al. 2021). There were insufficient dental findings to be included in the meta-analysis. This was the first attempt to synthesise these data. UKCAT predicts performance in medical school although this relationship is weak. The Verbal Reasoning subtest is the most predictive of the individual subtests. Researchers have speculated this maybe because of the heavy science load of school leaver curriculum for the majority of applicants (McManus, Dewberry et al. 2013). This raises issues for the test content and structure, perhaps supporting the need for verbal reasoning to be weighted differently to the other subtests given this finding.

Two of the larger predictive validity studies confirm that the test has incremental validity (though small) over school leaver qualifications (McManus, Dewberry et al. 2013, Tiffin, Mwandigha et al. 2016). Tiffin also confirms that the test continues to predict performance over the five years of medical programmes, whilst the ability of school leaver qualifications declines over the same period.

Stronger relationships with performance at medical school tended to be observed in smaller studies (single and dual centre) rather than those drawing on national datasets. Both types of study are of course valid. The power of large studies with

consistent demographic markers and outcomes is important but local studies can analyse assessment outcomes at a more granular level.

Whilst the fact the test predicts performance at medical school is reassuring, it is stressed in Chapter 5 that this relationship is weak. To an extent this might be predicted. The ability of school leaver qualifications to predict university performance is widely recognised and continues to be the key selection criteria. Restriction of range impacts on the power of selection criteria to predict outcomes. This is a high performing group of applicants (which restricts range in the first place) and as admission is based on UKCAT scores at least to some extent, the range of scores of admitted students is further restricted. Whilst correcting for range restriction is possible this is not always undertaken by researchers. Reporting relationships with and without correction allows readers to interpret results more fully.

Given the diverse audience interested in the selection literature it is unfortunate perhaps that findings in many of these studies are not always interpreted for a lay audience. Selection to medicine and dentistry is often under the spotlight; perhaps if applicants and their advisers understood the efforts undertaken to establish a research evidence base, they might be more accepting of processes.

Both test content and the candidate population change over time and so there is a need for predictive validity studies to take place at reasonable intervals in the organisation's development. For example, the Decision Making subtest had been operational in the test since 2016 but no studies have as yet investigated its ability to predict performance in medical and dental programmes. The creation of the United Kingdom Medical Education Database (UKMED) provides opportunities for full cohort studies to take place with the potential to investigate relationships between test performance and both undergraduate and postgraduate outcomes. Smaller local studies will continue to be important for the reasons outlined above.

The absence of reliable outcome markers against which to measure the performance of the UKCAT SJT remains a challenge, perhaps signalling the need for different approaches to measure the validity of this subtest. More recently the Work Psychology Group have started working with medical and dental schools to investigate the relationships between the SJT and local selection outcomes (e.g. interview scores) which has some potential to provide additional evidence regarding to utility of this subtest.

8.4.2 Does UKCAT predict anything else?

UKMED has facilitated studies looking more broadly at the ability of selection criteria to predict outcomes in postgraduate training and clinical performance. Higher UKCAT scores have been associated with successful completion of programmes by graduates and with the FPAS Educational Performance measure (EPM) and SJT (Garrud and McManus 2018). UKCAT predicted performance in MRCS examinations (Ellis, Brennan et al. 2021).

There are numerous studies currently underway using UKMED data which may throw further light on the ability of the test to predict these longer term markers.

8.4.3 Non-cognitive tests

A number of non-cognitive tests were trialled in the early years of UKCAT and then withdrawn from the test in 2010. Subsequent analysis suggested these subtests did not predict outcomes in medical school (MacKenzie, Dowell et al. 2017, Finn, Mwandigha et al. 2018).

The evidence base specific to the UKCAT SJT is limited. Small relationships have been identified between the UKCAT SJT and a bespoke supervisor rating (Patterson, Cousans et al. 2017). However the broader literature around the use of SJTs in selection is more positive although much of this evidence is based around their use in postgraduate selection (Webster, Paton et al. 2020). The authors note the challenge in identifying appropriate outcome measures against which to judge SJTs; they also raise an issue around the extent to which individual performance in traits may change over time.

8.4.4 Widening Access

Performance in the UKCAT is sensitive to a range of demographic factors including socio-economic class. There is some evidence that the cognitive subtest is less sensitive than A-levels (James, Yates et al. 2010, Tiffin, McLachlan et al. 2014) and that the UKCAT SJT is less sensitive to socio economic class than other selection criteria (Lievens, Patterson et al. 2016).

One study suggested that candidates from lower socio-economic groups were more likely to receive an offer from schools using the test strongly in selection (Tiffin, Dowell et al. 2012). However a later study found no evidence that the introduction of the UKCAT impacted positively on the demographics of applicants admitted to medical school (Mathers, Sitch et al. 2016). In fact there is some evidence that changes in selection processes in general have had little impact on candidate demographics (Fielding, Tiffin et al. 2018). An understanding of how socio economic class impacts on both school leaver qualification and UKCAT performance has led many universities to require lower scores in the test for candidates from widening access backgrounds (Greatrix and Dowell 2020).

8.5 How has the UK admission landscape changed since 2006 and what have been the drivers around this?

Chapter 6 draws on information collected from annual interviews with admission tutors since 2006, to describe how use of the UKCAT in selection has changed over time. Inevitably this provides a broader view of how selection to medicine and dentistry has shifted over this period.

UKCAT is now undoubtedly a significant feature in selection to medicine and dentistry in the UK, used by 30 out of 38 medical schools and all but two dental schools in the 2021 admission cycle. Medical schools have overtime used the test with increasing impact on selection decisions. In 2018 the mean average UKCAT

weighting for medical schools using the factor method was 39% (dentistry 30%); 13 medical schools used a threshold score to identify those candidates to interview. Many schools use the test more than once in selection process with an increasing number of schools using the SJT bandings at some point with the most common usage to exclude the lowest performing candidates. As reported in Chapter 7, admission tutors, whilst really supportive of the SJT's inclusion in the test, remain concerned around the evidence base to support its use more strongly.

Chapter 7 reports outcomes from qualitative analysis of interviews with current admission tutors. Whilst it is clear from these interviews that UKCAT is now an embedded feature of selection processes, the impact of the test on selection decisions is perhaps less clear. For some admission tutors the very existence of a threshold score pointed to the test having a significant impact on selection outcomes. However, some tutors felt the impact was perhaps less significant given the large number of candidates being interviewed as a proportion of their applicant numbers. Threshold scores might be regarded as having a greater impact than a factor approach; however if GSCE (or equivalent) scoring fails to discriminate between high achieving candidates, the use of UKCAT as a factor could be greater than that intended. In the absence of a common and evidence based understanding of UKCAT's impact on selection, messaging to candidates can be confusing and result in them over (or under) emphasising the impact of the test in selection.

Change has been driven by government (particularly around widening participation), regulators (with the GMC increasingly articulating its expectation around selection) and the literature. The Medical School's Council Selecting for excellence report (Medical Schools Council 2014), drew together this evidence to provide guidance to the sector. As reported in Chapter 7, current admission tutors note the impact of this document on change in selection. Whilst not providing a template for selection, the report's advice to medical schools was to combine both academic achievement and aptitude tests within selection and to move towards the use of multiple mini interviews (replacing more traditional structured or semi-structured interviews). Chapter 6 notes that by 2018 all medical schools in the UK were using an admission test as part of the selection process and most were using MMIs. The systematic review (Patterson, Knight et al. 2016) that followed this report provided additional evidence for this approach, at the same time supporting the use of SJTs to assess non-cognitive traits.

By 2018 all but a handful of schools had ceased to use personal statements within selection processes. There had been a growing unhappiness with the importance of personal statements in selection processes given the limited evidence regarding their utility and concern about authenticity. There is no doubt that the UKCAT provided selectors with an alternative tool to replace personal statements in which they had greater confidence. At the same time, reviewing personal statements was time consuming and costly. The attraction of a simple tool to discriminate between

candidates must have been very attractive. In Chapter 7 admission tutors comment that UKCAT allows them to discriminate between candidates efficiently and objectively and to defend their decisions.

Over this period there has both a focus on widening access in selection processes alongside investment by universities in supporting applicants with widening access backgrounds through the application cycle. By 2018 a significant majority of schools were adjusting their selection criteria/processes for widening access candidates whilst at the same time many schools were providing alternative routes into medicine and dentistry (such as foundation programmes) for these applicants. A growing number of schools adjust UKCAT criteria for this applicant group with some schools not even using the test in selection for widening access candidates/routes.

UKCAT has developed links with Consortium university access teams and national organisations supporting these activities (e.g. Realising Opportunities, Sutton Trust) and provides support in various ways to these applicant groups. This support has more recently shifted from the provision of hard copy materials to electronic information and attendance at online events. Given UKCAT's limited resources, online presentations and workshops has created an opportunity to reach far higher numbers of potential UKCAT candidates.

Since 2006 new medical schools have been created (Aston, Anglia Ruskin, Kent Medway, Sunderland and Edge Hill). A further set of schools are currently looking to recruit from the international market with the aspiration of training home students. The number of places at medical school has therefore increased over time. For 2018 entry UCAS reported that the number of applicants to medicine courses has risen significantly, by 8% to 20,730 (+1,520 people), ending a three-year trend of decreases, following a peak of 22,740 applicants in 2014. For 2020 entry this figure had increased to 23,710; in 2021, no doubt impacted by the COVID pandemic the figure was 28,690. The national picture therefore is one of increasing application numbers offset to an extent by increased student numbers (2018 intake 7,765, 2020 intake 9,450 source Office for Students) following Government action to increase numbers in training. The 2021 intake figure was further adjusted upwards (to 8,032) following the ongoing impact of the pandemic on A-level and equivalent grades. Whether following the pandemic there is political will to increase training numbers further remains to be seen.

Candidate numbers are likely however to continue to be impacted by the significant projected increase in school leavers over the next ten years. Therefore, even putting the pandemic to one side there is a pattern of increased applicants alongside a convergence in selection processes. Medical and dental schools are largely selecting candidates for interview using a combination of academic achievement and admission test outcomes. This may result in high achieving candidates being invited to more than one interview, having the knock-on effect of schools needing to interview more applicants and make more offers. This trend may not be sustainable and requires analysis to inform what is actually happening.

The potential shift to some sort of post-qualification admission system (as signalled by the current Government), whilst not without its own challenges, might assist schools with this issue.

Chapter 6 whilst providing a thorough overview of selection processes for medicine and dentistry, at the same time illuminates the complexities of different selection processes. Working with the UKCAT Consortium over the years has given me a unique insight into how complex that system is. Selection processes are at the discretion of individual medical and dental schools who also make their own decisions about the information they provide to applicants and in what form. Whilst efforts have been made to synthesise these data (https://www.medschools.ac.uk/studying-medicine/making-an-application/entry-requirements, https://www.ucat.ac.uk/results/how-universities-use-the-ucat/), it is really difficult for candidates to make informed decisions in a shifting environment with incomplete data. Undoubtedly some candidates are better informed than others which may give them an edge in their applications. Medical and dental schools ought to continually review how they provide information to applicants and whether greater consistency in presentation of this information would be of benefit.

8.6 What did the UKCAT Consortium set out to achieve and to what extent has it been successful?

8.6.1 Has UKCAT helped achieve greater fairness in selection?

As commented on above, it is not entirely clear what was meant by fairness in terms of organisational goals. Certainly, discussions in 2006 suggest a desire to shift away from selection criteria which had no evidence base, and which were regarded as being significantly influenced by social background and schooling such as personal statements. Fairness may also have linked to the need to be able to discriminate in a more objective way between high achieving candidates. At the same time the focus on UKCAT being a test of innate abilities possibly pointed to a desire to test traits which might have been deemed (again) less influenced by social background, schooling and coaching.

As reported in Chapter 7, admission tutors felt that the addition of a tool with greater objectivity into selection processes in itself increased fairness. Whilst commenting on fairness at a global level in this way, many tutors remained concerned about fairness in regard to some candidate subgroups and widening access candidates in particular. It is perhaps inevitable that if the test is widely used as a screening tool that this would impact (in the absence of any other adjustments) on some candidate subgroups more than others. This is discussed below.

Candidates have been sceptical from the outset about how fair the UKCAT actually is (Samuel 2008, Cleland, French et al. 2011, Lambe, Greatrix et al. 2016); although candidates who had scored better in the test were more likely to consider the test to be a fair way of discriminating between applicants.

8.6.2 Has UKCAT contributed to widening participation?

In Chapter 7 tutors recognised as positive the work UKCAT has undertaken in terms of financial support for candidates from low income families (through the bursary scheme) and the provision of extensive free preparation materials. The work undertaken in both these areas is noted in Chapter 4 with progress recently in reaching out to widening access candidates through online presentations being particularly valuable.

Some admission tutors, however reported using the test differently for widening access candidates reflecting a view that this candidate group remain disadvantaged in terms of performance in the test. Chapter 4 presented data showing that 76% of candidates are from SEC 1 and that this proportion has hardly shifted over time. With increasing candidate numbers, there is at least an increase in the actual number of candidates from the lower socio economic groups. However as presented in Figure 9 in Chapter 4, on average, candidates from SEC4&5 perform worse than those in SEC1. In 2020 candidates in SEC5 scored on average 113 scale points lower than candidates in SEC1. There are also performance differences observed between ethnic groups and those candidates for whom English is not a first language.

As commented on in Chapter 4, the demographic make up of test takers (compared to the UK population) and the extent to which different demographic markers interact and impact on subgroup differences in the test is not well enough understood. Without this information and additional investigation of demographic impact at subtest, item type and item level, it is not possible to start to identify developments to the test that might impact on these differences.

Alongside this, qualitative approaches to better understand how different candidate groups approach and prepare for the test may assist in understanding the data and improve advice on test preparation. The availability of performance data by subgroup has provided universities with the information they need to make decisions about how to use the test for widening access candidates. The move to contextualising the use of test scores (in a similar way to school leaver qualifications) has been informed by these discussions.

Linked to both discussions around fairness and widening access are concerns regarding the perceived and actual effects of coaching on performance impacting on these performance differences.

8.6.3 Has UKCAT improved the research evidence base around selection to medicine and dentistry?

Chapter 5 comments on the growing evidence base around medical selection in particular. A systematic review published in 2002 (Ferguson, James et al. 2002) drew on evidence from only 8 (of 51) articles from the UK. Chapter 5 includes reference to at least 30 articles where the focus is specifically on the UKCAT. There is no doubt that UKCAT has been a catalyst for research activity, with the

organisation directly funding a number of studies and supporting others with relevant data.

UKMED grew out of early collaborations between UKCAT and the Medical School's Council (MacKenzie, Cleland et al. 2016, MacKenzie, Dowell et al. 2017) having demonstrated the potential of linking UKCAT and Foundation Year data. Building on this work, UKCAT urged the Medical Schools Council and GMC to consider developing a database to contain data relating to medical education and the workforce; drawing on datasets from application, through performance in undergraduate and postgraduate training through to professional outcomes.

Building on original pilot data, the UKMED database at the outset contained data only from early UKCAT test takers (2007 and 2008) (Dowell, Cleland et al. 2018) but has rapidly grown to link data from many providers (The Higher Education Statistics Agency, UK Clinical Aptitude Test, Graduate Australian Medical Schools Admissions Test, General Medical Council, UK Foundation Programme Office, Postgraduate Deaneries and Local Education and Training Boards, UK National Recruitment Offices). The database is managed by the General Medical Council. UKCAT is represented on both the Advisory Board and Research Sub-Group.

Researchers may apply to undertake research on UKMED data which is supported through a safe haven ensuring the security and anonymity of data. At time of writing 12 studies had been published by researchers working on UKMED data. Whilst many of these are studies reported elsewhere in this chapter, other work includes analysis regarding speciality choice (Gale, Lambe et al. 2017, Kumwenda, Cleland et al. 2018, Lambe, Gale et al. 2019) and recruitment (Santana and Chalkley , Kumwenda, Cleland et al. 2018, Kumwenda, Cleland et al. 2018, Woolf, Jayaweera et al. 2019). A further 29 studies are underway looking at a diverse range of topics (https://www.ukmed.ac.uk/accepted_applications) including a number focussing on demographics and performance in the specialties, equality and diversity issues and predictors of non-standard outcomes (e.g. GMC sanctions). UKCAT is working closely with the United Kingdom Medical Applicant Cohort Study (United Kingdom Medical Applicant Cohort Study n.d.) which is using UKMED data together with questionnaire data obtained from medical applicants to better understand how candidates make medical school choices when applying.

In Chapter 7, admission tutors commented positively on the role UKCAT had had in supporting research and improving the evidence base around selection. It is interesting that some admission tutors were able to speak very confidently about the literature and impact on local practice. Other tutors, though aware in broad terms of the evidence base were less confident. As the evidence base grows this task is going to be even more challenging; perhaps suggesting that organisations such as UKCAT, the Medical/Dental Schools Councils and widening access organisations ought to collaborate to provide a repository of relevant studies alongside interpretation of findings.

Work related to dental selection is much harder to come by although UKCAT has recently supported two studies focussing on widening access in dentistry and predictive validity of selection criteria. No doubt some of the medical findings can be generalised to dental selection but the applicant group and selection criteria are different and warrant investigations of their own.

8.7 Other considerations

Even though UKCAT is now regarded as a fixture in selection to medicine and dentistry, I am conscious that some of the very criticisms levelled at the test from the start remain challenges for the organisation.

Competition for places at medical school in particular has grown, making it increasingly hard to select from high performing candidates. However, these candidates very understandably often report to UKCAT that they do not feel it is fair that their actual and predicted academic achievement is not sufficient for them to be selected; and that their performance in a two-hour test could be the deciding factor.

Applying to medicine and dentistry is already complex with additional requirements around work experience and interviews. Adding into that mix, a test which the candidate must pay to take is seen as another barrier to getting to medical and dental school.

As commented above there was some disagreement amongst admission tutors when quantifying the impact of UKCAT on selection outcomes. Applicants however clearly think that the test plays a significant part in these decisions. In the absence of alternative methods of discriminating between candidates it is unclear how universities can change their reliance on the test.

A sizeable coaching industry has grown up around the UKCAT with numerous companies making claims about their ability to improve performance in the test. The extent to which the claims of these commercial companies are true is difficult to assess. Certainly, it is important to be well prepared for the UKCAT. Candidates need to be familiar with the different question types. Timings can be challenging in some of the subtests and candidates need to have prepared enough to be reasonably confident of reaching the end of the questions. UKCAT provides extensive preparation materials and advice to candidates. Surveys of candidates (reported in Chapter 4) suggest that the use of commercially available preparation materials has grown over time. 56% of survey respondents report using these in 2017 (Kulkarni S. 2020). Perhaps the utility of the advice and coaching provided by some of these companies has improved over time. It is possible that such materials and advice provide candidates with an edge they otherwise would not have. It is likely to be the case that being steered through the preparation journey may provide a benefit to some candidates over others left to navigate themselves. More likely however I think it gives those candidates a confidence in approaching the test that they might not otherwise have. If there is an intrinsic value placed on paid for

preparation materials, then it is really challenging to persuade candidates that freely available materials (of whatever quality and quantity) are sufficient.

The actual impact of preparation and coaching on test performance is impossible to assess (without candidates having tested prior to the intervention). However, reported group differences in test performance and accessing preparation materials are of concern.

