# UKCAT

## 2006 Annual Report

UK Clinical Aptitude Test

# UKCAT Annual Report 2006

## CONTENTS

# PartI:

# Introduction

# 1. Chairman's Introduction

Competition for places in medical and dental schools in the UK is fierce. During the last few years professionals involved in admissions have become increasingly unhappy with the processes used. The difficulty of differentiating between candidates with outstanding grades in public examinations, with very similar, carefully crafted personal statements and bland references called into question the fairness and validity of the available systems. Increasing reliance was being placed upon 'measurable' criteria such as starred GCSEs and A level grades, with the suggestion that A grade at A level might be subdivided further, by the introduction of an A* or by giving actual percentage marks.

This concentration upon purely academic results ignores the other factors needed for successful clinical practice. We are also all too well aware that academic achievement is closely related to educational opportunity, and not always to academic potential. Although significant intelligence and intellectual ability is a prerequisite for success, there is little point in going on about common sense, problem solving skills, conscientiousness, empathy, resilience and all of the other myriad of skills and attitudes that are an advantage to the future doctor or dentist, if in the end entry depends upon getting 92% in chemistry A level rather than 91%. Something needed to be done, both to solve the logistic problems of fairly selecting the few from so many excellent candidates, and to address the issues around recognition of non-academic factors.

A group of academics and administrators responsible for admissions to medicine and dentistry for most of the Universities in the UK have formed a consortium to address the problems. Admissions tests for courses such as medicine have been used in the USA and Australasia for many years. Evidence for the predictive validity of these tests in selecting candidates who will become "good" doctors or dentists is lacking, but no more so than with all of the other factors currently used in the admissions process. Based upon experience internationally and in other professional areas we have designed a new selection test, the UKCAT, together with a planned research and auditing programme designed to identify predictive validity and to allow modifications to be made to the admissions processes as evidence becomes available.

The UKCAT was introduced in 2006, and although there were some minor logistic problems, the project was generally successful and has been widely accepted. This has been in no small part due to the hard work and commitment of individuals within UKCAT and Pear-

sonVue. As the project rolls forward there will be increasing experience, confidence and evidence, all allowing sensible and fair modifications to be made to the admissions processes with the promise of more rational selection of tomorrow's doctors and dentists.

PROFESSOR IAN JOHNSON
*Chair, UKCAT Consortium*

# 2. Summary

**Introduction** The UKCAT is a new entrance test for applicants to medical and dental schools, derived from existing selection tests that are already used for selection in job applications and in undergraduate law schools. The test is an appraisal of aptitudes, not of knowledge. The UKCAT was developed and delivered by Pearson Vue and its associates, in collaboration with representatives of the participating medical and dental schools (Pearson Vue and its partners were already experienced in developing and delivering other aptitude tests, including the entrance test for law schools). It was selected by representatives from all those medical and dental schools that were part of the consortium at the outset (23 universities), in an open tender process that included comparison with other tests currently used in undergraduate medical selection (there were no entrance tests for dental schools at the time). It is managed by a Board, elected from among the representatives of the participating medical and dental schools. The Board is answerable to the whole consortium of schools, whose representatives meet twice in each year [section 8, page 34].

A key feature of the UKCAT is a research programme to monitor undergraduate progress through medical and dental schools, and to see how well the test predicts student performance during training and, eventually, in the early years after qualification [section 7, page 33].

**This Document** The purposes of this document are to describe the test and its management, to present a summary of the information that we have available from the test after its first year of use, and to explain the future directions of development of the test and of its associated research programme. The document consists of three parts:

- this summary;

- an overview of the UKCAT, presented as a series of answers to common questions that we have received from students, medical school representatives and journalists [page 10];

- a more detailed analysis of our experience of the first year of the test and of our strategy for future development [Part 3, page 19].

**The Test** In its first year, the test consisted of four sections: verbal reasoning, quantitative reasoning, abstract reasoning and decision analysis. The first three sections were derived from existing aptitude tests used for selection in other contexts: the decision analysis section, designed to appraise the ability to make judgements under conditions of increasing complexity and ambiguity, was relatively new. It had already been under development by Pearson Vue's associates with the aim of simulating the real world more effectively; this test was seen to fit UKCAT's needs and it was chosen in order to broaden the cognitive skills being tested [section 4, page 20]. The test was delivered on-line, in

any one of eight permutations, in order to allow us to offer the test at the candidates' convenience while maintaining the security of the

6

test [section 5, page 22]. The scores varied very little (about 1.5%) between different permutations of the test [page 31].