UKCAT aims to provide sufficient advice and preparation materials to support candidates in familiarising themselves with test content and to put in place effective strategies to perform well. Whilst accepting that the availability of paid for preparation materials potentially creates an uneven playing field for candidates there is a debate to be had regarding the extent to which UKCAT, as the test provider, should seek to fully redress this imbalance.

The understanding of the impact of coaching on aptitude tests is limited and complex. UKCAT might wish to consider whether further interpretation of the literature might help counter the claims of coaching companies and help Consortium universities understand this impact better.

8.8 The Future: How might the organisation and the test develop?

Looking to the future against a backdrop of the COVID pandemic is difficult. Whilst online proctored delivery of the test was not without challenges, it is now not inconceivable that soon this could be the main delivery mechanism for the test. This may well be a preference on the part of many candidates, removing travel costs and offering greater flexibility and perhaps being more cost effective. However, whilst the technology supporting online delivery will no doubt continue to develop, concerns are likely to remain about the integrity of test content and increased opportunities to cheat. UKCAT will only know whether such concerns are valid or only perceived by going down that path. I would be concerned that consistently supporting online delivery for candidates testing at home is more difficult, especially when many problems are likely to be created by the candidate's own testing environment and internet connectivity.

Recent increases in candidate numbers in themselves represents a challenge both to UKCAT and more broadly in selection to medicine and dentistry. Candidate numbers have increased by 10,000 since 2017. In is unclear what is driving this trend though it might be speculated that during the pandemic the presence of health stories and visibility of health professionals might have created additional demand. It is also unclear whether this increase will be sustained as the UK emerges from the pandemic. I am acutely aware that whilst many administrative processes can be scaled up, there is a danger that the focus on customer service is lost whilst processing large candidate volumes. Some processes/deadlines almost certainly require adjusted for these new candidate numbers if they are to be sustained.

In 2020 and 2021 due to the COVID pandemic, pupils in the UK did not sit GCSEs and A-levels (and equivalent). Reliance on teacher calculated grades in 2020 and teacher assessed grades in 2021 led to significant grade inflation. This created challenges for medical and dental selection with more applicants than anticipated achieving their grades and being admitted to courses. Short term fixes were put in place by increasing student numbers and attempting to redistribute successful applicants around medical schools. However, even before this development the UK government had announced its intention to move towards some sort of post-qualification admission process (DFE 2021). Such developments have been discussed for some time but there has been increasing disquiet about discrepancies between predicted and achieved grades and the impact this has on widening access right across higher education. The government has consulted on these proposals and intended to present a response in summer 2021.

Radical change to selection processes, will create both challenges and opportunities for medical and dental schools in general and UKCAT in particular. In terms of broader selection, universities will welcome the opportunity to select candidates based on achieved grades. However, the selection process for health programmes requires interviews to take place and it is not clear how this might fit into any post-qualification admission framework. For UKCAT this may mean a change in the testing window (with testing perhaps taking place later in the academic year) and potentially fewer candidates depending on test timings. As mentioned in Chapter 4, in its response to the government consultation, UKCAT commented on the flexibility in being able to deliver the test at a time required by medical and dental schools. In addition, the response noted the potential difficulties of testing clashing more directly with preparation for public examinations.

8.9 Thesis Recommendations

The following provides a summary of recommendations made throughout this thesis. Operationalising these recommendations has the support of the UKCAT Chair. The intention is to present these recommendations to the UKCAT Board and Consortium to feed into future work and contribute to the further development of organisational aims.

8.9.1 Test Content

- UKCAT should consider the findings from the systematic review in relation to the future development and use of the UKCAT itself. Verbal Reasoning clearly plays a great part in the relationships observed, supporting perhaps a case for this subtest to have a higher weighting or be treated differently in selection.

- Given ongoing drivers for change in the test (e.g. speededness, utility of the SJT) UKCAT should actively engage with Consortium universities regarding future test content and its structure so that they have a stake in development.

- Additional analysis needs to take place on an annual basis to help further understand candidate demographics alongside subgroup performance differences within the test. The extent to which changes to test content might reduce such differences ought to be explored.

8.9.2 Research

- UKCAT should commission a large cohort study every five years to ensure that validity evidence remains up to date.
- Those undertaking future cohort studies should be advised to:
    - Provide consistent detail regarding assessment outcomes being investigated.
    - Interpret findings clearly for the benefit of selectors (and test takers).
    - Consider performance in the middle years of medical school as fewer studies have looked at years three and four.
    - Focus on the lack of evidence regarding dentistry.
    - Undertake analysis of relevant sub-group differences (e.g. age, gender, international, widening access) regarding prediction of university outcomes.
    - Include analysis which adjusts for range restriction or at least comment explicitly on the limiting impact of not undertaking such analysis.
- Studies investigating the extent to which the UKCAT might compensate for lower A-level achievement and the potential impact on opening up routes to widen access ought to be considered.
- Data collection regarding the use of the test in selection by universities ought to be expanded to include greater detail regarding adjustments made for widening access candidates.
- Work is required to better understand the impact UKCAT is having on selection decisions to support medical and dental schools in refining selection processes. This may include investigating the extent to which the convergence of selection processes is impacting on conversion rates and interview volumes.
- A summary of the literature around the impact of coaching on aptitude testing ought to be undertaken to better understand the impact on overall test performance and that of candidate subgroups. Such work can further inform selection and potentially counter some of the claims of commercial organisations.

- Qualitative approaches to help better understand the performance differences in the test between candidate subgroups ought to be considered.
- UKCAT should work with other national organisations to create a repository of relevant selection research, providing appropriate interpretation for admission tutors, applicants and their advisors.
- Further work with the Consortium is required to understand and improve the evidence base around the UKCAT SJT.

8.9.3 Candidates

- Working with Consortium universities, UKCAT should seek to improve and make more consistent information provided to candidates regarding the use of the test by universities in selection.
- Further consideration ought to be given as to how to challenge claims of commercial coaching companies regarding how their services can impact on candidate test outcomes.
- UKCAT should consider in the light of the now established commercial coaching industry, the extent to which UKCAT should develop preparation materials and the focus of such materials.

8.9.4 General

- UKACT should continue to keep the record of test and organisational development (as summarised in Chapters 4) updated to inform future research where an understanding of changes in the test over time would be helpful.

8.10 Conclusion

This thesis set out to evaluate the impact UKCAT has had on selection to medicine and dentistry. To investigate impact in a broad sense, information has been presented regarding the development of the organisation and the test itself. The literature has been explored specifically regarding the test's predictive validity and more generally in terms of UKCAT specific research outcomes. How selection has changed since 2006 had been described drawing on information collected from medical and dental schools on an annual basis. Admission tutor views on UKCAT's impact and the extent to which the organisation has achieved its original aims has been explored qualitatively.

The test is now firmly embedded in selection processes and plays a significant part in decisions made regarding most applicants to undergraduate medical and dental training. There is evidence that the test is a more valid selection criteria than the scoring of personal statements which was a common selection tool prior to the introduction of the test. The test provides universities with an effective tool for discriminating between the many high performing candidates.

The test in itself does not appear to widen participation and universities are increasingly using the test differently in selection for this candidate subgroup.

Advances in technology brings with it opportunities in terms of test delivery whilst at the same time demanding greater vigilance regarding candidate fraud. The landscape around selection continues to develop, impacted by COVID, increasing applicant numbers and the potential of a post-qualification framework for admissions around the corner.

References

Adam, J., M. Bore, R. Childs, J. Dunn, J. McKendree, D. Munro and D. Powis (2015). "Predictors of professional behaviour and academic outcomes in a UK medical school: A longitudinal cohort study." Medical teacher 37(9): 868-880.

Adam, J., M. Bore, J. McKendree, D. Munro and D. Powis (2012). "Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study." BMC medical education 12(1): 69.

Adam, J., J. Dowell and R. Greatrix (2011). "Use of UKCAT scores in student selection by U.K. medical schools, 2006-2010." BMC medical education 11.

Adam, J., J. Dowell and R. Greatrix (2011). "Use of UKCAT scores in student selection by UK medical schools, 2006-2010." BMC Med Educ 11.

Adrian Husbands, J. D., Matthew Homer, Robert McAndrew and Rachel Greatrix (2018). Exploring the Relationship between the UKCAT Situational Judgement Test and the Multiple Mini Interview.

Albanese, M. A., M. H. Snow, S. E. Skochelak, K. N. Huggett and P. M. Farrell (2003). "Assessing personal qualities in medical school admissions." Academic Medicine 78(3): 313-321.

Arulampalam, W., R. Naylor and J. Smith (2004). "Factors affecting the probability of first year medical student dropout in the UK: a logistic analysis for the intake cohorts of 1980–92." Medical Education 38(5): 492-503.

Arulampalam, W., R. A. Naylor and J. P. Smith (2007). "Dropping out of medical school in the UK: explaining the changes over ten years." Medical Education 41(4): 385-394.

Barkley, R. (1976). "The dilemma of dental school admissions: a possible solution." The Journal of the American College of Dentists 43(2): 119.

Blackmur, J. P., N. I. Lone, O. D. Stone, D. J. Webb and N. Dhaun (2016). "A Cross-sectional Study of Current Doctors' Performance in a Modified Version of a Medical School Admission Aptitude Test: The UKCAT." Medicine 95(18): e3506.

Bowen, G. A. (2009). "Document Analysis as a Qualitative Research Method." Qualitative Research Journal 9(2): 27-40.

Bowman, N. A. (2012). "Effect Sizes and Statistical Methods for Meta-Analysis in Higher Education." Research in Higher Education 53(3): 375-382.

Boyd, M. A., W. Teteruck and G. Thompson (1980). "Interpretation and use of the dental admission and aptitude tests." Journal of Dental Education 44(5): 275-278.

Braun, V. and V. Clarke (2006). "Using thematic analysis in psychology." Qualitative Research in Psychology 3(2): 77-101.

Carrothers, R. M., S. W. Gregory Jr and T. J. Gallagher (2000). "Measuring emotional intelligence of medical school applicants." Academic Medicine 75(5): 456-463.

Cassidy, J. (2008). "Medical education: UKCAT among the pigeons." BMJ: British Medical Journal 336(7646): 691.

Cleland, J. (2015). "Exploring versus measuring: considering the fundamental differences between qualitative and quantitative research." Researching medical education: 1-14.

Cleland, J., J. Dowell, J. McLachlan, S. Nicholson and F. Patterson (2012). "Research report: Identifying best practice in the selection of medical students (literature review and interview survey)." London: General Medical Council.

Cleland, J., F. French and P. Johnston (2011). "A mixed-methods study identifying and exploring medical students's views of the UKCAT." Med Teach 33.

Cleland, J., J. McKimm, R. Fuller, D. Taylor, J. Janczukowicz and T. Gibbs (2020). "Adapting to the impact of COVID-19: Sharing stories, sharing practice." Medical teacher 42(7): 772-775.

Cleland, J. A., S. Nicholson, N. Kelly and M. Moffat (2015). "Taking context seriously: explaining widening access policy enactments in UK medical schools." Medical Education 49(1): 25-35.

Cochrane, L. R., David; Guthrie, Bruce; Dowell, Jon. (2013). "Socioeconomic status of applicants to UKCAT Consortium Medical Schools 2009-2012."

Cohen, D. J. and B. F. Crabtree (2008). "Evaluative criteria for qualitative research in health care: controversies and recommendations." The Annals of Family Medicine 6(4): 331-339.

Conger, J. J. and R. H. Fitz (1963). "Prediction of Success in Medical School." Academic Medicine 38(11): 943-948.

Council, G. M. (2008). The Consensus Statement on the Role of the Doctor.

Council, G. M. (2014). The Consensus Statement on the Role of the Doctor.

Creswell, J. W. (2013). Research design: Qualitative, quantitative, and mixed methods approaches, Sage publications.

Cunningham, C., F. Patterson and J. Cleland (2019). "A literature review of the predictive validity of European dental school selection methods." European Journal of Dental Education 23(2): 73-87.

Curtis, S. and D. Smith (2020). "A comparison of undergraduate outcomes for students from gateway courses and standard entry medicine courses." BMC Medical Education 20(1): 1-14.

Davis Jr, N. S. (1903). "Requirements for admission to medical schools." Quarterly Bulletin of the Northwestern University Medical School 5(2): 241.

Department of Health (2004). "Medical School: Delivering the Doctors of the Future."

Deubert, L., M. Smith, S. Downs, C. Jenkins and D. Berry (1975). "The selection of dental students. A pilot study of an assessment of manual ability by practical tests." British dental journal 139(9): 357-361.

DFE (2021). Post-Qualification Admissions Reform. D. f. Education, HMSO.

DiCicco-Bloom, B. and B. F. Crabtree (2006). "The qualitative research interview." Medical Education 40(4): 314-321.

Dixon, R. A. (1940). "Trends In Dental Education." Journal of the National Medical Association 32(1): 30-33.

Dowell, J., J. Cleland, S. Fitzpatrick, C. McManus, S. Nicholson, T. Oppé, K. Petty-Saphon, O. S. King, D. Smith, S. Thornton and K. White (2018). "The UK medical education database (UKMED) what is it? Why and how might you use it?" BMC Medical Education 18(1): 6.

Dowell, J., M. A. Lumsden, D. Powis, D. Munro, M. Bore, B. Makubate and B. Kumwenda (2011). "Predictive validity of the personal qualities assessment for selection of medical students in Scotland." Medical teacher 33(9): e485-e488.

Dowell, J., B. Lynch, H. Till, B. Kumwenda and A. Husbands (2012). "The multiple mini-interview in the UK context: 3 years of experience at Dundee." Medical Teacher 34(4): 297-304.

Drummond, J. and R. Duguid (1997). "Student drop-out from UK dental schools." British dental journal 182(9): 347-349.

Duguid, R. (1989). "Applications and admissions to UK dental schools through UCCA: 1968-1987." British dental journal 167(10): 337-340.

Duguid, R. (1994). "Recent changes in UK dental school admissions." British dental journal 177(6): 213-216.

Duguid, R. and J. Drummond (2000). "The admission of students to UK Dental Schools–recent trends (1983–1998)." European Journal of Dental Education 4(2): 71-76.

Edwards, J. C., C. L. Elam and N. E. Wagoner (2001). "An admission model for medical schools." Academic Medicine 76(12): 1207-1212.

Edwards, J. C., E. K. Johnson and J. B. Molidor (1990). "The interview in the admission process." Academic Medicine 65(3): 167-177.

Ellis, R., P. Brennan, D. S. Scrimgeour, A. J. Lee and J. Cleland (2021). "Performance at medical school selection correlates with success in Part A of the intercollegiate Membership of the Royal College of Surgeons (MRCS) examination." Postgraduate Medical Journal: postgradmedj-2021-139748.

Esmail, A., P. Nelson, D. Primarolo and T. Torna (1995). "Acceptance into medical school and racial discrimination." BMJ : British Medical Journal 310(6978): 501-502.

Eva, K., J. Rosenfeld, H. Reiter and G. Norman (2004). "An admissions OSCE: the multiple mini-interview." Med Educ 38.

Ferguson, E., D. James and L. Madeley (2002). "Factors associated with success in medical school: systematic review of the literature." BMJ 324(7343): 952-957.

Ferguson, E., A. Sanders, F. O'Hehir and D. James (2000). "Predictive validity of personal statements and the role of the five-factor model of personality in relation to medical training." Journal of Occupational and Organizational Psychology 73(3): 321-344.

Fernando, N., G. Prescott, J. Cleland, K. Greaves and H. McKenzie (2009). "A comparison of the United Kingdom Clinical Aptitude Test (UK-CAT) with a traditional admission selection process." Medical Teacher 31(11): 1018-1023.

Fielding, S., P. A. Tiffin, R. Greatrix, A. J. Lee, F. Patterson, S. Nicholson and J. Cleland (2018). "Do changing medical admissions practices in the UK impact on who is admitted? An interrupted time series analysis." BMJ Open 8(10): e023274.

Finn, G. M., L. Mwandigha, L. W. Paton and P. A. Tiffin (2018). "The ability of 'non-cognitive' traits to predict undergraduate performance in medical schools: A national linkage study." BMC Medical Education 18(1).

Foley, J. and K. Hijazi (2015). "Predictive value of the admissions process and the UK Clinical Aptitude Test in a graduate-entry dental school." British dental journal 218(12): 687-689.

Francis, R. (2013). Report of the Mid Staffordshire NHS Foundation Trust public inquiry: executive summary, The Stationery Office.

Fredericks, M. A. and P. Mundy (1967). "The relationship between social class, average grade in college, medical college admission test scores, and academic achievement of students in a medical school." Academic Medicine 42(2): 126-133.

Funkenstein, D. H. (1962). "Failure to graduate from medical school." Academic Medicine 37(6): 588-603.

Funkenstein, D. H. (1970). "Current medical school admissions: the problems and a proposal." Academic Medicine 45(7): 497-509.

Gale, T. C., P. J. Lambe and M. J. Roberts (2017). "Factors associated with junior doctors' decisions to apply for general practice training programmes in the UK: secondary analysis of data from the UKMED project." BMC medicine 15(1): 220.

Gallagher, J., A. Calvert, V. Niven and L. Cabot (2017). "Do high tuition fees make a difference? Characteristics of applicants to UK medical and dental schools before and after the introduction of high tuition fees in 2012." British dental journal 222(3): 181.

Gangler, P., J. de Vries, L. Akota, I. Balciune, P. Berthold and M. Gajewska (2002). "Student selection and the influence of their clinical and academic enviroment on learning. University of Witten Herdeke Germany, University of Manitoba Canada." Eur J Dent Educ 6: 8-26.

Garrud, P. and I. C. McManus (2018). "Impact of accelerated, graduate-entry medicine courses: a comparison of profile, success, and specialty destination between graduate entrants to accelerated or standard medicine courses in UK." BMC Medical Education 18(1): 250.

General Medical Council (1993). Tomorrow's doctors: Recommendations on undergraduate medical education, General Medical Council London.

General Medical Council (2003). Tomorrow's Doctors.

General Medical Council (2009). Tomorrow's Doctors: Outcomes and Standards for Undergraduate Medical Education.

General Medical Council (2013). Good Medical Practice.

Goldbeck-Wood, S. (1996). "Choosing tomorrow's doctors." BMJ : British Medical Journal 313(7053): 313-313.

Goldberg, L. R. (1993). "The structure of phenotypic personality traits." American psychologist 48(1): 26.

Greatrix, R. and J. Dowell (2020). "UKCAT and medical student selection in the UK – what has changed since 2006?" BMC Medical Education 20(1): 292.

Greatrix, R. and R. McAndrew (2021). UCAT and dental student selection in the UK – what has changed?, University of Nottingham.

Greatrix, R., S. Nicholson and S. Anderson (2021). "Does the UKCAT predict performance in medical and dental school? A systematic review." BMJ open 11(1): e040128.

Greene, J. C., V. J. Caracelli and W. F. Graham (1989). "Toward a conceptual framework for mixed-method evaluation designs." Educational evaluation and policy analysis 11(3): 255-274.

Griffin, B., S. Carless and I. Wilson (2013). "The effect of commercial coaching on selection test performance." Medical teacher 35(4): 295-300.