**Reliability** The reliability of the scores in the four sections has been found to be good (Cronbach's alpha> 0.7) in three sections of the test, and slightly lower in the new Decision Analysis section [page 30]. We are working to improve the reliability of this new section.

**Fairness** The fairness of the test has been assessed by comparing the total test scores by age, sex, ethnic group and socio-economic group. Overall the distorting effect of these factors on the measurement of the defined aptitudes seems small [pages 28 to 29]. At a more detailed level, a differential item function (DIF) analysis has been performed to check that individual questions did not discriminate against particular groups identified by age, sex and ethnic group: again, the effects on any particular group seem very small [page 30].

**Accessibility** The UKCAT is a test of aptitude, not of knowledge of any curriculum: in principle, that ought to mean that candidates from any educational background are competing on equal terms and that the advantage from specific teaching for the test is minimized.

The UKCAT Consortium charges a fee for the test (to cover the costs of development and delivery of the test), but also makes available bursaries to candidates who meet specified criteria. In 2006, 1009 candidates (about 5%) received bursaries to allow them to take the test without charge [page 23].

**Extended Test** Candidates with dyslexia or other disabilities were offered longer times to complete the test. In 2006, 357 candidates (1.9%) took up this option.

**Delivery** The practical delivery of the test is through a network of testing centres administered by Pearson Vue: candidates book a time and place to take the test through an on-line registration service. There were very few practical problems with this aspect of the test: some candidates complained about the on-line calculator (provided as an aid to arithmetic in the quantitative reasoning section), and this problem has been addressed for 2007; and a few candidates, particularly those who registered late, were unable to arrange the times and places of their first choice, but all candidates were accommodated before the end of the testing period [section 5, page 22].

**Development** There is a constant programme of new item development in order to "refresh" the tests on a regular basis; this strategy includes the testing of new questions as non-scoring components of the test.

**The Future** A new, fifth, section of the test was launched in 2007, aiming to appraise non-cognitive skills which are considered important in the practice of medicine and dentistry – namely integrity, empathy and resilience. Also in 2007, the Research Working Group began to collect data on the progress of candidates through their undergraduate courses, a process which will eventually form the basis for assessing the validity of the test as a selection tool [section 7, page 33].

# Part II:

# Overview

# 3. Answers to Common Questions

## Why are we doing this?

### Justification for the UKCAT

The UKCAT was conceived to improve the fairness and objectivity of the Admissions process for clinical subjects (medicine and dentistry initially, though some vet schools have also expressed an interest in using the test). It arose partly from a widespread feeling that A-levels, which were the only other objective indicator of student performance used by most medical and dental schools at the time, were failing to discriminate between candidates at the upper end of the scale of academic ability. The failure in capacity to discriminate seemed to be progressively worsening, and there seemed to be no immediate plan to provide a better test. There was a feeling among some admissions tutors that our selection processes could not be completely fair as long as so many applicants were scoring similar grades at A-level, and when the only practical discriminator – the number of A-grade results obtained – might have depended more on the calibre and resources of candidates' schools than on the candidates' own 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.

In addition, it was hoped that the test might eventually be used to assess characteristics not easily assessed by other means.

## Why not a selection test for university entrance, rather than just for specific subjects?

The participating schools were keen to establish a test as quickly as possible, and at the time no general selection test for university entrance was available or close to fruition. An additional reason for considering a separate test for medicine and dentistry was the feeling that selection for these courses might look at qualities that are not necessarily required for other subjects: the new fifth section of the test, initiated in 2007, is our first step in that direction.

## Why didn't we just use an existing entrance test?

The schools involved in the development of the UKCAT wished to begin with a clean sheet, to set their own criteria, and to have control over the test themselves, with a view to experimenting with new selection techniques.