Griffin, B., G. L. Horton, L. Lampe, B. Shulruf and W. Hu (2021). "The change from UMAT to UCAT for undergraduate medical school applicants: impact on selection outcomes." Medical Journal of Australia 214(2): 84-89.

Haggerty, M. (1918). "Tests of applicants for admission to University of Minnesota Medical School." Journal of Educational Psychology 9(5): 278.

Hamilton, J. D. (1972). "The selection of medical students at McMaster University." Journal of the Royal College of Physicians of London 6(4): 348-351.

Hanlon, K., G. Prescott, J. Cleland and R. Kmackenzie (2011). "Does Ukcat Predict Performance in the First Year of an Integrated Systems Based Medical School Curriculum?" Medical Education 45: 87.

Harasym, P. H., W. Woloschuk, H. Mandin and R. Brundin-Mather (1996). "Reliability and validity of interviewers' judgments of medical school candidates." Academic Medicine 71(1): S40-42.

Harris, A. J. (1937). "The relative significance of measures of mechanical aptitude, intelligence, and previous scholarship for predicting achievement in dental school." Journal of Applied Psychology 21(5): 513-521.

Haynes, K. (2012). "Reflexivity in qualitative research." Qualitative organizational research: Core methods and current challenges: 72-89.

Higher Education Funding Council for England (2018). New medical training places announced under expansion programme.

Hill, J. K. (1959). "Assessment of intellectual promise for medical school." Academic Medicine 34(10): 959-964.

Hoad-Reddick, G. and T. MacFarlane (1999). "An analysis of an admissions system: can performance in the first year of the dental course be predicted?" British dental journal 186(3).

Horton, R. C. (1986). "Admission to medical school: from audit to action." JRSM 79.

Howe, A., P. Campion, J. Searle and H. Smith (2004). "New perspectives—approaches to medical education at four new UK medical schools." Bmj 329(7461): 327-331.

Hughes, P. (2002). "Can we improve on how we select medical students?" Journal of the Royal Society of Medicine 95(1): 18-22.

Husbands, A. and J. Dowell (2013). "Predictive validity of the Dundee multiple mini-interview." Medical education 47(7): 717-725.

Husbands, A., A. Mathieson, J. Dowell, J. Cleland and R. MacKenzie (2014). "Predictive validity of the UK clinical aptitude test in the final years of medical school: a prospective cohort study." BMC medical education 14(1): 88.

James, D. and C. Chilvers (2001). "Academic and non-academic predictors of success on the Nottingham undergraduate medical course 1970–1995." Medical Education 35(11): 1056-1064.

James, D., J. Yates and S. Nicholson (2010). "Comparison of A level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study." BMJ 340.

Jaques, H. (2010). "UKCAT." BMJ 341: c5647.

Johnson, R. B. and A. J. Onwuegbuzie (2004). "Mixed Methods Research: A Research Paradigm Whose Time Has Come." Educational Researcher 33(7): 14-26.

Julian, E. R. (2005). "Validity of the Medical College Admission Test for Predicting Medical School Performance." Academic Medicine 80(10): 910-917.

Kalis, B. L., J. J. Tocchini and P. R. Thomassen (1962). "Correlation study between personality tests and dental student performance." The Journal of the American Dental Association 64(5): 656-670.

Kelly, L. M. and M. Cordeiro (2020). "Three principles of pragmatism for research on organizational processes." Methodological innovations 13(2): 2059799120937242.

Kelly, M. E., F. Patterson, S. O'Flynn, J. Mulligan and A. W. Murphy (2018). "A systematic review of stakeholder views of selection methods for medical schools admission." BMC Medical Education 18(1): 139.

Kennedy, I. (2001). The Report of the Public Inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995, Stationery Office.

King, N., C. Cassell and G. Symon (2004). "Using templates in the thematic analysis of text." Essential guide to qualitative methods in organizational research 2: 256-270.

Knights, J. A. and B. J. Kennedy (2006). "Medical school selection: Screening for dysfunctional tendencies." Med Educ 40(11): 1058-1064.

Knights, J. A. and B. J. Kennedy (2007). "Medical school selection: impact of dysfunctional tendencies on academic performance." Med Educ 41(4): 362-368.

Kramer, G. A. (1986). "Predictive validity of the dental admission test." Journal of Dental Education 50(9): 526-531.

Kreiter, C. D., P. Yin, C. Solow and R. L. Brennan (2004). "Investigating the reliability of the medical school admissions interview." Advances in Health Sciences Education 9(2): 147-159.

Kulkarni S., P. J., Sitch A. (2020). An assessment of the impact of formal preparation activities on performance in the University Clinical Aptitude Test (UCAT): a national study, Institute of Applied Health Research, University of Birmingham.

Kumar, K., C. Roberts, I. Rothnie, C. Du Fresne and M. Walton (2009). "Experiences of the multiple mini-interview: a qualitative analysis." Medical education 43(4): 360-367.

Kumwenda, B., J. Cleland, R. Greatrix, R. K. MacKenzie and G. Prescott (2018). "Are efforts to attract graduate applicants to UK medical schools effective in increasing the participation of under-represented socioeconomic groups? A national cohort study." BMJ Open 8(2): e018946.

Kumwenda, B., J. Cleland, G. Prescott, K. Walker and P. Johnston (2019). "Relationship between sociodemographic factors and specialty destination of UK trainee doctors: a national cohort study." BMJ open 9(3): e026961.

Kumwenda, B., J. A. Cleland, G. J. Prescott, K. Walker and P. W. Johnston (2018). "Relationship between sociodemographic factors and selection into UK postgraduate medical training programmes: a national cohort study." BMJ open 8(6): e021329.

Kumwenda, B., J. A. Cleland, G. J. Prescott, K. A. Walker and P. W. Johnston (2018). "Geographical mobility of UK trainee doctors, from family home to first job: a national cohort study." BMC medical education 18(1): 314.

Kumwenda, B., J. A. Cleland, K. Walker, A. J. Lee and R. Greatrix (2017). "The relationship between school type and academic performance at medical school: a national, multi-cohort study." BMJ open 7(8).

Lala, R., D. Wood and S. Baker (2013). "Validity of the UKCAT in applicant selection and predicting exam performance in UK dental students." Journal of dental education 77(9): 1159-1170.

Lambe, P., R. Greatrix, K. Milburn, J. Dowell, K. S. Way and D. Bristow (2016). Candidates' views on the UK Clinical Aptitude Test: Results from the Post 2012 UKCAT Survey.

Lambe, P., R. Greatrix, K. Milburn, J. Dowell, K. S. Way and D. Bristow (2016). Do differentials in access to advice and support at UK schools on preparation for the UK Clinical Aptitude Test disadvantage some candidate groups?

Lambe, P., E. Kay and D. Bristow (2016). "Exploring uses of the UK Clinical Aptitude Test-situational judgement test in a dental student selection process." European Journal of Dental Education.

Lambe, P., C. Waters and D. Bristow (2012). "The UK Clinical Aptitude Test: Is it a fair test for selecting medical students?" Medical teacher 34(8): e557-e565.

Lambe, P., C. Waters and D. Bristow (2013). "Do differentials in the support and advice available at UK schools and colleges influence candidate performance in the medical school admissions interview? A survey of direct school leaver applicants to a UK medical school." Medical Teacher 35(9): 731-739.

Lambe Paul, B. D. (2016). Can a Situational Judgement Test replace a standardised interview as part of the process to select undergrdauate medical students? OTTAWA2016. Australia.

Lambe, P. J., T. C. Gale, T. Price and M. J. Roberts (2019). "Sociodemographic and educational characteristics of doctors applying for psychiatry training in the UK: secondary analysis of data from the UK Medical Education Database project." BJPsych bulletin 43(6): 264-270.

Lavelle, E., J. Vuk and C. Barber (2013). "Twelve tips for getting started using mixed methods in medical education research." Medical Teacher 35(4): 272-276.

Layton, W. L. (1953). "Predicting success in dental school." Journal of Applied Psychology 37(4): 251.

Leech, N. L. and A. J. Onwuegbuzie (2009). "A typology of mixed methods research designs." Quality & Quantity 43(2): 265-275.

Liberati, A., D. Altman, J. Tetzlaff, C. Mulrow, P. Gøtzsche, J. Ioannidis, M. Clarke and P. Devereaux (2009). "The PRISMA statement for reporting systematic and meta-analyses of studies that evaluate interventions: explanation and elaboration." PLoS Medicine 6(7): 1-28.

Lievens, F., F. Patterson, J. Corstjens, S. Martin and S. Nicholson (2016). "Widening access in selection using situational judgement tests: evidence from the UKCAT." Med Educ 50.

Lingard, L., M. Albert and W. Levinson (2008). "Grounded theory, mixed methods, and action research." Bmj 337: a567.

Little, J. M., H. H. Gee and M. R. Novick (1960). "A study of the medical college admission test in relation to academic difficulties in medical school." Academic Medicine 35(3): 264-272.

Lumsden, M. A., M. Bore, K. Millar, R. Jack and D. Powis (2005). "Assessment of personal qualities in relation to admission to medical school." Medical education 39(3): 258-265.

Lynch, B., R. MacKenzie, J. Dowell, J. Cleland and G. Prescott (2009). "Does the UKCAT predict Year 1 performance in medical school?" Medical Education 43(12): 1203-1209.

Lynch, C., R. McConnell and A. Hannigan (2006). "Dental school admissions in Ireland: can current selection criteria predict success?" European Journal of Dental Education 10(2): 73-79.

MacKenzie, R. K., J. A. Cleland, D. Ayansina and S. Nicholson (2016). "Does the UKCAT predict performance on exit from medical school? A national cohort study." BMJ open 6(10): e011313.

MacKenzie, R. K., J. Dowell, D. Ayansina and J. A. Cleland (2017). "Do personality traits assessed on medical school admission predict exit performance? A UK-wide longitudinal cohort study." Advances in Health Sciences Education 22(2): 365-385.

Mackenzie, R. S. (1967). "Predictive validity of a biographical inventory in higher education." Journal of Applied Psychology 51(6): 544.

Mann, K. and A. MacLeod (2015). "Constructivism: learning theories and approaches to research." Researching medical education: 49-66.

Marley, J. and I. Carman (1999). "Selecting medical students: a case report of the need for change." Med Educ 33(6): 455-459.

Marrin, M. L., K. A. McIntosh, D. Keane and M. L. Schmuck (2004). "Use of the Paired-Comparison Technique to Determine the Most Valued Qualities of the McMaster Medical Programme Admissions Process." Advances in Health Sciences Education 9(2): 129-135.

Mathers, J., A. Sitch and J. Parry (2016). "Longitudinal assessment of the impact of the use of the UK clinical aptitude test for medical student selection." Medical education 50(10): 1033-1044.

McAndrew, R., J. Ellis and R. A. Valentine (2017). "Does a selection interview predict year 1 performance in dental school?" European Journal of Dental Education 21(2): 108-112.

McAndrew, R. and R. Greatrix (2014). "The UKCAT test: developments, research and its use by dental schools in the UK." British dental journal 216(4): 191-194.

McGaghie, W. C. (1990). "Perspectives on medical school admission." Academic Medicine 65(3): 136-139.

McGaghie, W. C. (1990). "Qualitative variables in medical school admission." Academic Medicine 65(3): 145-149.

McGrath, E. J. (1945). "What Subjects Should Pre-dental Students Study?" The Journal of Higher Education 16(7): 351-357.

McKinley, A., M. Stevenson and K. Steele (2017). "Widening Participation To The Medical Course At Queens University Belfast." The Ulster medical journal 86(2): 119-122.

McLachlan, J. C., G. Finn and J. Macnaughton (2009). "The conscientiousness index: a novel tool to explore students' professionalism." Academic Medicine 84(5): 559-565.

McManus, I., C. Dewberry, S. Nicholson, J. Dowell, K. Woolf and H. Potts (2013). "Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: meta-regression of six UK longitudinal studies." BMC Med 11.

McManus, I., C. Dewberry, S. Nicholson and J. S. Dowell (2013). "The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools." BMC medicine 11(1): 244.

McManus, I., D. Powis, R. Wakeford, E. Ferguson, D. James and P. Richards (2005). "Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school." BMJ 331.

McManus, I. and P. Richards (1984). "Audit of admission to medical school: III--applicants' perceptions and proposals for change." BMJ 289(6455): 1365-1367.

McManus, I. and P. Richards (1986). "Prospective survey of performance of medical students during preclinical years." Br Med J (Clin Res Ed) 293(6539): 124-127.

McManus, I., P. Richards, B. Winder, K. Sproston and V. Styles (1995). "Medical school applicants from ethnic minority groups: identifying if and when they are disadvantaged." BMJ 310(6978): 496-500.

McManus, I., K. Woolf and J. Dacre (2008). "Even one star at A level could be “too little, too late” for medical student selection." BMC Med Educ 8.

McManus, I., K. Woolf, J. Dacre, E. Paice and C. Dewberry (2013). "The Academic Backbone: Longitudinal Continuities in Educational Achievement from Secondary School and Medical School to MRCP(UK) and the Specialist Register in UK Medical Students and Doctors." BMC Med 11.

McManus, I. C. (1982). "A-level grades and medical school admission." Br Med J (Clin Res Ed) 284(6329): 1654-1656.

McManus, I. C. and P. Richards (1984). "Audit of admission to medical school: I--Acceptances and rejects." Br Med J (Clin Res Ed) 289(6453): 1201-1204.

McManus, I. C. and P. Richards (1984). "Audit of admission to medical school: II--Shortlisting and interviews." Br Med J (Clin Res Ed) 289(6454): 1288-1290.

Medical Schools Council (2009). Medical students: professional values and fitness to practise.

Medical Schools Council (2010). Guiding Principles for the Admission of Medical Students.

Medical Schools Council (2013). Selecting for Excellence: End of Year Report – 2013. London, The Medical Schools Council.

Medical Schools Council (2014). Selecting for Excellence Final Report. Selecting for Excellence Group.

Milburn, A. (2009). "Unleashing aspiration: the final report of the panel on fair access to the professions."

Mills, A. J. and J. Helms Mills (2018). "Archival research." The Sage handbook of qualitative research methods in business and management. London: SAGE.

Mitchell, K. J. (1990). "Traditional predictors of performance in medical school." Academic Medicine 65(3): 149-158.

Monroe, A., E. Quinn, W. Samuelson, D. M. Dunleavy and K. W. Dowd (2013). "An Overview of the Medical School Admission Process and Use of Applicant Data in Decision Making: What Has Changed Since the 1980s?" Academic Medicine 88(5): 672-681.

Morgan, D. L. (2014). "Pragmatism as a paradigm for social research." Qualitative inquiry 20(8): 1045-1053.

Moss, F. (1930). "Scholastic Aptitude Tests for Medical Students." Academic Medicine 5(2): 90-110.

Mwandigha, L. M. (2017). Evaluating and extending statistical methods for estimating the construct-level predictive validity of selection tests, University of York.

Mwandigha, L. M., P. A. Tiffin, L. W. Paton, A. S. Kasim and J. R. Böhnke (2018). "What is the effect of secondary (high) schooling on subsequent medical school performance? A national, UK-based, cohort study." BMJ Open 8(5): e020291.

Nicholson, S. (2005). The benefits of aptitude testing for selecting medical students.

Nicholson, S. (2018). "Diversifying the medical workforce requires a step change in widening participation-Putting the cat amongst the pigeons!" MedEdPublish 7.

Onwuegbuzie, A. J. and N. L. Leech (2005). "On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies." International journal of social research methodology 8(5): 375-387.

Oosterveld, P. and O. ten Cate (2004). "Generalizability of a study sample assessment procedure for entrance selection for medical school." Med Teach 26(7): 635-639.

Papadakis, M. A., A. Teherani, M. A. Banach, T. R. Knettler, S. L. Rattner, D. T. Stern, J. J. Veloski and C. S. Hodgson (2005). "Disciplinary Action by Medical Boards and Prior Behavior in Medical School." New England Journal of Medicine 353(25): 2673-2682.

Parry, J., J. Mathers and A. Stevens (2006). "Admissions processes for five year medical courses at English schools: review." BMJ 332.

Paton, L. W., P. A. Tiffin, D. Smith, J. S. Dowell and L. M. Mwandigha (2018). "Predictors of fitness to practise declarations in UK medical undergraduates." BMC Medical Education 18(1).

Paton, P. A. T. L. W. (2017). Exploring the validity of the 2013 UKCAT SJT- prediction of undergraduate performance in the first year of medical school: Summary Version of Report.

Patterson, F., F. Cousans, H. Edwards, A. Rosselli, S. Nicholson and B. Wright (2017). "The Predictive Validity of a Text-Based Situational Judgment Test in Undergraduate Medical and Dental School Admissions." Academic Medicine.

Patterson, F., A. Knight, J. Dowell, S. Nicholson, F. Cousans and J. Cleland (2016). "How effective are selection methods in medical education? A systematic review." Med Educ 50.

Patterson, F., C. Roberts, M. D. Hanson, W. Hampe, K. Eva, G. Ponnamperuma, M. Magzoub, A. Tekian and J. Cleland (2018). "2018 Ottawa consensus statement: Selection and recruitment to the healthcare professions." Medical Teacher 40(11): 1091-1101.

Patterson, F., L. Zibarras and V. Ashworth (2016). "Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100." Medical Teacher 38(1): 3-17.

Pearson VUE (2019). UK Clinical Aptitiude Test (UKCAT) Consortium, UKCAT Examination Technical Report, Testing Interval: 1 July 2018 - 2 October 2018.

Pearson VUE (2020). University Clinical Aptitude Test (UCAT) Consortium Executive Summary 2019.

Peterson, S. (1946). "Who should enter dental school?" The Journal of the American Dental Association 33(1): 58-63.

Peterson, S. (1947). "Dental aptitude testing program: A report of progress." The Journal of the American Dental Association 35(3): 175-184.

Peterson, S. (1948). "Forecasting the success of freshman dental students through the aptitude testing program." The Journal of the American Dental Association 37(3): 259-265.

Peterson, S. (1951). "The dental aptitude testing program will become nation-wide for the 1951 entrants." Science Education 35(2): 119-121.

Powis, D., J. Hamilton and J. Gordon (2004). "Are graduate entry programmes the answer to recruiting and selecting tomorrow's doctors?" Medical education 38(11): 1147-1153.

Powis, D. A., R. L. B. Neame, T. Bristow and L. B. Murphy (1988). "The objective structured interview for medical student selection." British Medical Journal (Clinical research ed.) 296(6624): 765-768.

Powis, D. A., T. C. Waring, T. Bristow and D. L. O'Connell (1992). "The structured interview as a tool for predicting premature withdrawal from medical school." Aust N Z J Med 22(6): 692-698.

Ranney, R. R., M. B. Wilson and R. B. Bennett (2005). "Evaluation of applicants to predoctoral dental education programs: review of the literature." Journal of Dental Education 69(10): 1095-1106.

Razack, S., S. Faremo, F. Drolet, L. Snell, J. Wiseman and J. Pickering (2009). "Multiple mini-interviews versus traditional interviews: stakeholder acceptability comparison." Medical education 43(10): 993-1000.

Roberts, G. D. and A. M. Porter (1989). "Medical student selection--time for change: discussion paper." J R Soc Med 82(5): 288-291.

Roemer, R. (1965). "Nine year validity study of predictors of medical school success." The Journal of Educational Research 59(4): 183-185.

Rypins, H. (1931). "Paradoxical Trends." The Journal of Higher Education 2(8): 427-432.

Samuel, D. G. (2008). "UKCAT among the pigeons: UKCAT among the pigeons. Undermining access." BMJ (Clinical research ed.) 336(7648): 788.

Santana, I. R. and M. Chalkley "A sequential analysis of the specialty allocation process in the UK. Empirical evidence from the UKMED database."

Sartania, N., G. Haddock and M. Underwood (2018). "The challenges of widening access to the medical profession: how to facilitate medical careers for those at a genuine disadvantage." MedEdPublish 7.

Sartania, N., J. D. McClure, H. Sweeting and A. Browitt (2014). "Predictive power of UKCAT and other pre-admission measures for performance in a medical school in Glasgow: a cohort study." BMC medical education 14(1): 116.

Scheetz, J. P. (1987). "Predicting graduation from dental school using admissions data." Journal of dental education 51(5): 250-251.

Schwartz (2004). Fair admissions to higher education: recommendations for good practice. London: Department for Education and Skills.

Schwartzman, A. E., R. C. A. Hunter and J. G. Lohrenz (1962). "Factors Related to Medical School Achievement." Academic Medicine 37(8): 749-759.

Searle, J. and J. McHarg (2003). "Selection for medical school: just pick the right students and the rest is easy!" Medical education 37(5): 458-463.

Seccombe, W. (1932). "Survey of the Dental Curriculum." The Journal of Higher Education 3(5): 241-245.

Seyan, K., T. Greenhalgh and D. Dorling (2004). "The standardised admission ratio for measuring widening participation in medical schools: analysis of UK medical school admissions by ethnicity, socioeconomic status, and sex." BMJ 328(7455): 1545-1546.

Shaw, D. L., D. M. Martz, C. J. Lancaster and R. M. Sade (1995). "Influence of medical school applicants' demographic and cognitive characteristics on interviewers' ratings of noncognitive traits." Academic Medicine 70(6): 532-536.

Shoemaker, H. A. and J. H. Rohrer (1948). "Relationship Between Success in the Study of Medicine and Certain Psychological and Personal Data." Academic Medicine 23(3): 190-201.

Spratley, M. (1990). "Aptitude testing and the selection of dental students." Australian dental journal 35(2): 159-168.

Srikathirkamanathan K, C. S., MacManus BN (2018). "The relationship between UKCAT scores and finals exam performance for widening access and traditional entry students."

Stalnaker, J. M. (1950). "Medical College Admission Test." Academic Medicine 25(6): 428-434.

Stephenson, L. and T. Stephenson (2010). "Selecting tomorrow's doctors—not a level playing field." BMJ 341.

Steven, K., J. Dowell, C. Jackson and B. Guthrie (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): 11.

Stewart, F., J. Drummond, L. Carson and G. H. Reddick (2004). "The future of the profession—a survey of dental school applicants." British dental journal 197(9): 569-573.

Story, M. and A. Mercer (2005). Selection of medical students: an Australian perspective, Wiley Online Library.

Tavakol, M. and J. Sandars (2014). "Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part II." Medical Teacher 36(10): 838-848.

Thompson, C. E. (1944). "Personality and interest factors in dental school success." Educational and Psychological Measurement 4(1): 295-304.

Tiffin, P., L. Webster and J. McLachlan (2013). Forecasting the impact of the mode of use of the UKCAT on medical school entrant demographics.

Tiffin, P. A., J. S. Dowell and J. C. McLachlan (2012). "Widening access to UK medical education for under-represented socioeconomic groups: modelling the impact of the UKCAT in the 2009 cohort." BMJ 344.

Tiffin, P. A., J. C. McLachlan, L. Webster and S. Nicholson (2014). "Comparison of the sensitivity of the UKCAT and A Levels to sociodemographic characteristics: a national study." BMC medical education 14(1): 7.

Tiffin, P. A., L. M. Mwandigha, L. W. Paton, H. Hesselgreaves, J. C. McLachlan, G. M. Finn and A. S. Kasim (2016). "Predictive validity of the UKCAT for medical school undergraduate performance: a national prospective cohort study." BMC Medicine 14(1): 140.

Tiffin, P. A. and L. W. Paton (2019). "Does 'online confidence' predict application success and later academic performance in medical school? A UK-based national cohort study." BMJ Open 9(12): e034437.

Tiffin PA, P. L. (2017). "Exploring the validity of the 2013 UKCAT SJT- prediction of undergraduate performance in the first year of medical school: Summary Version Report."

Tinkelman, S. N. (1954). "Predicting academic success in New York state medical and dental schools." Academic Medicine 29(3): 35-41.

Travers, R. M. and W. L. Wallace (1950). "Inconsistency in the predictive value of a battery of tests." Journal of Applied Psychology 34(4): 237.

Turner, R. and S. Nicholson (2011). "Can the UK Clinical Aptitude Test (UKCAT) select suitable candidates for interview?" Medical education 45(10): 1041-1047.

Turner, R. and S. Nicholson (2011). "Reasons selectors give for accepting and rejecting medical applicants before interview." Medical Education 45(3): 298-307.

UKCAT Consortium (2006). UKCAT 2006 Annual Report.

UKCAT Consortium (2008). UKCAT: 2007 Annual Report.

United Kingdom Medical Applicant Cohort Study. (n.d.). "United Kingdom Medical Applicant Cohort Study." Retrieved 2 March, 2020, from https://ukmacs.wordpress.com/.

Webb, S. C. (1956). "The prediction of achievement for first year dental students." Educational and Psychological Measurement 16(4): 543-548.

Webster, E. S., L. W. Paton, P. ES Crampton and P. A. Tiffin (2020). "Situational judgement test validity for selection: a systematic review and meta-analysis." Medical Education.

Weinberg, E. and J. F. Rooney (1973). "The academic performance of women students in medical school." Academic Medicine 48(3): 240-247.

Weiss, I. (1952). "Prediction of academic success in dental school." Journal of Applied Psychology 36(1): 11.

Whittemore, R., S. K. Chase and C. L. Mandle (2001). "Validity in qualitative research." Qualitative health research 11(4): 522-537.

Woolf, K., H. Jayaweera, E. Unwin, K. Keshwani, C. Valerio and H. Potts (2019). "Effect of sex on specialty training application outcomes: a longitudinal administrative data study of UK medical graduates." BMJ open 9(3): e025004.

Wright, S. R. and P. M. Bradley (2010). "Has the UK Clinical Aptitude Test improved medical student selection?" Medical education 44(11): 1069-1076.

Yates, J. and D. James (2006). "Predicting the “strugglers”: a case-control study of students at Nottingham University Medical School." BMJ 332(7548): 1009-1013.

Yates, J. and D. James (2010). "The value of the UK Clinical Aptitude Test in predicting pre-clinical performance: a prospective cohort study at Nottingham Medical School." BMC medical education 10(1): 55.

Yates, J. and D. James (2013). "The UK clinical aptitude test and clinical course performance at Nottingham: a prospective cohort study." BMC medical education 13(1): 32.

Zimmermann, S., D. Klusmann and W. Hampe (2017). "Correcting the predictive validity of a selection test for the effect of indirect range restriction." BMC Medical Education 17(1): 246.

Appendices

Appendix A: UKCAT Consortium Membership

University of Aberdeen	2006 -
Brighton and Sussex Medical School	2006 – 2014 (left Consortium)
Cardiff University	2006 -
University of Dundee	2006 -
University of East Anglia	2006 -
University of Edinburgh	2006 -
University of Glasgow	2006 -
Hull York Medical School	2006 -
Keele University	2006 -
King's College London	2006 -
University of Leeds	2006 – 2014 (left Consortium)
University of Leicester	2006 -
University of Manchester	2006 –
Newcastle University	2006 -
University of Nottingham	2006 -
University of Oxford	2006 -
Peninsula Medical School	2006 – 2011 (to Plymouth and Exeter)
Queen Mary, University of London	2006 -
University of Southampton	2006 -
University of St Andrews	2006 -
St George's, University of London	2006 –
University of Aberdeen	2007 –
Imperial College London	2007 – 2014 (left Consortium)
Queens University Belfast	2007 –
University of Sheffield	2007 –
Warwick University	2007 –
University of Exeter	2012 -
University of Plymouth	2012 -
University of Central Lancashire	2014 (left Consortium)
University of Birmingham	2015 –
University of Liverpool	2015 -
University of Bristol	2016 –
Aston University	2017 –
Anglia Ruskin University	2018 –
Edge Hill University	2019 –
University of Sunderland	2019 –
Kent and Medway Medical School	2019 -
Brunel University	2021 -
Worcester University	2021 -
Chester University	2021 -

Appendix B: UKCAT Board Membership

Christopher Stevens	*December 2004 – April 2008*
David Gordon	*December 2005 – September 2006*
Katie Petty Saphon	*November 2005 -*
Mary Ann Lumsden	*November 2005 – January 2008*
Ian Johnson	*November 2005 – December 2008*
Sandra Nicholson	*November 2005*
Jane Adam	*November 2005 – December 2010*
Paul Dennis	*November 2005 – December 2009*
Martyn Annis	*November 2005 – December 2013*
Anthony Weetman	*December 2005 – December 2007, August 2009 – July 2013*
Malcolm Jones	*December 2005 – June 2007*
John Tooke	*July 2006 – August 2009*
Nigel Siesage	*December 2006 – December 2009, June 2011 -*
Barbara Chadwick	*June 2007 – November 2010*
David Yates	*December 2007 – December 2010*
Jon Dowell	*June 2008 – December 2014*
Anthony Warrens	*February 2009 – December 2011*
John McLachlan	*February 2009 – December 2017*
Lyndon Cabot	*February 2010 -*
Mark Thomason	*November 2010 – September 2012*
Martine Lowes	*December 2010 – December 2016*
Keith Steele	*December 2010 – December 2013*
Brigitte Scammell	*December 2011 – December 2014*
Iain Cameron	*August 2013 – July 2016*
Robert McAndrew	*September 2012 –*
Brian Pollard	*January 2014 – December 2016*
Adrian Husbands	*January 2014 – December 2016*
Paul Teulon	*January 2015 – December 2019*
Fiona Stewart	*January 2015 -*
Jenny Higham	*August 2016 – July 2019*
Gay Fagan	*January 2017 – December 2019*
Christine Kay	*January 2017 – December 2019*
Angela Kubacki	*January 2017 – December 2019*
Kim Piper	*January 2018 – February 2021*
Mandy Hampshire	*January 2018 –*
Ruth Valentine	*January 2018 – February 2021*
Enamul Ahsan	*January 2020 -*
Joanna Batt	*January 2020 -*
Victoria Mays	*January 2020 - February 2021*
Jayne Parry	*January 2020 -*
Emma Paton	*March 2021 -*

Natalie Cope	*March 2021 -*
Nana Sartania	*March 2021 -*
Ian Fussell	*August 2021 -*

Appendix C: UKCAT Test Content

VERBAL REASONING

The Verbal Reasoning subtest requires candidates to think about information presented in passages and to determine whether conclusions can be drawn from the information presented. Prior knowledge is not required to answer questions and indeed, passages of text are purposefully selected on the basis that it would be very unlikely for candidates to have come across the topic previously.

In 2021 this subtest was 22 minutes long and comprised 40 scored and 4 unscored items.

Between 2006 and 2009 (inclusive) only one item type was used in this subtest. Candidates were presented with passages of text, each associated with 4 items. Their task was to read each passage of text and decide whether the subsequent statement followed logically. Candidates had three answer options. An example of such an item can be found below.

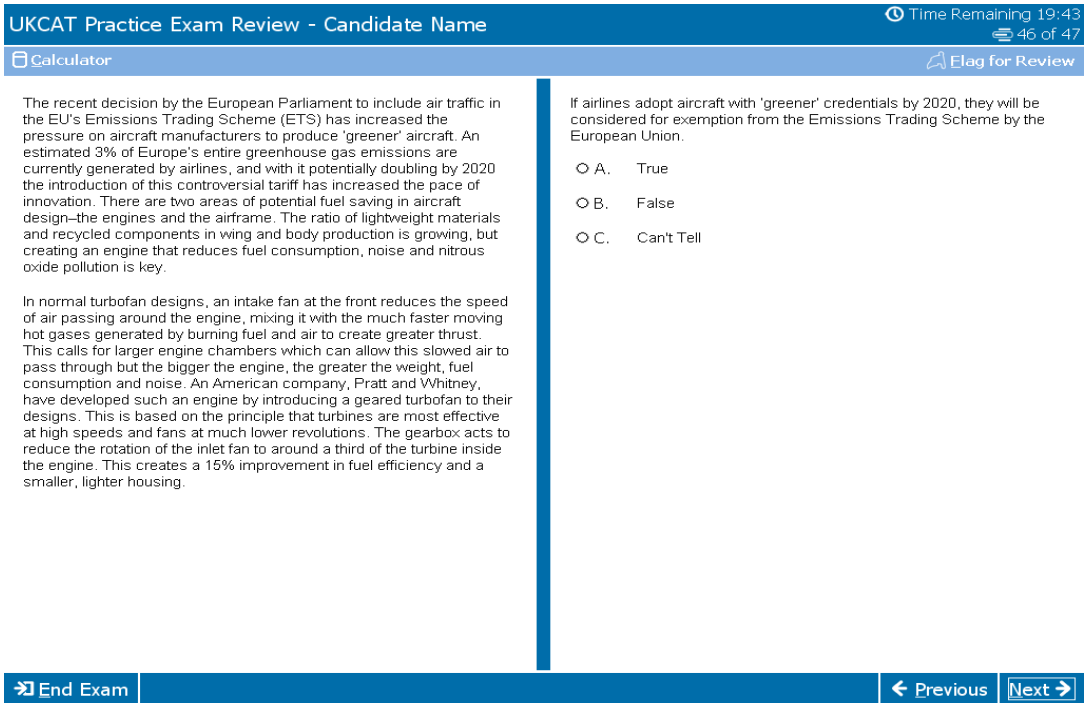


Figure 37 Verbal Reasoning Item 1 (Source: UKCAT)

In 2010 alternative verbal reasoning item types were trialled. The new items required candidates to read passages of text and respond to items such as 'If the above information is true, which of the following must also be true?' or 'Which of the following, if true, would weaken the argument presented in the text?' and 'Which of the following statements best supports the speaker's claims?'. It was anticipated that these items would better discriminate between candidates. An example of such an item can be found below.

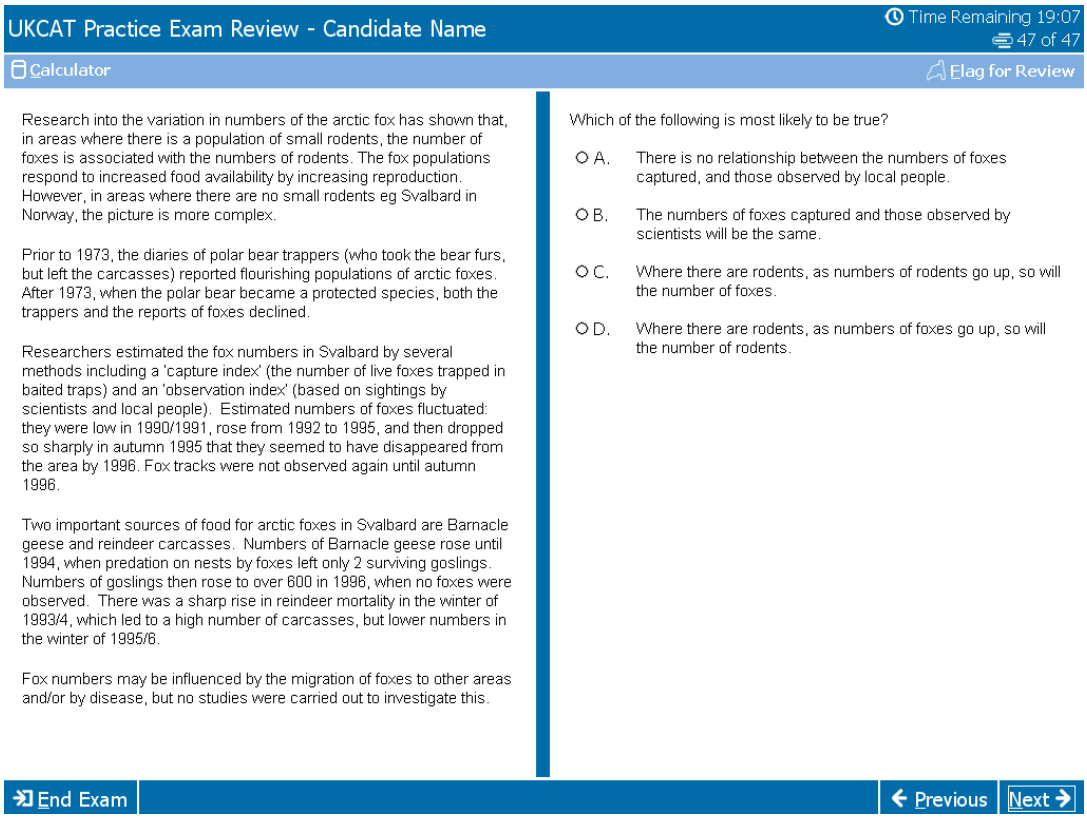


Figure 38 Verbal Reasoning Item 2 (Source: UKCAT)

Further pre-testing of these item types took place in 2011. The new items performed better than the existing true/false/can't tell items. In addition, 'close reading' items appeared to discriminate between candidates better than 'reasoning' items. These items were used operationally in 2012 with the balance continuing to increase in favour of such items in subsequent years.

QUANTITATIVE REASONING

The Quantitative Reasoning subtest intends to assess candidate ability to use numerical skills to solve problems. Given that little can be assumed regarding candidate qualifications beyond GSCE, this subtest only assumes mathematical ability to the standard of a good pass at GCSE. In fact, items are less to do with numerical facility and more to do with problem solving. Candidates were able to use handheld calculators in the early testing years. As a result of technical issues around their use, they were later replaced with a simple on-screen calculator embedded in the test itself.

Figure 39 UKCAT On-screen Calculator (Source: UKCAT)

In 2021 this subtest was 25 minutes long and comprised 3 scored and 4 unscored items.

Items presented in quantitative reasoning require candidates to solve problems by extracting relevant information from tables, graphs and other numerical presentations. Initially, for each table/graph candidates were presented with four items related to that image. An example of such an item is included below.