The tests in common use at the time of inception of the UKCAT were the GAMSAT, MSAT, UMAT and BMAT, all used by medical schools (no dental school used an entrance test at the time). The companies involved in producing those tests were free to tender for the production of the UKCAT; some did so, and their proposals were compared on an equal basis with the other tenders.

## How does the UKCAT compare with existing tests?

We don't yet know how the UKCAT compares with other entrance tests in terms of its ability to predict candidates' performance in medical and dental training. The UKCAT consortium has no plan at present to compare selection with the UKCAT with selection by other tests, but of course individual schools may wish to make such comparisons if they are in a position to do so. There is a common core of aims in all entrance tests, but some other tests for medicine have specific aims that are not part of the remit for the UKCAT: for example, the UKCAT specifically does not test scientific knowledge at either GSCE or A-level.

## Why is there no science component, given that we were trying to improve on A-levels as a selector?

Scientific knowledge is already assessed by school-leaving exams, which have been shown to have a degree of correlation with performance on medical courses. However, the schools participating in the UKCAT consortium felt that they did not wish to develop this assessment further, for a number of reasons.

First, the medical and dental schools felt that an appraisal of factual learning didn't seem to be a good way to select their students; and if we were to stray into appraisal of scientific ability beyond factual learning, we would risk making the test inaccessible to some candidates (a criticism often levelled at the former Oxbridge entrance exams).

Second, we felt that there was no point in retesting material already tested by A-levels and similar exams. We accept that a science test might be useful to some graduate-entry courses, with a wide base of candidates; but we felt that for the majority of schools involved in the UKCAT a science component would be an unwelcome complication at this stage.

Third, we wanted to ensure that the UKCAT allowed all candidates to compete on equal terms, irrespective of their educational background. This aim meant that we should concentrate on measuring aptitude, rather than knowledge: schools that require a measure of educational attainment should use another test (such as A-levels) alongside the UKCAT.

The UKCAT tests a wide range of mental processes, and we believe this approach allows us to be fair to candidates from all backgrounds, as well as to medical and dental schools with a wide variety of admissions policies, not all of which focus exclusively on science qualifications.

## How is the test managed?

The overall strategy for UKCAT is set by the consortium of the participating medical and dental schools, each of which sends a representative to full meetings of the consortium twice a year. Responsible to this body is a board of eight members elected by the consortium, plus four members appointed to represent the Council of Heads and Deans of Dental Schools and the Medical Schools Council [section 8, page 34] (modified in 2008 to nine elected and three appointed members): this board meets at least four times per year, and takes decisions or makes recommendations on more detailed management issues. Each year three members of the board stand down and elections are held for the places (resigning members of the Board are eligible for re-election).

The test itself is managed and delivered on a day-to-day basis by Pearson Vue, which has contracted the setting of the test to a company with extensive experience in setting and marking tests in the context of other academic and employment selection procedures.

Oversight of the setting and delivery of the test on behalf of the consortium is provided by two sub-committees of the Board: the Test Delivery Group (which monitors the delivery of the test, ensures that there are adequate places available and oversees the distribution of the results), and the Test Development Group, which guides the development of the test and ensures that the questions are consistent with the aims of the consortium. Both of these groups consist of members of the Board, afforced by other consortium members and external advisers; they are chaired by members of the Board, who are themselves representatives of participating medical and dental schools.

## Is it a good test?

### A good test of what? Aims of the UKCAT

The test aims to select students who will perform well in medical or dental school and who will eventually make good doctors or dentists. The definition of a good clinical practitioner may be (justifiably) nebulous, and in the first instance we aim to establish whether the UKCAT can identify candidates who will fail during their undergraduate studies. The UKCAT consortium has established a research programme to follow the progress of all students who sit the test as they progress through medical and dental school. By this means, and by allowing feedback from this longitudinal study to the Test Development Group, we hope to develop and use the tests in the most effective way for enhancing the selection of applicants (i.e., for identifying those who are more likely to qualify and those who are more likely to perform well once they start practising their profession).