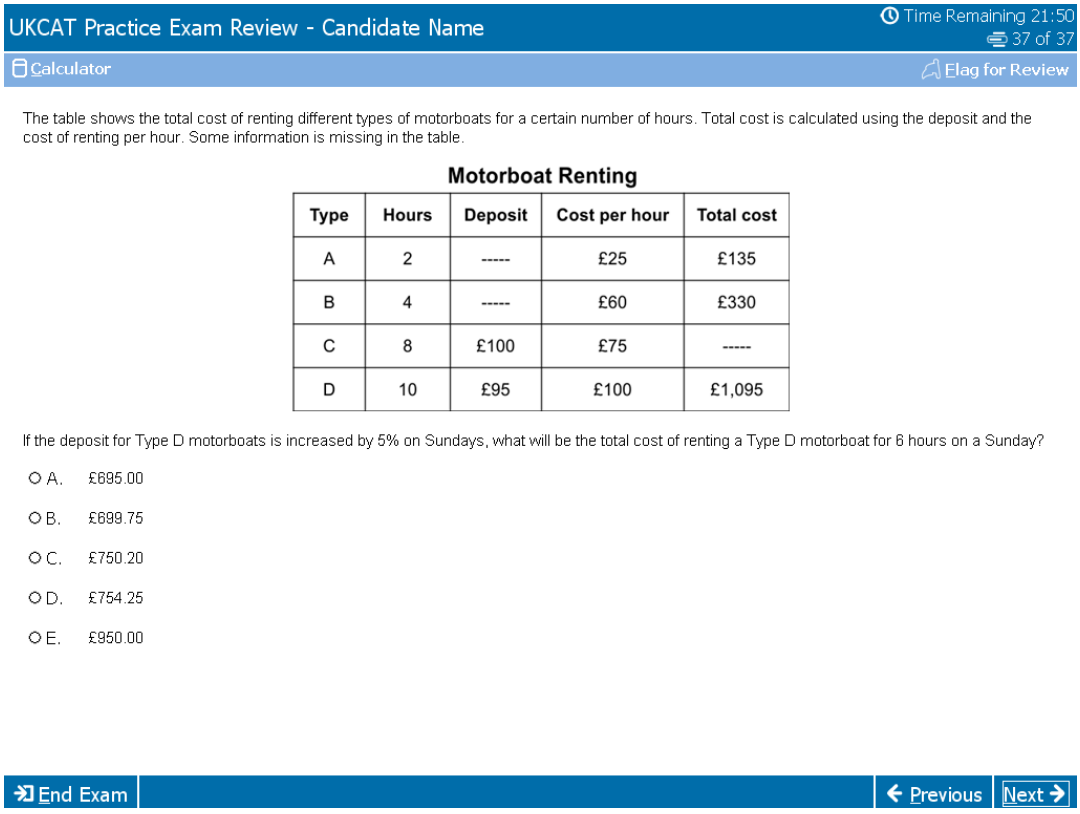


Figure 40 Quantitative Reasoning Item (Source: UKCAT)

The Quantitative Reasoning subtest had been repeatedly highlighted as the most *speeded* subtest. That is, this is the subtest that most candidates fail to finish. This has resulted in several adjustments over the years to item counts and subtest timings.

In 2012, stand-alone items were trialled for this subtest. These items performed similarly to existing 'testlets' and allowed for greater flexibility in trialling. The number of such items were increased in 2013.

Despite this being a speeded subtest, candidate mean average scores in this subtest have drifted upwards over time (from 597 in 2006 to 695 in 2017) resulting in quantitative reasoning impacting disproportionately on the total score. As a result, this subtest was rescaled in 2018 reducing the mean average that year to 658. Further rescaling of this subtest was put on hold in 2020 as it was deemed appropriate, given the challenges faced by COVID 19, not to introduce change to the test in that year.

ABSTRACT REASONING

Abstract Reasoning assesses candidates' ability to identify patterns amongst abstract shapes. Items are intended to be ambiguous leading some candidates to incorrect conclusions. The intention is that candidates query their judgements as they move through this subtest; different items require different approaches to arrive at the correct answer.

In 2021 this subtest was 13 minutes long and comprised 50 scored and 5 unscored items.

Originally, the Abstract Reasoning subtest comprised one item type with candidates presented with two sets of six shapes labelled "Set A" and "Set B". The shapes in Set A had a common connection as did the set of shapes in Set B. Therefore, initially candidates needed to establish what connected the shapes in Sets A and B. They were then required to determine whether a series of (five) test shapes belonged to Set A, Set B, or Neither. An example is provided here.

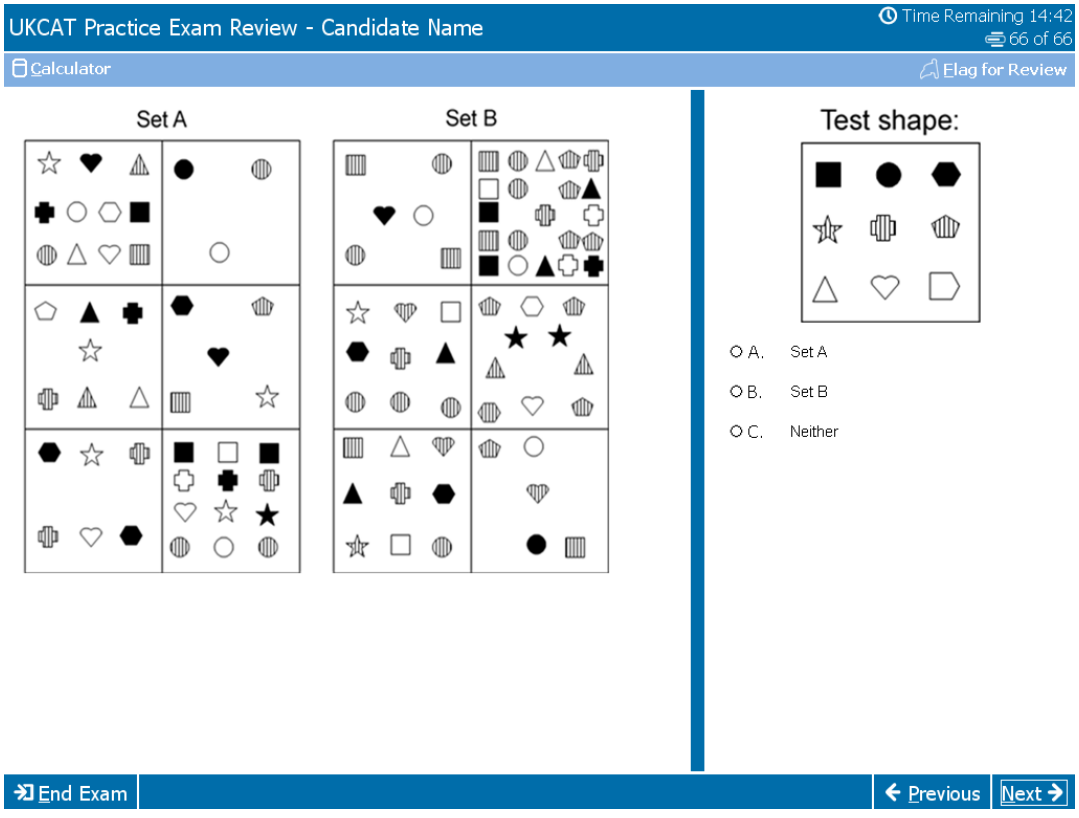


Figure 41 Abstract Reasoning Item 1 (Source, UKCAT)

Following a review of test content in 2012, it was agreed to trial new item types in abstract reasoning. Whilst the new items looked very similar to the original ones, they required candidates to take different approaches. Examples of the new items are included below. In one set of items, candidates were presented with a series of shapes and asked to select the next shape in the series.

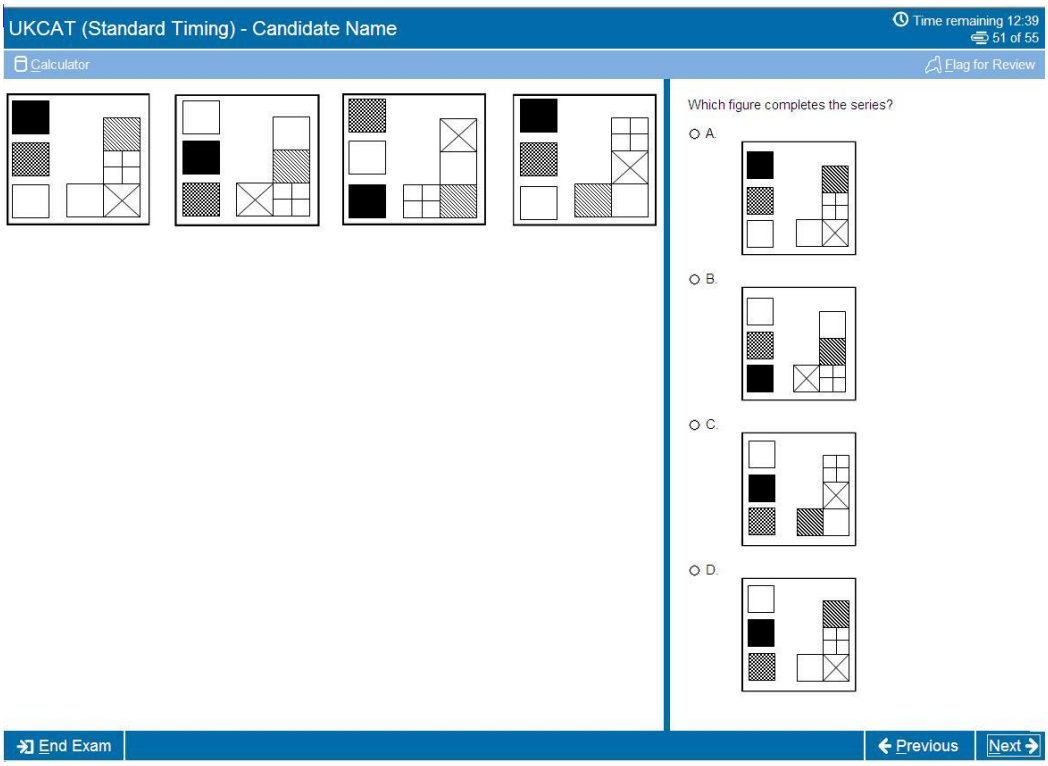


Figure 42 Abstract Reasoning Item 2 (Source, UKCAT)

A further item type presented candidates with a statement, involving a group of shapes, and candidates were asked to determine which shape completed the statement.

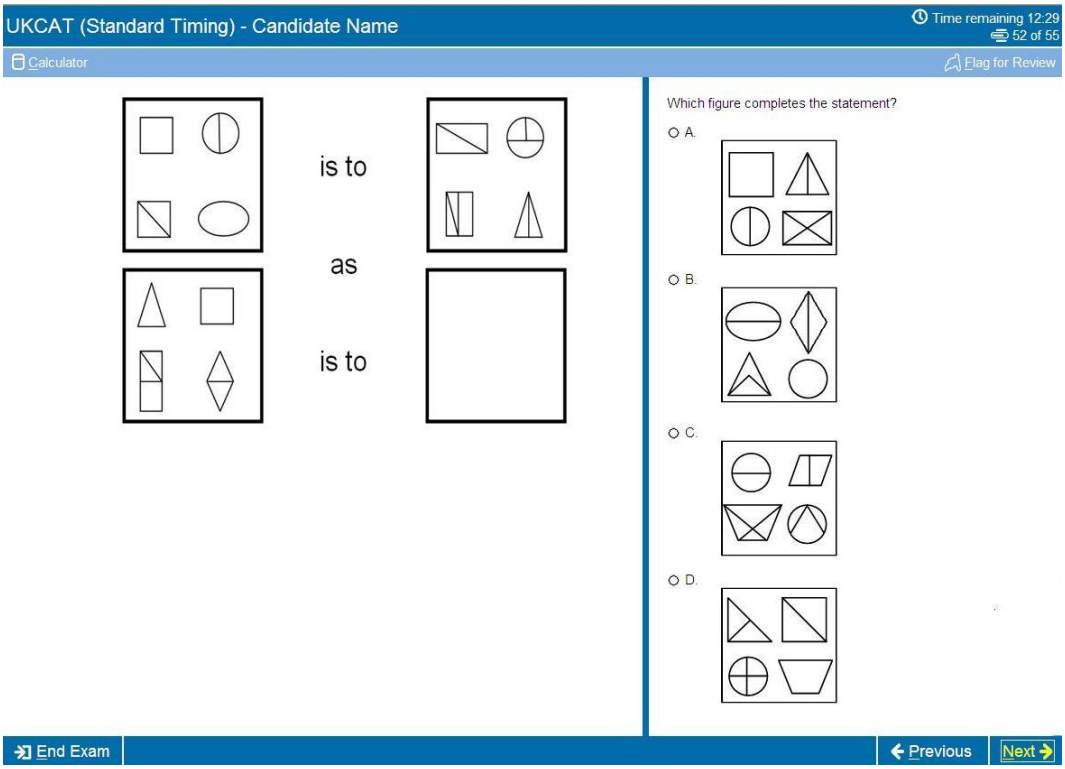


Figure 43 Abstract Reasoning Item 3 (Source: UKCAT)

In the final new item, candidates were presented with two sets of shapes labelled "Set A" and "Set B" and asked to select which of the four response options belonged to Set A or Set B.

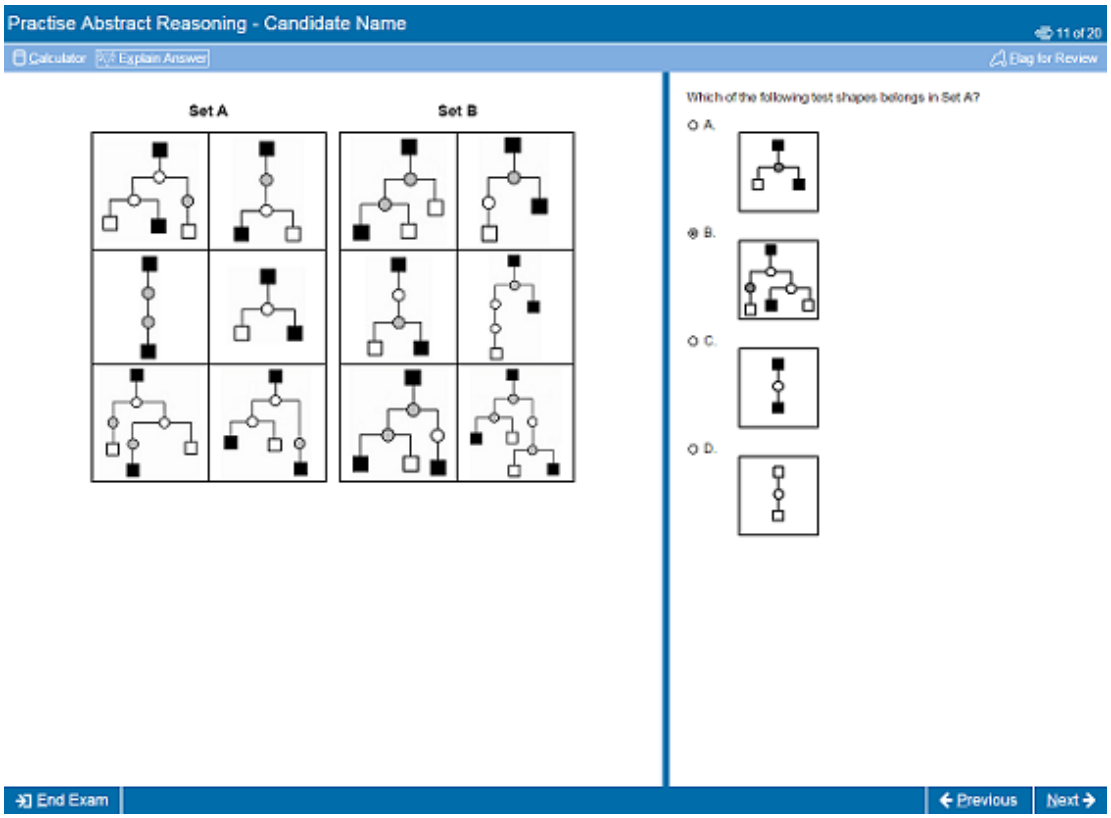


Figure 44 Abstract Reasoning Item 4 (Source: UKCAT)

Following trialling in 2013 and 2014, these new item types were introduced into the live test from 2015.

Scale scores in this subtest had drifted upwards over time (perhaps due to coaching effects mentioned below). As for quantitative reasoning rescaling of this subtest was again put on hold in 2020.

There has been some concern that this subtest is the most coachable of the subtests. Certainly, it would appear to be the subtest with most explicit advice given to candidates about how to approach items. In an unpublished study abstract reasoning is shown to be the subtest to show greatest shifts in score when retesting and to benefit most from both the use of commercial preparation materials and additional preparation time (Kulkarni S. 2020). It is also perhaps the subtests with least face validity. In 2021 the Research and Development Group asked Pearson VUE to explore alternative approaches to this subtest.

DECISION ANALYSIS

The Decision Analysis subtest was developed specifically for UKCAT with the intention of assessing candidate ability to make decisions in situations of uncertainty. It was removed from the test in 2015.

The subtest used a coding approach, requiring candidates to make informed judgements with information that was intentionally incomplete, complex, and ambiguous. The order of test items was fixed as candidates moved through the subtest with decisions intending to require increasing degrees of judgement.

Candidates were presented with a scenario which related to the discovery of a coded language. Most items required candidates to identify the best interpretation of a code. For some items, candidates were required to create codes for words or phrases.

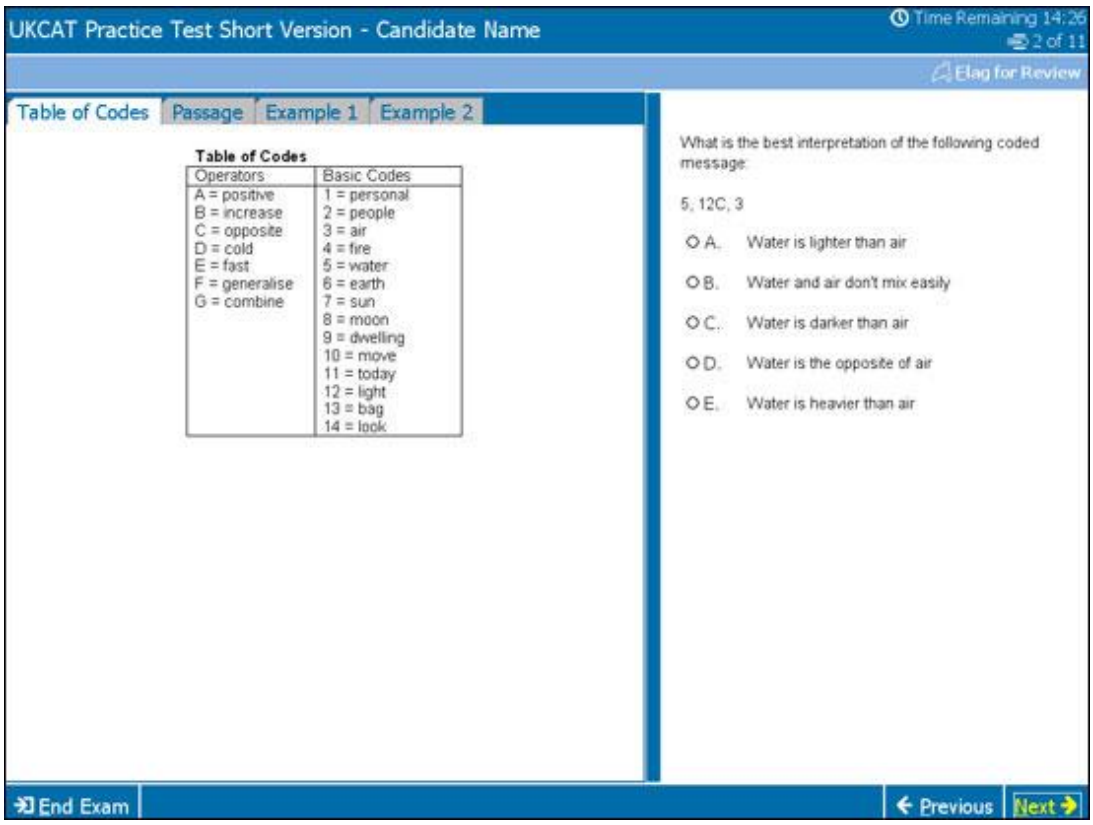


Figure 45 Decision Analysis Item 1 (Source: UKCAT)

The Decision Analysis subtest provoked most debate in Test Development meetings over the years with lower reliabilities in this subtest highlighted early on. The construct of the test provided limited opportunities to address this issue in the early years.

In May 2009, in response to the challenges posed by the subtest, Pearson VUE proposed that UKCAT consider an alternative test of critical thinking. It was decided not to explore this option, but to consider further how to improve the current subtest. One action was to increase the number of items in the subtest through trialling with the intention of improving item performance.

In 2013 it was agreed to trial the inclusion of confidence ratings in this subtest. Candidates were asked to provide information as to how confident they were that their response to an item was correct. Whilst it was originally intended that

candidates be given qualitative feedback on their answers to confidence ratings, this proved challenging, and it was agreed that candidates would not get feedback.

How these items were presented is illustrated below.

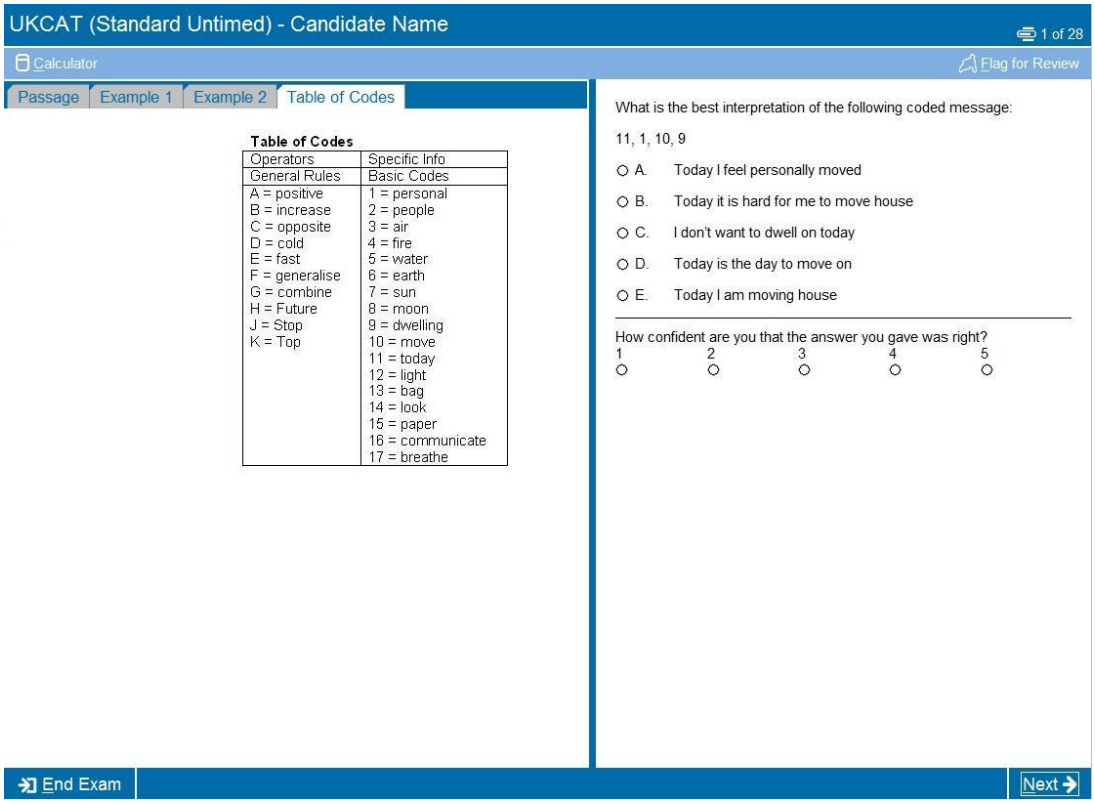


Figure 46 Decision Analysis Item 2 (Source: UKCAT)

Whilst initial analysis of confidence ratings showed reasonable distribution, and appeared to be measuring an independent trait, significant floor and ceiling effects were observed. Confidence bias appeared to be most useful when the difficulty of the test mirrored candidate ability (Tiffin and Paton 2019). A final review of the Confidence Rating data confirmed the existence of a ceiling and floor effect. Significant gender differences existed across confidence bias scores which would be difficult to take account of. It was agreed that confidence ratings would not be used further but that the UKCAT Research Group be invited to consider studies which could investigate/validate the use of this data further.

In 2013, one of the scenarios used in the subtest did not perform as expected requiring rescaling and rescoring for some candidates. Mean average scores for the subtest increased significantly (and unexpectedly) creating issue for universities, particularly those using threshold scores in selection. Reliabilities for this subtest remained much lower than for the other subtests. The challenge around improving this subtest was further impacted on by the fact that UKCAT did not own its intellectual property in the subtest, impacting on the cost of developing additional operational and practice items. Considering these issues, Pearson VUE were asked to investigate alternatives to the subtest. The subtest was removed in 2015 and a new Decision Making subtest trialled in 2016.

DECISION MAKING

In 2021 this subtest was 32 minutes long and comprised 26 scored and 3 unscored items.

A review of the literature had revealed a huge amount of information regarding testing of decision making and clinical decision making. UKCAT agreed to develop a bespoke measure of decision making utilising a variety of different item types, both to test a broader definition of decision making and to mitigate coaching effects. In reviewing potential content, Pearson VUE had focussed upon test constructs related to abilities identified in the Selecting for Excellence report (Medical Schools Council 2014) most relevant to decision making: problem solving, dealing with uncertainty and managing risk. As a result of this review, the recommended constructs for inclusion were deductive reasoning, evaluating arguments and statistical & figural reasoning.

It was agreed to utilise multiple approaches within the subtest to reduce coaching effects. Item types included syllogisms, logical puzzles, recognising assumptions, interpreting information and drawing conclusions, venn diagrams and probabilistic & statistical reasoning. The intention was to investigate overlap between item types over time. Drag-and-drop items were introduced resulting in a more innovative look to the subtest.

Example items are included below.

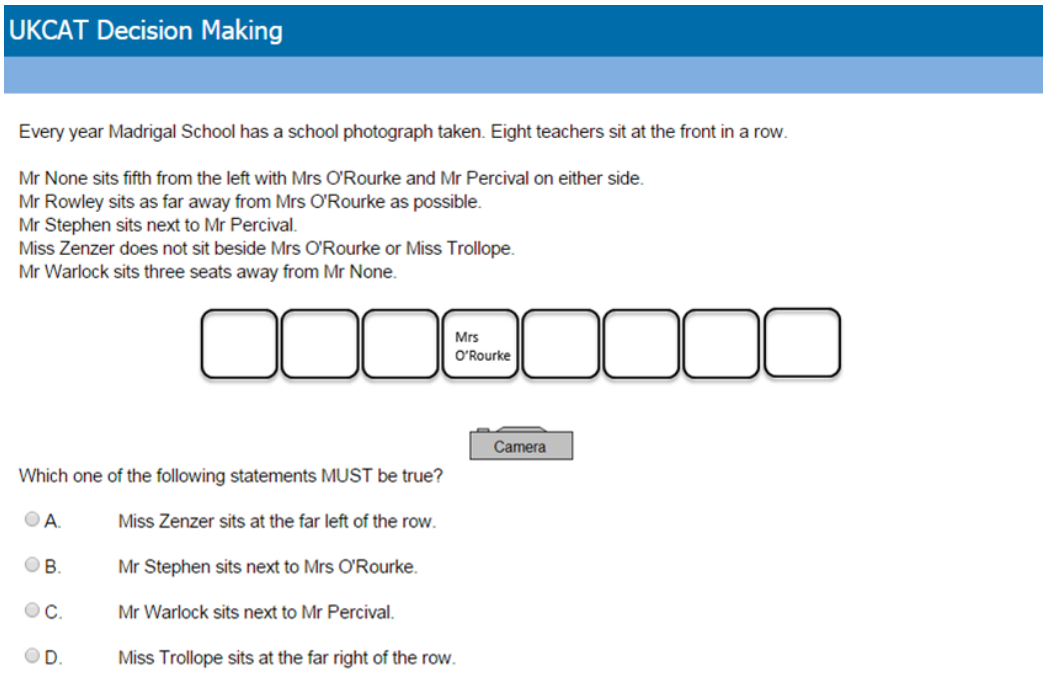


Figure 47 Deductive Reasoning Logic Item (Source: UKCAT)

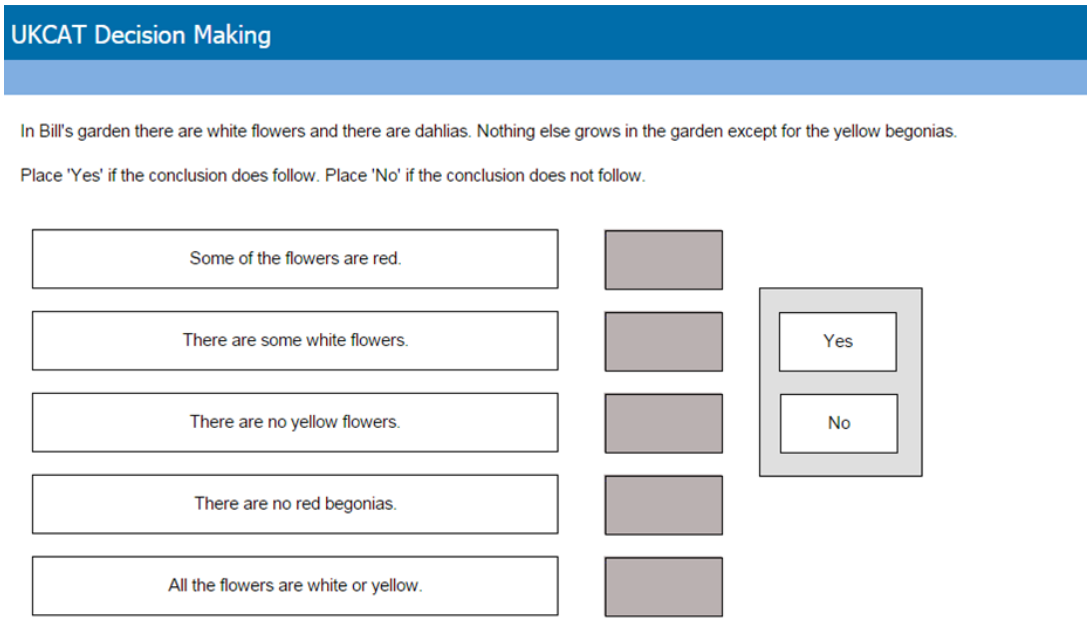


Figure 48 Deductive Reasoning Syllogism Item (Source: UKCAT)

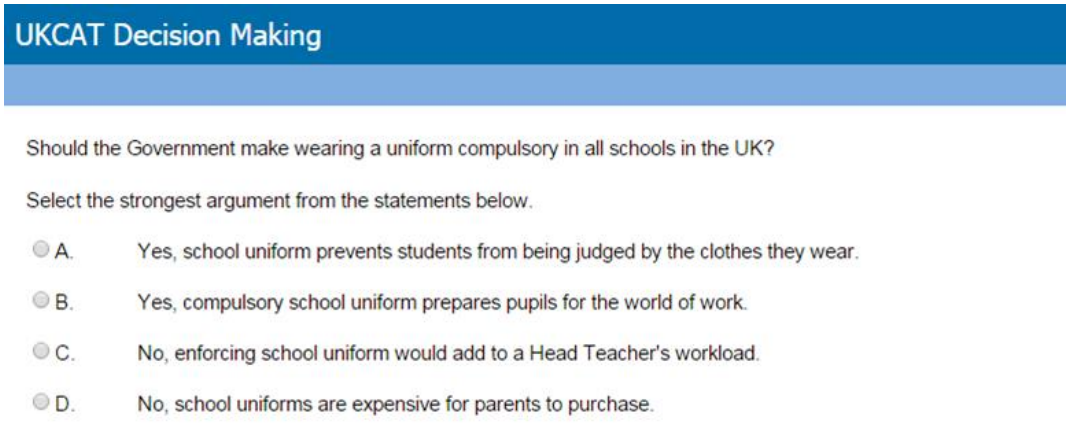


Figure 49 Evaluating Arguments Item (Source: UKCAT)

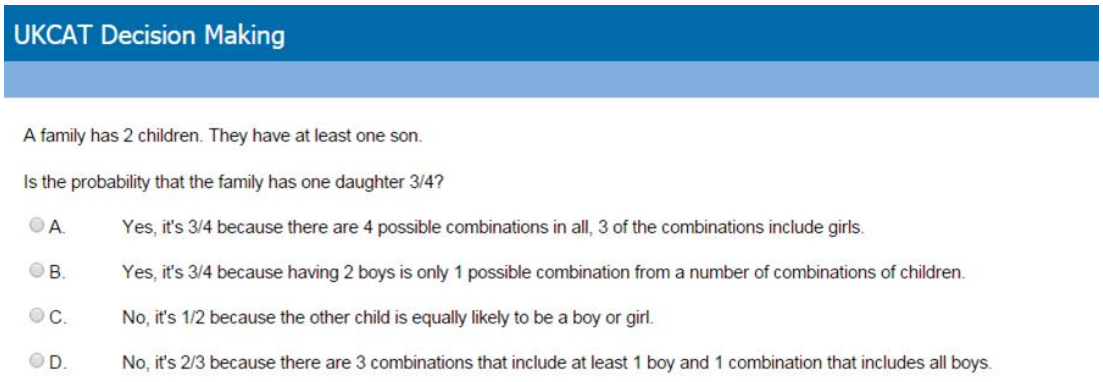


Figure 50 Probabilistic Reasoning Item (Source: UKCAT)

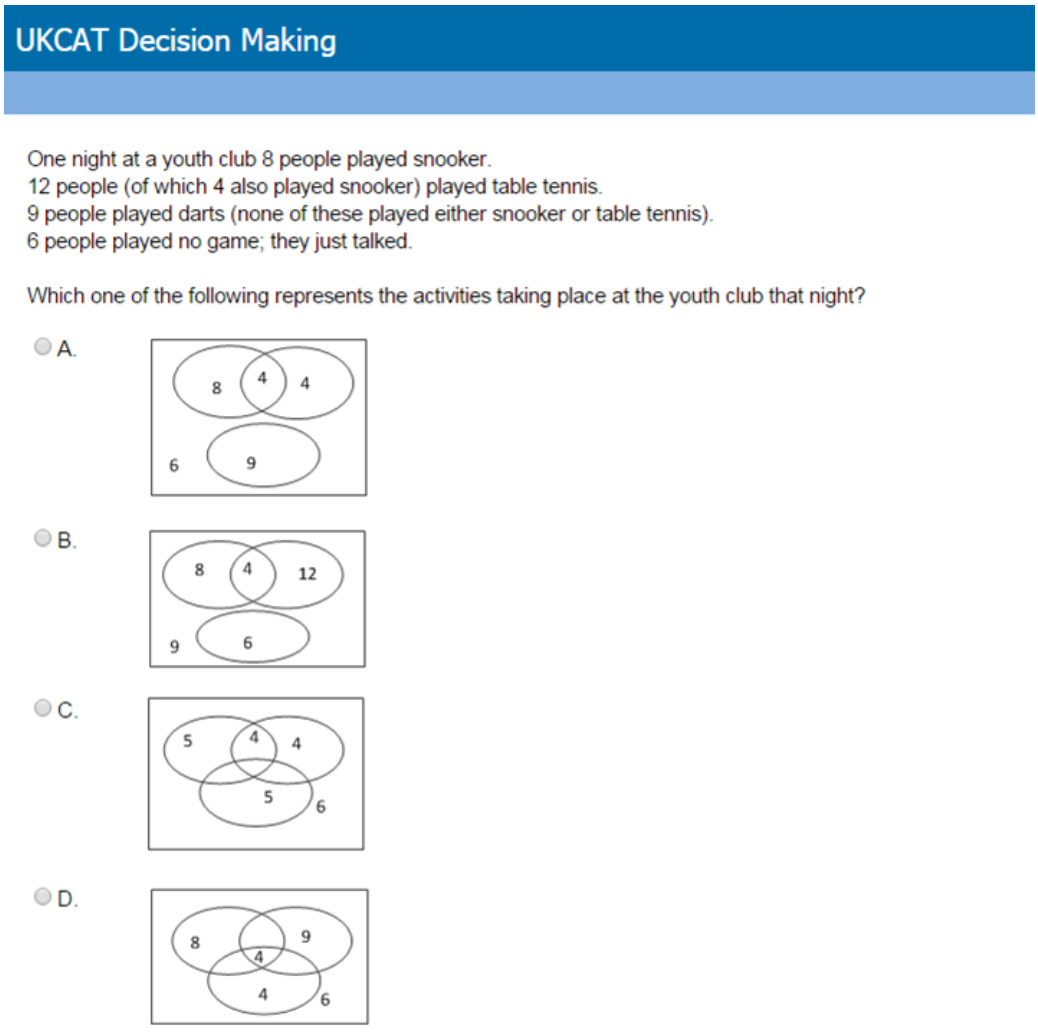


Figure 51 Venn Diagram Item (Source: UKCAT)

The subtest was trialled in 2016. To mitigate concerns that candidates would not take the test seriously, previously calibrated quantitative reasoning items were placed within the "Interpreting information and drawing conclusions" section,

allowing analysis to take place regarding item performance against candidate ability. This approach provided a better understanding of potential anticipated shifts in scores in 2017 (when the test was operationalised). Items generally performed well and there was no need to make significant changes to the subtest in 2017 when it became operational.

CREATING AND DEVELOPING SUBTEST 5 (THE BEHAVIOURAL TEST)

The Early Days

From its inception, UKCAT committed to developing a test which assessed (alongside cognitive ability) the traits required to make good doctors and dentists. This section is referred to in documents initially as Paper B, Section 5, the non-cognitive test, or the behavioural test.

In 2007 it was agreed to trial items intended to assess non-cognitive traits of empathy, integrity and robustness and three different instruments were piloted with candidates:

- MEARS (Managing Emotions and Resilience Scales)
- ITQ100 (Interpersonal Traits Questionnaire)/NACE (Narcissism, Aloofness, Confidence and Empathy)
- IVQ49 (Interpersonal Values Questionnaire)/MOJAC (a measure of ethical orientation).

In addition, abridged versions of ITQ (ITQ50) and IVQ (IVQ33) were combined and piloted.

Candidates did not receive results from this subtest nor were they communicated to universities.

In 2010, the number of versions of the Section 5 tests was reduced to 3 (combined ITQ/ITV, MEARS and SA12) with the intention of moving towards the development of a single non-cognitive test in 2011. Proposals for a combined test were considered following 2010 testing. UKCAT remained concerned about how such a test might be used operationally in the absence of research evidence. It was agreed therefore to withdraw this subtest in 2011 and evaluate existing data.

Subsequent analysis of the data concluded the tests had limited potential to predict performance in undergraduate assessments, also noting the difficulties in using self-report tools in high stakes testing (Finn, Mwandigha et al. 2018).