### How do we know how good it is?

Choosing good quality tests at the beginning is a necessary but not sufficient step. As with all selection procedures, good follow-up will enable the process to be checked and improved. UKCAT is committed to monitoring the performance of these tests as part of a long-term validation exercise. Three of the four tests being used by UKCAT were compiled from items already pre-tested in other contexts so that the characteristics of the questions were known and the tests would have predictably high reliability (which subsequent analysis has confirmed). The fourth test was chosen because it was designed to measure a different kind of cognitive and problem solving skill. Choosing this range of tests of cognitive ability rather than additional measures of science or other attainments was a deliberate policy in order to broaden the criteria for selection beyond the abilities already assessed by A-Levels. However, such tests are still likely to focus on the abilities required for managing the academic challenges and problem solving elements of professional practice. Most people recognize that even this combination does not cover the full range of skills and qualities required from practising medics and dentists, and so we are introducing additional tests to broaden the assessment still further into the realm of the more behavioural elements required – the fifth section of the test introduced in 2007 is the first step in this direction.

## How do we know that it is fair?

Part of the question of fairness depends on what exactly we are aiming to select for, but obviously everyone will wish to be reassured that the test does not contain an unfair bias against particular groups of people. We have analysed the overall scores of the 2006 test to look for evidence of bias against either sex or against any particular ethnic or socio-economic group: the results of our analysis are included in section 6, beginning on page 28. Our interpretation of this analysis is that, although the mean scores are (as expected) not identical between all groups, the test does not discriminate unfairly against any particular group. In addition to this analysis of overall scores, Pearson Vue conducted an analysis on our behalf (based on a process called DIF (Differential Item Functioning) analysis, which has become a standard way for evaluating bias in test items) of the relative performance of the two sexes and of different ethnic groups and age groups on each question in the test database to see whether any individual questions were poorly answered by particular groups. Their conclusion was that there were very small biases in a small number of questions in the database, but that they did not produce a significant effect on the final score [page 30]. Nevertheless, work has begun to replace or redesign those questions that showed any bias in the test.

An additional check on the quality of the test comes from the insertion of non-scoring questions into the test: these are questions that we are trying out, to see whether they are fair. They do not contribute to the candidates' scores during the trial year, but if they prove satisfactory then they may be used (as normal, scoring questions) in a future year.

## Why don't all candidates take the same test?

All candidates do take the same test, in the sense that they all answer questions from the same database, which are designed (and tested) to be equivalent. However, if all candidates were to face exactly the same questions then the only way to make the test secure would be to insist that all candidates sit it at the same time, which would pose difficulties for some candidates and would also mean that we should have to find a large number of venues and trained supervisors to be available at the same time. The variation in questions means that we can allow candidates to sit the test at different times over an extended period without compromising the security of the test.

Obviously the variation in the test opens the possibility that some combinations of questions may in practice be easier than others: we have addressed this question [page 31] and have found that the variation in scores between the lowest- and highest-scoring combinations in the test is small – of the order of 11/2%.

## How can we be sure that there were enough places for candidates wishing to take the test?

The delivery of the test was supervised by the Test Delivery Group, working closely with Pearson Vue, whose testing centres we used. We had some indication of the number of candidates to expect, and of their geographical distribution, from UCAS data of previous years' admissions. Pearson Vue's testing centres covered the areas we needed, and appeared to provide sufficient capacity: we also benefited from the considerable experience that Pearson Vue has accumulated in delivering similar tests for other institutions. In the event, there were very few problems connected with the delivery of the test (these are detailed in section 5, page 24).

## How do we know that the candidate who takes the UKCAT is the same as the candidate who writes the UCAS form and who turn s up for interview?

Candidates are required to bring a form of identification with them to the test centre when they take test, together with the printout of their registration for the test. The identification must include a photograph and a date of birth, and well as the full name: passport, photocard driving licence, government-issue identity card are examples of accept- able identification. Candidates who arrive without an accepted form of identification are not allowed to take the test. To that extent, we know that the candidates who take the test are who they say they are. Those universities that require similar forms of identification from candidates coming to interview will be reassured that the candidate in the interview is the same as the candidate who took the test.

## Medical and dental schools are not obliged to use the UKCAT in a standard way: why not?

We felt that we had no right to tell medical and dental schools how to use the results of the UKCAT, particularly since the test was new and and none of us could have had any special experience of how it would predict student performance. We have some information on how schools did use it in practice, and it seems that there was variation in the weight given to the test score, and in the point of the admissions process at which the test score was used to influence the outcome. However, there is already some variation in the way in which medical and dental schools use A-level results and data from the UCAS form during their admissions processes, and it would seem unreasonable to insist that the UKCAT should be used in a uniform way when other data are not. We regard these variations as healthy, and as part of our collective wish to find out how best to improve our admissions processes.

## Should we be considering the four scores separately, or just looking at the total?

At this stage, we would not feel able to give firm directions to schools on how they use the test. One reason for having four sections to the UKCAT is to ensure that the broad spectrum of candidates' abilities is fully sampled (so that, for example, we are not giving a particular advantage to candidates who are particularly strong in any one subject area); and to that extent the total score may represent the fairest mark. However, we will continue to provide separate scores for the four sections of the test because we believe that the four sections may be testing different things, and some schools may wish to pay particular attention to the specific aptitudes being tested. Our research into the predictive value of the test may eventually show that one section is a particularly good predictor of progress in medical school, but it would be too early for us to make such a claim at present.

We have looked at the correlation between the scores for the four sections of the test [page 31], and although the scores are not completely independent, the correlation is not strong; this result implies that candidates do have particular strengths and weaknesses which are showing in the separate sections of the test.

## How can we know that the UKCAT is not unfairly distorting the admissions process?

Our evidence is that most schools used the UKCAT fairly lightly in its first year (2006), and that no school has substantially modified its procedures because of the new test. In this sense, the UKCAT has provided some additional information but has not caused a fundamental revision of admissions processes. It seems unlikely, therefore, that there is much distortion of existing processes.

One area where there has been some distortion is that a small number of those medical schools that did not use an entrance test in 2006 experienced a large increase in the number of applicants, while some schools using the UKCAT for the first time saw a small fall in numbers [page 32], but it is hard to see how we could have prevented such an effect.

## How much support is available to candidates trying to register for the UKCAT?

For 2006, we provided a very detailed web site, where the registration process was explained step by step. The same web site offered some sample questions, which candidates were encouraged to try by way of practice before the test. In addition, candidates could contact the UKCAT administration by e-mail if they had difficulties with the registration process.

In practice, the administrative staff were overwhelmed by e-mails, almost all asking questions that were answered on the web site. It became clear that written instructions on the web were not adequate for some candidates, and for 2007 we have implemented a telephone support line. This service is provided by Pearson Vue, and the cost of the service will be recovered from the registration fee for the test: the provision of this service is one of the reasons for the small increase in registration fee for 2007.

In response to feedback on the practice questions, we are developing this resource further, both by expanding the number of questions available and by offering a timed test, so that it will be easier for candidates to practise in conditions similar to those that they will encounter at the test centre.

## How can we justify charging a fee for a university entrance test?

It costs money to provide the test, and the money has to come either from the candidates or from the universities themselves. The universities could not have afforded to bear the whole annual cost of the UKCAT (though they did make a financial contribution to the start-up costs in 2006), and so we have to charge a fee to the candidates. In 2006, our main expenses were the direct costs of the test (paid to Pearson Vue to cover the costs of setting the questions, administering the test and distributing the results); legal fees to set up the consortium; the costs of establishing the research group to validate the UKCAT; and the cost of the bursaries, to ensure that candidates who could not afford the fee were not prevented from taking the test (Financial Statement, page 39). In future years, we anticipate that the legal costs will fall and the research costs will rise as the programme to develop the test and to validate it becomes established (for instance, management of the database is likely to incur costs). We may also need to take on more staff to help with the administration of the company. However, we do not expect to see large increases in the cost to the candidates.

We offered bursaries to candidates who were in receipt of income support and/or educational maintenance allowances; the bursaries were taken up by 877 candidates (4.7% of all candidates). Details are given on

# Part III:

# Detailed Analysis

# 4. Question Development

## 4.1 Foundations

The UKCAT was fortunate in being able to build on the foundation of a large database of questions, already tried in contexts similar to our own intended use: Pearson Vue had chosen as an associate an experienced player in the process of selection at the undergraduate level through its involvement in the LNAT, an aptitude test similar in concept and in aim to the UKCAT, and which had been running for two years at the time the UKCAT started. The company's other experience included selection for employment in the civil service and in the armed forces.