The UKCAT Situational Judgement Test

Included within the tender process for the award of the new test delivery and development contract (to run from 2012) was a requirement for tenderers to include proposals for a test of non-cognitive traits to be developed within the test. Pearson VUE were successful in securing the contract and proposed working with the Work Psychology Group (https://www.workpsychologygroup.com/) to develop the UKCAT Situational Judgement Test (SJT).

In 2021 this subtest was 27 minutes long and comprised 63 scored and 6 unscored items.

Role analysis took place using a literature review, semi-structured stakeholder interviews and a validation questionnaire. The traits identified for inclusion in the 2021 test were integrity, perspective taking (empathy) and team involvement.

The test was trialled with all candidates in 2012 and became operational in 2013. Rather than using scale scores, candidates scores were allocated into 4 bands with Band 1 being the highest performing candidates. Candidates received their results by band together with a feedback statement as follows:

Band 1	**Those in Band 1 demonstrated an excellent level of performance, showing similar judgement in most cases to the panel of experts.**
Band 2	Those in Band 2 demonstrated a good, solid level of performance, showing appropriate judgement frequently, with many responses matching model answers.
Band 3	Those in Band 3 demonstrated a modest level of performance, with appropriate judgement shown for some questions and substantial differences from ideal responses for others.
Band 4	The performance of those in Band 4 was low, with judgement tending to differ substantially from ideal responses in many cases.

Candidates were presented with a scenario and asked to comment on either the appropriateness of a response from an individual in that scenario, or on how important certain considerations were in deciding how to respond. Example items are below.

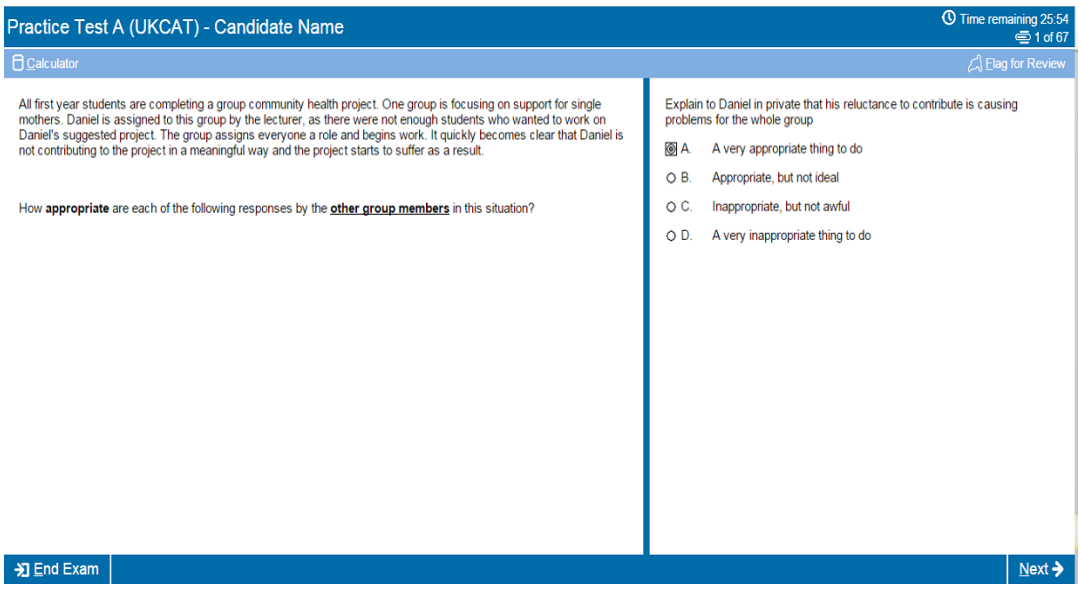


Figure 52 SJT Appropriateness Item (Source: UKCAT)

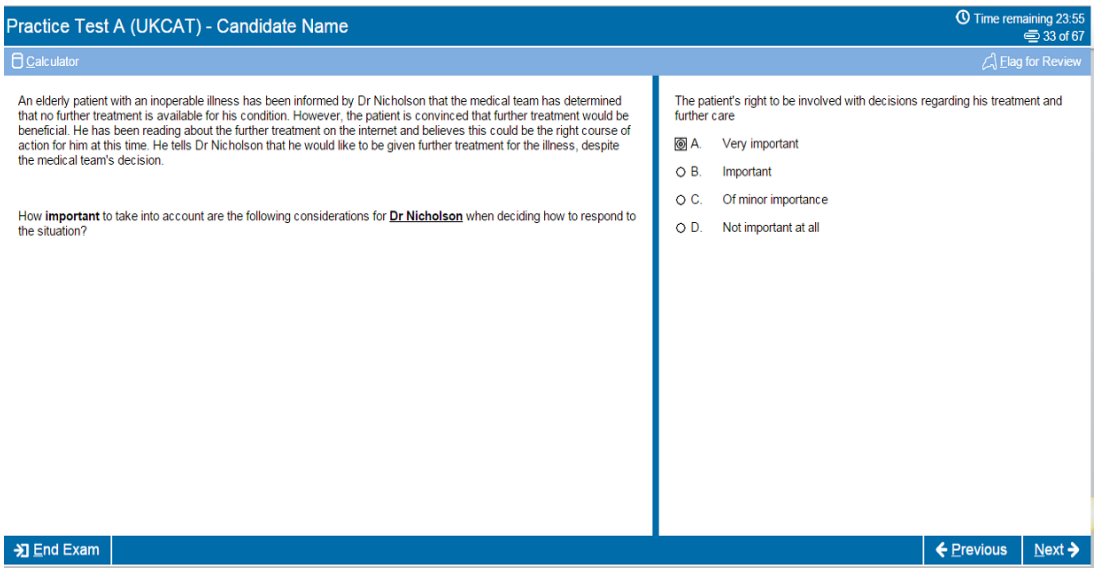


Figure 53 SJT Importance Item (Source: UKCAT)

In 2015, items including the additional traits of resilience and adaptability were trialled within the test. These were incorporated into the operational test in 2016.

At a Consortium in July 2015, the Work Psychology Group presented initial outcomes from a validation study which had compared candidates' SJT scores with tutor ratings collected using a bespoke questionnaire. This work was later published (Patterson, Cousans et al. 2017). The study provided positive early evidence of the predictive validity of the SJT. The authors concluded that the findings supported the use of this subtest as a selection tool for screening out lower performing applicants, when used alongside other measures. This approach was further supported the following year in a presentation of descriptive statistics demonstrating how different responses to the subtest from band 4 candidates were to the overall candidate population.

It was agreed to trial some 'Pick 1 of 3' items in 2018 alongside other new items using a direct speech approach. Example items are provided below.

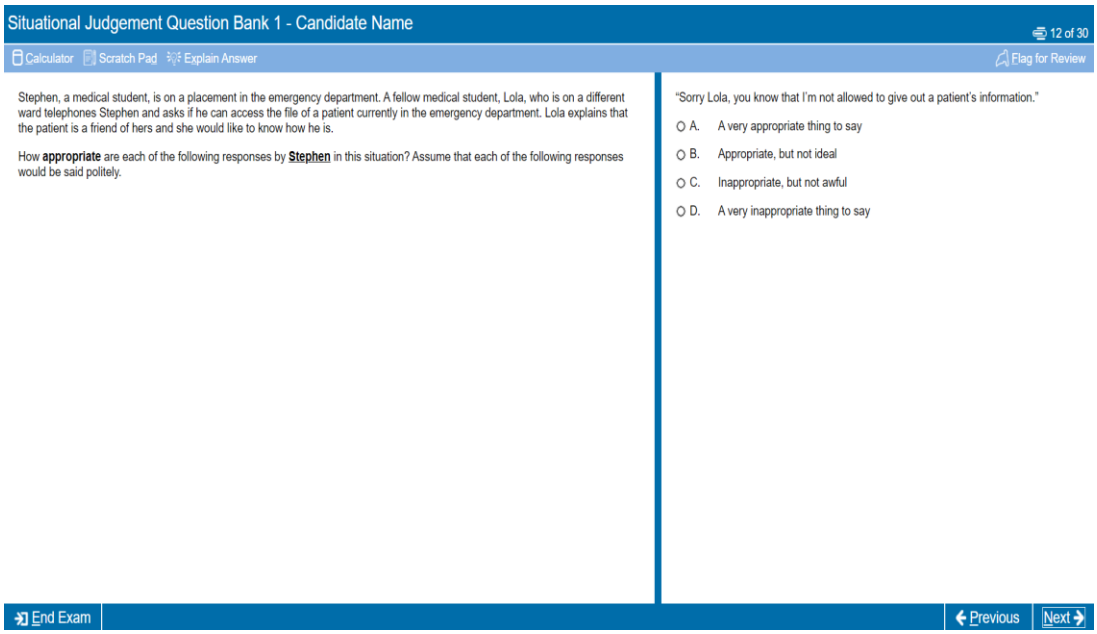


Figure 54 Direct Speech SJT Item (Source: UKCAT)

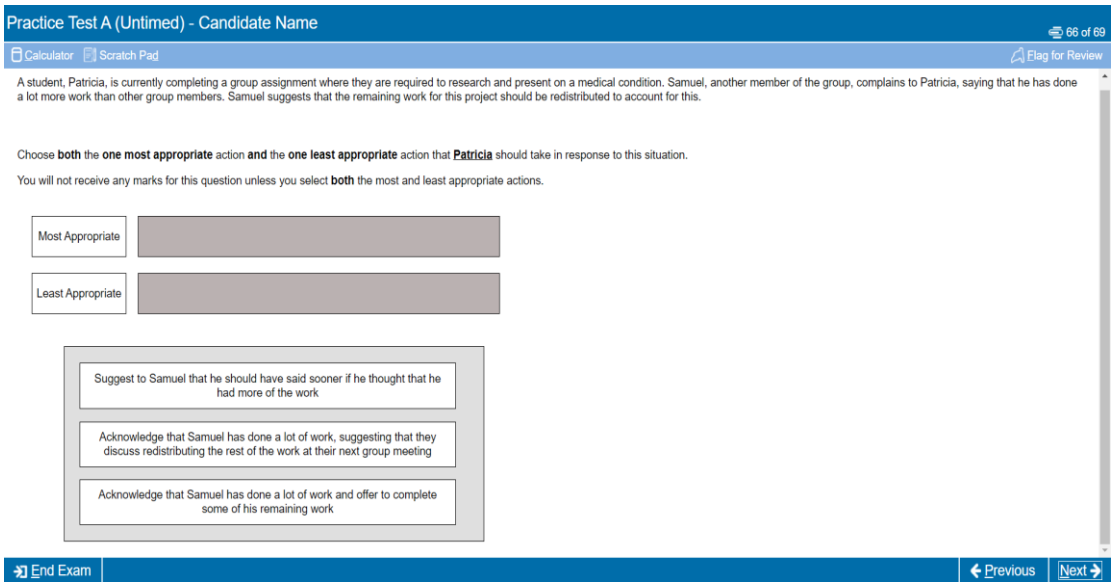


Figure 55 Drag and Drop SJT Item (Source: UKCAT)

Appendix D: PRISMA-P 2015 Checklist

This checklist has been adapted for use with protocol submissions to *Systematic Reviews* from Table 3 in Moher D et al: Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 2015 4:1

Section/topic	#	Checklist item	Information reported		Line number(s)
			Yes	No	
ADMINISTRATIVE INFORMATION					
Title					
Identification	1a	Identify the report as a protocol of a systematic review	☐		22
Update	1b	If the protocol is for an update of a previous systematic review, identify as such		☐	
Registration	2	If registered, provide the name of the registry (e.g., PROSPERO) and registration number in the Abstract		☐	
Authors					
Contact	3a	Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author	☐		4,5,8,9
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	☐	☐	459-465
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments		☐	
Support					
Sources	5a	Indicate sources of financial or other support for the review	☐		466
Sponsor	5b	Provide name for the review funder and/or sponsor	☐		466
Role of sponsor/funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	☐		466
INTRODUCTION					

Section/topic	#	Checklist item	Information reported		Line number(s)
			Yes	No	
Rationale	6	Describe the rationale for the review in the context of what is already known	☐		49-106
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)		☐	N/A
METHODS					
Eligibility criteria	8	Specify the study characteristics (e.g., PICO, study design, setting, time frame) and report characteristics (e.g., years considered, language, publication status) to be used as criteria for eligibility for the review	☐		110-134
Information sources	9	Describe all intended information sources (e.g., electronic databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage	☐		110-134
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	☐		111-116
STUDY RECORDS					
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	☐		144-147
Selection process	11b	State the process that will be used for selecting studies (e.g., two independent reviewers) through each phase of the review (i.e., screening, eligibility, and inclusion in meta-analysis)	☐		136, 459
Data collection process	11c	Describe planned method of extracting data from reports (e.g., piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	☐		144
Data items	12	List and define all variables for which data will be sought (e.g., PICO items, funding sources), any pre-planned data assumptions and simplifications	☐		157-189
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	☐		157-189
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis		☐	N/A

Section/topic	#	Checklist item	Information reported		Line number(s)
			Yes	No	
DATA					
Synthesis	15a	Describe criteria under which study data will be quantitatively synthesized		☐	N/A
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data, and methods of combining data from studies, including any planned exploration of consistency (e.g., I^2, Kendall's tau)		☐	N/A
	15c	Describe any proposed additional analyses (e.g., sensitivity or subgroup analyses, meta-regression)		☐	N/A
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	☐		240-244
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (e.g., publication bias across studies, selective reporting within studies)		☐	N/A
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (e.g., GRADE)		☐	N/A

Appendix E: Survey of Medical Schools' use of the UKCAT (2018)

Medical School	
Name of respondent	
Role of respondent	
Telephone number	
Email	
Date and time of interview	
Interview duration	
Draft document sent for checking	
Final document received	

Basic statistics	**Response in 2017**	**Response in 2018**
Number of places (home)		
Number of places (overseas)		
Number of applications		
Number of interviews		
Number of offers		
Standard offer		

	Question	**Response in 2017**	**Response in 2018**
1	Please describe your selection process this year. It may be useful to use the following headings:		
	Receipt of form		
	Screening academic credentials		
	Reading PS and Ref		
	Scoring		
	Invitation for interview		
	Format of interview		

	Question	Response in 2017	Response in 2018
	Outcome of interview		
	Recommendation for offer		
	Making of offer		
	Offer level		
2	How did you use the UKCAT result in your selection process?		
2a	Was this used in pre-selection?		
	Was this used in selection for interview?		
	Was this used after interview?		
2b	Did you use it for assessment or selection of any specific subgroups? (e.g. borderline, WP, mature or disabled). Please specify and describe.		
3	Did you make any other changes to your selection process this year?		
4a	Do you think using the UKCAT has affected the profile of the candidates you have selected this year, compared with previous years? If so, how?		
4b	Do you have any evidence to support this?		
5a	Have you analysed the effect of using the UKCAT on the profile of your selected candidates? If so, how?		
5b	Do you intend to analyse effect of using the UKCAT on the profile of your selected candidates? If so, how?		
6	Are you anticipating any significant changes to your		

	Question	Response in 2017	Response in 2018
	admission processes next year including the use of the UKCAT? Please comment on whether you will be using the SJT results within your admission processes this year.		

Appendix F: Research Ethics Consent Annual Medicine Interviews

University of Nottingham
UK | CHINA | MALAYSIA

Faculty of Medicine & Health Sciences
Research Ethics Committee
Faculty Hub
Room E41, E Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham, NG7 2UH
Email: FMHS-ResearchEthics@nottingham.ac.uk

11 July 2019

Rachel Greatrix
UCAT Chief Operating Officer
c/o Dean's Office
B Floor, Medical School
School of Medicine
QMC Campus
Nottingham University Hospitals
Nottingham
NG7 2UH

Dear Mrs Greatrix

Ethics Reference No: 325-1905 – **please always quote**
Study Title: UKCAT and medical student selection in the UK - what has changed since 2006?

Thank you for your e-mail query dated 9 July 2019. Implied consent in this instance would be appropriate because it is information that is routinely collected as part of the service provided. In this case the participant is not the subject of data being collected for research but rather relaying routine information required on behalf of the organisation who are partners in the UK wide service. The routine data provided informed for the evaluation of the UKCAT service.

Yours sincerely

Professor Ravi Mahajan
Chair, Faculty of Medicine & Health Sciences Research Ethics Committee

Appendix G: Research Ethics Consent Annual Dentistry Interviews

Faculty of Medicine & Health Sciences
Research Ethics Committee
Faculty Hub
Room E41, E Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham, NG7 2UH
Email: FMHS-ResearchEthics@nottingham.ac.uk

11 March 2021

Rachel Greatrix
UKCAT Chief Operating Officer
c/o Dean's Office
B Floor, Medical School
School of Medicine
QMC Campus
Nottingham University Hospitals
Nottingham
NG7 2UH

Dear Mrs Greatrix

Ethics Reference No: 204-0221 – please always quote
Study Title: UKCAT and Dental student selection in the UK - what has changed?

Thank you for your e-mail query dated 26 February 2021. Implied consent in this instance would be appropriate because it is information that is routinely collected in this way as part of the business as usual service provided. In this case the participant is not the subject of the interview or data being collected for research but rather they are reporting routine information required on behalf of the organisation who are partners in the UK wide service. The routine data provided informed for the evaluation of the UKCAT service.