Clinical practice requires the ability to handle a vast amount and a broad range of different types of information. Verbal, quantitative and abstract reasoning tests have been shown consistently to predict the ability to do this well across a wide range of occupations. It was therefore a natural choice to include such tests in the UKCAT procedure. However, the full range of cognitive skills required by doctors and dentists extends beyond these areas. One area of particular interest is the ability to handle incomplete and ambiguous information. Pearson Vue's associate company was in the process of developing a different kind of test to assess such abilities as a result of demand from the business world where this is the reality – people rarely have all the information they would like and it is not always consistent. The consortium decided to trial this new test alongside the more traditional reasoning tests in line with its belief that a broader set of cognitive skills could enhance the effectiveness of the selection procedure.

## 4.2 Initial Testing

An initial subset of questions for the verbal, quantitative and abstract tests was chosen from items that had been developed and trialled extensively on UK undergraduates across a wide range of disciplines. This meant that UKCAT could use new tests but nevertheless could be confident that the tests would conform to minimum standards of reliability – and subsequent analysis has confirmed the predictions for these tests.

## 4.3 Calibration

The aim of the test designers was to produce scores for each section of the test that would differentiate adequately within the target group – the tests should be neither too easy nor too difficult. To this end, questions from the established database (which had previously been used in other tests, intended for other groups) were calibrated for the target population of prospective medical and dental students: the calibration was based on the first 3000 applicants to sit the whole test, which gave us a sufficiently large sample for calibration. A rough calibration had been made on the basis of the initial screening tests in the previous autumn, but in order to refine the calibration the scores were temporarily withheld from the first 3000 candidates who took the test until the calibration exercise, based on these first 3000 scores, had been completed. This exercise provided confirmation that the tests were functioning as expected. The calibration was complete by the beginning of September, and from 8th September candidates were given their scaled scores immediately after they had completed the test. Those candidates who had taken the test before 8th September were notified of their scaled scores within 10 days of the completion of the calibration exercise. The scale used gives a mean of 600 and a range of 300 to 900 for each section of the test.

# 5. Test Delivery

The delivery of the test – that is, its day-to-day administration and physical delivery to the candidates – was managed by Pearson Vue, and supervised by the Test Delivery Group, a subcommittee of the UKCAT Board [section 8, page 37].

## 5.1 Practical Arrangements

Delivery was through an existing network of testing centres owned and managed by Pearson Vue: there were about 150 centres in the UK, and at least one centre in each of 65 other countries (including all countries of the EU). Candidates were expected to take the test if there was a test centre within their country of residence or in the country where they were receiving their education.

In practice, 30 candidates were exempted from the test; of these, 20 were exempted for broadly medical reasons or reasons related to disability, and the remainder for reasons of access to a test centre.

The size of the network of test centres meant that almost all candidates within the UK were within forty miles of a centre. For the area that was less well served by the network – the north of Scotland – a mobile test centre was provided, visiting advertised points on specific dates.

## 5.2 Timing and Item Count

The test included four sections, each timed separately (so that a candidate could not use time saved on one section to make more time available on another section. The item count (i.e., the number of questions) for each section, and the time allowed (in minutes) for each section, are shown below.

Three of the four sections of the test included questions that were not intended for inclusion in the final score. These questions were new items, being tested for possible inclusion in the database of questions (and which might therefore appear, perhaps in modified form, in tests in future years). The decision analysis section contained no non-scoring items.

| Section | Total items | Pretest items | Time (mins.) |
|---|---|---|---|
| Verbal reasoning | 44 | 4 | 21† |
| Quantitative reasoning | 40 | 4 | 21 † |
| Abstract reasoning | 65 | 5 | 15† |
| Decision analysis | 26 | 0 | 29† |

† – For each section of the test, candidates were allowed an additional 1 minute to read the instructions for the section, in addition to the times shown here.

## 5.3 Extended Test

Candidates with dyslexia or other disabilities were offered longer times to complete the test. In 2006, 357 candidates (1.9%) took up this option.