Yours sincerely

pp

Dr John Williams, Associate Professor in Anaesthesia and Pain Medicine
Chair, Faculty of Medicine & Health Sciences Research Ethics Committee

Appendix H: Medical School Use of the UKCAT to select applicants for interview over time

	Number of Medical Schools												
Year	**No use**	**Borderline**	**Factor**	**Factor (inc SJT)**	**Factor and SJT Threshold**	**Threshold (actual and convenience)**	**Threshold (actual)**	**Threshold (actual) and factor**	**Threshold (actual) and SJT**	**Threshold (convenience)**	**Threshold (convenience) and SJT**	**Trade Off**	**Total**
2007	12	4	5				1			1			23
2008	6	4	8				3			4		1	26
2009	6	3	8				1			7		1	26
2010	6	2	7					1		8		1	25
2011	6	1	7				1	1		8		1	25
2012	4	1	8				3	1		7		1	25
2013	3	1	9				2	2		8			25
2014	1	2	7		2	1	3	2		7			25
2015	1	1	7			1	3	3		7			23
2016	2		7	1	1		3	2		8	1		25
2017	2		8	1	1	1	1	2	1	7	1		25
2018	3		7	1	2	1	1	2	1	7	1		26

Appendix I: Threshold and Factor Methods over time (Medicine)

1 Mean Average Threshold Scores and UKCAT Scores over time (N = number of medical schools)

	Actual Threshold		Convenience Threshold		All Candidates
Year	**Average Score**	**N**	**Average Score**	**N**	**Average Score**
2007	2350	1	2790	1	2375
2008	2403	4	2115	3	2401
2009	2600	1	2491	7	2430
2010	2250	2	2599	8	2505
2011	2319	4	2659	7	2489
2012	2397	6	2625	6	2476
2013	2511	6	2614	7	2516
2014	2591	6	2805	7	2643
2015	2468	6	2467	8	2505
2016	2369	5	2619	9	2531
2017	2449	4	2586	9	2523
2018	2463	4	2544	9	2540

2 Mean Average Weightings of Factors used in selection for interview over time (N = number of medical schools)

	UKCAT		Academic Score		Personal Statement and Reference		SJT	
Year	**N**	**Average Weighting %**	**N**	**Average Weighting %**	**N**	**Average Weighting %**	**N**	**Average Weighting %**
2007	4	26.11	9	29.38	9	58.02		
2008	7	22.48	9	38.56	9	58.04		
2009	7	20.88	10	35.03	9	55.74		
2010	7	20.88	9	36.43	8	51.03		
2011	7	23.47	9	40.58	8	45.70		
2012	8	31.05	10	44.35	7	43.39		
2013	10	37.58	10	44.85	5	33.23		
2014	10	40.50	9	53.23	4	31.46		
2015	9	43.70	9	52.81	1	23.53		
2016	10	42.40	10	53.72	1	22.54	2	9.61
2017	11	37.89	11	50.14	2	47.90	2	9.92
2018	11	39.39	11	50.81	2	36.11	2	10.28

3 Mean Average Weightings of Factors used in selection for offer over time

	Interview Score		UKCAT		Academic		Personal Statement and Reference		SJT	
Year	**N**	**Average Weighting %**	**N**	**Average Weighting %**	**N**	**Average Weighting %**	**N**	**Average Weighting %**	**N**	**Average Weighting %**
2007	23	91.12	2	8.36	3	44.33	3	33.17		
2008	24	89.82	3	8.91	3	40.64	3	33.17		
2009	24	89.82	3	8.91	3	40.64	3	31.31		
2010	23	91.83	4	9.43	2	49.92	2	29.87		
2011	23	91.38	4	20.78	2	46.62	2	31.82		
2012	23	89.44	4	20.55	2	46.62	3	36.37		
2013	23	89.64	4	24.53	2	46.62	3	29.55		
2014	24	90.18	5	23.93	2	50.00	3	15.00	1	15.00
2015	22	95.45	2	35.00	1	50.00	1	15.00	1	15.00
2016	24	94.14	3	28.10	1	50.00	1	15.00	3	13.75
2017	24	93.68	3	28.10	1	50.00	1	15.00	4	13.09
2018	25	90.93	4	27.32	2	50.00	1	15.00	4	13.09

Appendix J: Expressions of Interest Information Sheet

Admission Tutor Views of the Impact of the UKCAT on Medical and Dental Selection

Rachel Greatrix, School of Medicine, University of Nottingham

Rachel.greatrix@nottingham.ac.uk

0115 8230041

Invitation

You are being invited to take part in the above research project. Before you decide whether you would like to take part, it is important you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask me if there is anything that is not clear or if you would like more information. Thank you for reading this.

What is the project's purpose?

The aim of this project is to better understand how Admission Tutors view the impact UKCAT has had on selection locally and why this has occurred. To this end, interviews will explore:

- How and to what extent the UKCAT impacts on selection locally;
- How usage has changed (if at all) and what the drivers have been around this change;
- What the challenges are to using UKCAT in selection;
- How admission tutors would like UKCAT to develop to better support selection.

The project forms one part of my PhD which is looking at the broader impact of the UKCAT, drawing on information from the literature, archived materials and my annual interviews with tutors regarding the use of the test.

Why have I been invited to take part?

You have been invited to express an interest in taking part because as Admission Tutor (or equivalent) you will have knowledge about how selection works locally, how the UKCAT is used and (in some cases) how this use has changed over time. You will be key to influencing selection processes locally and may have views on how the UKCAT might develop in the future.

Please note I will be interviewing a *sample* of those who volunteer to take part in this study to ensure I speak to a representative group of admission tutors.

Do I have to take part?

It is up to you to decide whether or not you would like to take part. If you do, then please reply to my email to confirm this. You can still withdraw at any time. You do not have to give a reason.

What will happen if I offer to take part in this study?

Once I have responses from admission tutors I will identify a representative sample to interview. If you are part of that sample I will contact you to arrange a time convenient to you when we can conduct the interview. At that point I will let you have a copy of a consent form to complete.

Interviews will take place over the phone and will be recorded. Interviews will last at most an hour. There are no other commitments or lifestyle restrictions associated with participating.

If you are not part of the selected sample then I will let you know.

What are the possible disadvantages and risks of taking part?

Participating in the research is not anticipated to cause you any disadvantages or discomfort. The potential physical and/or psychological harm or distress will be the same as any experienced in everyday life.

What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will have a beneficial impact on the future development of UKCAT, ensuring that the organisation and the test develop in response to consortium views. Results will be shared with participants.

What happens if the research study stops earlier than expected?

Should the research stop earlier than planned and you are affected in any way we will tell you and explain why.

What if something goes wrong?

If you have any complaints about the project in the first instance you can contact any member of the research team. They should acknowledge your concern and give you an indication of how they intend to deal with it.

If you remain unhappy and wish to complain formally, you can do this by contacting the FMHS Research Ethics Committee Administrator, Faculty Hub, Medicine and Health Sciences, E41, E Floor, Medical School, Queen's Medical Centre Campus, Nottingham University Hospitals, Nottingham, NG7 2UH or via E-mail: FMHS-ResearchEthics@nottingham.ac.uk.

Will my taking part in this project be kept confidential?

Any information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified or identifiable in reports or publications. Your institution will also not be identified or identifiable.

Data collected may be shared in an anonymised form to allow reuse by the research team and other third parties. These anonymised data will not allow any individuals or their institutions to be identified or identifiable.

Will I be recorded, and how will the recorded media be used?

Interviews will be recorded. Recordings will be transcribed (digitally) and analysed using software in order that themes arising out of the interviews can be identified.

What type of information will be sought and why is the collection of this information relevant for achieving the research project's objectives?

Interviews will focus on how UKCAT is used locally and how this has changed. Views on the test and thoughts about future development will also be of interest.

What will happen to the results of the research project?

Results of the research will be published and will form part of my PhD thesis. Interviewees will not be identified in any report or publication. Institutions will not be identified in any report or publication. Participants can request a copy of reports resulting from the research.

Who is organising and funding the research?

The project forms part of my PhD which is funded by the KCAT Consortium.

Who has ethically reviewed the project?

This project has been ethically approved by the University of Nottingham, Faculty of Medicine and Health Sciences Research Ethics Committee. tbc

Contacts for further information

Rachel Greatrix, WD1419, Medical School, QMC, Nottingham NG7 2UH Tel: 0155 8230041 rachel.greatrix@nottingham.ac.uk

Professor Sandra Nicholson, Centre for Medical education, Queen Mary University of London s.nicholson@qmul.ac.uk

Appendix K: Participant Information Sheet

University of Nottingham Ethics Committee Approval ID: XXXXXX NOTE CHANGE

Admission Tutor Views of the Impact of the UKCAT on Medical and Dental Selection

Rachel Greatrix, School of Medicine, University of Nottingham

Rachel.greatrix@nottingham.ac.uk

0115 8230041

Invitation

You are being invited to take part in the above research project. Before you decide to do so, it is important you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask me if there is anything that is not clear or if you would like more information. Thank you for reading this.

What is the project's purpose?

The aim of this project is to better understand how Admission Tutors view the impact UKCAT has had on selection locally and why this has occurred. To this end, interviews will explore:

- How and to what extent the UKCAT impacts on selection locally;
- How usage has changed (if at all) and what the drivers have been around this change;
- What the challenges are to using UKCAT in selection;
- How admission tutors would like UKCAT to develop to better support selection.

The project forms one part of my PhD which is looking at the broader impact of the UKCAT, drawing on information from the literature, archived materials and my annual interviews with tutors regarding the use of the test.

Why have I been chosen?

You have been chosen because as Admission Tutor (or equivalent) you will have knowledge about how selection works locally, how the UKCAT is used and (in some cases) how this use has changed over time. You will be key to influencing selection processes locally and may have views on how the UKCAT might develop in the future.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part please keep a copy of this information sheet and indicate your agreement on the consent form. You can still withdraw at any time. You do not have to give a reason.

What will happen to me if I take part?

Once I have your consent form, I will contact you to arrange a time convenient to you when we can conduct the interview. Interviews will take place over the phone and will be recorded. Interviews will last at most an hour. There are no other commitments or lifestyle restrictions associated with participating.

What are the possible disadvantages and risks of taking part?

Participating in the research is not anticipated to cause you any disadvantages or discomfort. The potential physical and/or psychological harm or distress will be the same as any experienced in everyday life.

What are the possible benefits of taking part?

Whilst there are no immediate benefits for those people participating in the project, it is

hoped that this work will have a beneficial impact on the future development of UKCAT, ensuring that the organisation and the test develop in response to consortium views. Results will be shared with participants.

What happens if the research study stops earlier than expected?

Should the research stop earlier than planned and you are affected in any way we will tell you and explain why.

What if something goes wrong?

If you have any complaints about the project in the first instance you can contact any member of the research team. They should acknowledge your concern and give you an indication of how they intend to deal with it.

If you remain unhappy and wish to complain formally, you can do this by contacting the FMHS Research Ethics Committee Administrator, Faculty Hub, Medicine and Health Sciences, E41, E Floor, Medical School, Queen's Medical Centre Campus, Nottingham University Hospitals, Nottingham, NG7 2UH or via E-mail: FMHS-ResearchEthics@nottingham.ac.uk.

Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified or identifiable in reports or publications. Your institution will also not be identified or identifiable.

Data collected may be shared in an anonymised form to allow reuse by the research team and other third parties. These anonymised data will not allow any individuals or their institutions to be identified or identifiable.

Will I be recorded, and how will the recorded media be used?

Your interview will be recorded. Recordings will be transcribed (digitally) and analysed using software in order that themes arising out of the interviews can be identified.

What type of information will be sought from me and why is the collection of this information relevant for achieving the research project's objectives?

The interviewer will ask you about how UKCAT is used locally and how this has changed. Your views on the test and thoughts about future development will also be of interest.

What will happen to the results of the research project?

Results of the research will be published and will form part of my PhD thesis. You will not be identified in any report or publication. Your institution will not be identified in any report or publication. If you wish to be given a copy of any reports resulting from the research, please ask me.

Who is organising and funding the research?

The project forms part of my PhD which is funded by the UKCAT Consortium.

Who has ethically reviewed the project?

This project has been ethically approved by the University of Nottingham, Faculty of Medicine and Health Sciences Research Ethics Committee. tbc

Contacts for further information

Rachel Greatrix, WD1419, Medical School, QMC, Nottingham NG7 2UH Tel: 0155 8230041 rachel.greatrix@nottingham.ac.uk

Professor Sandra Nicholson, Centre for Medical Education, Queen Mary University of London s.nicholson@qmul.ac.uk

Thank you for taking part in this research.

Appendix L: Consent Form

Admission Tutor Views of the Impact of the UKCAT on Medical and Dental Selection

Rachel Greatrix, School of Medicine, University of Nottingham

Rachel.greatrix@nottingham.ac.uk 0115 8230041

Research Participants name: XXXXXXXXXXXXXXXXXXXXX

Interviews will last approximately an hour. We do not anticipate any risk associated with participation, but you may stop the interview or withdraw from the research at any time.

Thank you for agreeing to take part in this research project. Ethical procedures for academic research undertaken in UK institutions require interviewees to explicitly agree to being interviewed and how information obtained in interviews is used. This consent form is necessary so that we are sure you understand the purpose of your involvement and that you have agreed to the conditions of participation. Please therefore read the accompanying **information sheet** and sign this form to certify you approve that:

- the interview will be recorded and a transcript produced
- the transcript of the interview will be analysed by RG as research investigator
- access to the interview transcript will be limited to RG and academic colleagues and researchers with whom she might collaborate as part of the research process
- summary interview content/direct quotations, made available through publication or other academic outlets will be anonymized so that you and your institution cannot be identified
- recordings will be destroyed once the transcript has been finalised
- any variation of these conditions will only occur with your explicit approval

By signing this form I agree that:

- I am voluntarily taking part in this project. I understand that I don't have to take part, and I can stop the interview at any time;
- The transcribed interview or extracts from it may be used as described above;
- I have read the Information sheet;
- I don't expect to receive any benefit or payment for my participation;
- I have been able to ask any questions I might have, and I understand that I am free to contact the researcher with any questions I may have in the future.

Participant Name _____

Signature _____ Date _____

Researcher's Name _____

Signature _____ Date _____

Contact Information

This research has been reviewed and approved by the University of Nottingham, Faculty of Medicine & Health Sciences Research Ethics Board. If you have any further questions or concerns about this study, please contact:

Name of researcher: Rachel Greatrix

Full address: UKCAT, WD1419, QMC, Nottingham NG7 2UH

Tel: 0115 82330041

E-mail: rachel.greatrix@nottingham.ac.uk

You can also contact Rachel Greatrix's supervisor:

Name of researcher: Susan Anderson

Full address: Room E127 Medical School, QMC, Nottingham NG7 2UH

Tel: 0115 8231575

E-mail: susan.anderson@nottingham.ac.uk

What if I have concerns about this research?

If you have any complaints about the project in the first instance you can contact any member of the research team. They should acknowledge your concern and give you an indication of how they intend to deal with it.

If you remain unhappy and wish to complain formally, you can do this by contacting the FMHS Research Ethics Committee Administrator, Faculty Hub, Medicine and Health Sciences, E41, E Floor, Medical School, Queen's Medical Centre Campus, Nottingham University Hospitals, Nottingham, NG7 2UH or via E-mail: FMHS-ResearchEthics@nottingham.ac.uk.

Appendix M: Interview Guide

Admission Tutor Views of the Impact of the UKCAT on Medical and Dental Selection

Rachel Greatrix, School of Medicine, University of Nottingham

Rachel.greatrix@nottingham.ac.uk 0115 8230041

The aim of this project is to better understand how Admission Tutors view the impact UKCAT has had on selection locally. Today's interview will explore:

- How and to what extent the UKCAT impacts on selection locally;
- How usage has changed (if at all) and what the drivers have been around this change;
- What the challenges are to using UKCAT in selection;
- How admission tutors would like UKCAT to develop to better support selection.

Note to interviewees:

- confirm receipt of consent form
- note discussion is being recorded
- note confidentiality
- note they are free to stop the interview/recording at any point
- do you have any questions about today's interview at this stage?

Question	**Potential Prompts**
Please confirm your name, University and role.	
How long have you been involved in medical/dental admissions and it what roles?	
Can you briefly describe how your University *currently* uses UKCAT in selection?	• At what stages is the UKCAT used? • What other criteria are used at that stage? • *Summarise use and check understanding if required*
Given how you use the UKCAT in selection, how big is the impact UKCAT has on selection decisions compared with other criteria?	• Check understanding of this question
UKCAT was first used by Universities in 2006.	

Are you able to describe how selection processes at your University have changed since then? How has the use of UKCAT changed during this period?	• Note some might not be familiar with this entire period. • *Summarise and check understanding if required. Focus might be on changes to use of personal statements.*

I am keen to know why those changes have taken place.
 	
What in your view has driven change in medical/dental selection locally?
 	
Are there specific drivers for change around the use of the UKCAT?
	• What external drivers?
	  ○ Selection Alliance
	  ○ Research outcomes
	• Internal drivers?
	  ○ Management intervention?
	  ○ Personal preferences?
	  ○ Defensibility
	  ○ Impact of numbers?
	  ○ Administrative ease of use
	
	Prompt around some of these drivers to obtain a view as to which are relevant.

Do you have a sense as to whether the use of UKCAT has changed the demographics in selection in anyway?	• Gender?
Do you anticipate your use of the UKCAT changing in the near future?	• In what way? • What current issues are driving change?
What is the biggest challenge you currently face in selection?	• What steps are being taken to overcome this?
The UKCAT originally set out to achieve greater fairness in selection, widen participation and improve the research evidence base around selection to medicine and dentistry. I would like to know to what extent you believe it has achieved these aims. • Do you think the UKCAT has led to greater fairness overall in selection?	

• Has has contributed to widening participation? • What about UKCAT's contribution to improving the research evidence base around selection to medicine and dentistry?	
How would you like to see UKCAT develop in future years?	• Would anything encourage you to use the UKCAT more strongly in selection?
Have you any other comments about how UKCAT has developed or its impact on selection?	• Test? • Consortium? • Support? *Encourage to contact me by email afterwards if anything further occurs to them.*

Appendix N: Coding Framework

Level 1	Level 2
Autonomy	
Applicant feedback	Applicant feedback (local)
	Applicant feedback (no)
	Applicant feedback (UKCAT)
Examples of change	
Fairness	Fairness (fairness)
	Fairness (UKCAT negative)
	Fairness (UKCAT neutral)
	Fairness (UKCAT positive)
Future change	Future change (local)
	Future change (UKCAT)
Other	Other (driver for change local)
	Other (driver for change)
	Other (impact local resource)
	Other (other)
	Other (Selection challenge attributes)
	Other (selection challenge local)
	Other (selection challenge time, resource)
	Other (test content)
Research	Research (confident knowledge)
	Research (driver for change)
	Research (negative)

Level 1	Level 2
	Research (positive)
	Research (unconfident knowledge)
UKCAT	UKCAT (driver for change)
	UKCAT (impact demographics)
	UKCAT (impact time, resource)
	UKCAT (organisation)
	UKCAT (SJT)
	UKCAT (use impact low)
	UKCAT (use impact)
	UKCAT (use impact, high)
	UKCAT (use impact, local)
	UKCAT (use impact, negative)
WP	WP (driver for change)
	WP (local)
	WP (negative)
	WP (neutral)
	WP (positive)
	WP (selection challenge)

Appendix O: Research Ethics Consent Admission Tutor Interviews

University of Nottingham
UK | CHINA | MALAYSIA

Faculty of Medicine & Health Sciences
Research Ethics Committee
Faculty Hub
Room E41, E Floor, Medical School
Queen's Medical Centre Campus
Nottingham University Hospitals
Nottingham, NG7 2UH
Email: FMHS-ResearchEthics@nottingham.ac.uk

24 February 2020

Rachel Greatrix
PhD Student Medical Education
WD1419, Medical School
QMC Campus
Nottingham University Hospitals
Nottingham
NG7 2UH

Dear Ms Greatrix

Ethics Reference No: 458-2001 **please always quote**	
Study Title: Admission Tutor Views of the Impact of the UKCAT on Medical and Dental Selection	
Chief Investigator/Supervisor: Susan Anderson, Professor of Anatomy, Life Sciences	
Lead Investigators/student: Rachel Greatrix, PhD Student, School Medicine	
Other Key investigators: Professor Sandra Nicholson, Centre for Medical education, Queen Mary University of London	
Proposed Start Date: 01/02/2020	**Proposed End Date:** 31/10/2020

Thank you for submitting the above application and the following documents were received:

- FMHS REC Application form and supporting documents version 1.0: 09/01/2020

These have been reviewed at the meeting on 24 January 2020 and are satisfactory. The project has been given a favourable opinion.

A favourable opinion has been given on the understanding that:

1. The protocol agreed is followed and the Committee is informed of any changes using a notice of amendment form (please request a form).
2. The Chair is informed of any serious or unexpected event.
3. An End of Project Progress Report is completed and returned when the study has finished (Please request a form).

Yours sincerely

Dr John Williams, Associate Professor in Anaesthesia and Pain Medicine
Chair, Faculty of Medicine & Health Sciences Research Ethics Committee