## 5.4 Distribution of Candidates

2018 candidates took the test in overseas centres. Outside the UK, the countries with the largest number of tests were Ireland (359 tests delivered), Malaysia (237), Canada (155), Singapore (135), Hong Kong (103) and the USA (101).

## □.□ Bursaries

Bursaries to cover the test fee were available to those candidates who applied for them and who were in receipt of educational maintenance allowance at the top rate, or who were members of a household of which at least one member was in receipt of income support. (From 2007, the second criterion changed: income support was a criterion only if it was paid to the candidate directly, rather than to someone in the candidate's household.) 1009 bursaries were awarded, of which 877 were actually redeemed by candidates taking the test: this figure represents 4.7% of all candidates taking the test. The cost of the bursaries was borne from the candidate registration fee.

Of the bursaries awarded, only 3 went to candidates resident overseas: the remainder went to candidates within the UK.

## 5.6 Difficulties, and Lessons for the Future

We encountered a small number of problems in this first season, which have resulted in slight changes to the delivery process for 2007. All of the problems were minor, though obviously in a high-stakes test any inconvenience or difficulty may cause distress to some candidates: we have tried to ensure that we have avoided these difficulties for the 2007 test.

The single commonest source of complaints from candidates was the calculator, which was provided as an on-screen (pop-up) facility during the test (similar to the ones available on most personal computers). We used this facility as a way to guarantee that all candidates were provided with exactly the same resources. However, some candidates apparently failed to realize that the facility existed, while others found it uncomfortable to use. We accept that this feature did not achieve the aims we had set for it, and for the 2007 tests all candidates will be issued with simple four-function hand-held calculators, all of the same design.

Towards the end of the testing period, some candidates experienced difficulty in finding appointments at local test centres at times that were convenient for them. This problem applied only to candidates who had registered late for the test, and particularly in the last few days before the registration deadline. We have tried hard to encourage candidates to register early during the 2007 testing session, to try to reduce the incidence of this problem (and there has always been a small financial incentive to encourage candidates to register early in the testing window). We had anticipated that, during this first year,

a large number of candidates might wait until the end of the testing period to register: we were able to extend the advertised testing period by two weeks to accommodate this late surge in candidates (as we had expected that we might have to do), though even so, a few candidates struggled to find places at convenient venues. We are grateful to Pearson Vue for their unstinting efforts to provide additional places above the "planned excess" in order to cope with these late registrants. In the event, all candidates who had registered by the deadline (or a few days beyond it) were accommodated. For 2007, we did our best to ensure that candidates understood the benefits of early registration, and the registration deadline was set two weeks ahead of the final date for the test, in an attempt to avoid a recurrence of this problem. No planned excess was built into the numbers for 2007, and we did not anticipate an extension of the testing period.

# 6. Statistics

## 6.1 The Candidates

### Registrations

22,182 candidates registered for the test; 18,542 candidates actually took and completed it. The difference includes 871 candidates who cancelled, 620 who failed to attend the test centre, 6 who started the test but did not complete it. In addition, a number of candidates registered more than once through the web site.

### Age

Most candidates who took the test were school-leavers, with more than three-quarters of all applicants being aged 16–19 at the time of taking the test. The commonest age for taking the test was 17 (57% of all candidates). In 2006, only three graduate-entry courses used the UKCAT: the age distribution is likely to change slightly in the future, because other graduate-entry medical courses started using the test in 2007. The age distribution for 2006 was:

| Age | Total | % |
|---|---|---|
| 16–19 | 15051 | 81.2 |
| 20–24 | 2472 | 13.3 |
| 25–34 | 809 | 4.4 |
| >34 | 143 | 0.8 |
| Other* | 67 | 0.3 |
| *Total* | *18542* | |

* – includes candidates who registered a date of birth that would have made them over 100 or less than 10 years old at the time of taking the test.

### Sex

As is often observed in applications to medical courses, there was a preponderance of female candidates:

| Sex | n | % |
|---|---|---|
| Female | 10476 | 56.5 |
| Male | 8066 | 43.5 |

## Ethnicity

Candidates were asked to report their own ethnicity, which we have simplified in this report into broad ethnic groupings. (The self-report categories of nationality and ethnicity were those tested and approved for use in the 2001 Census data.) Distribution of candidates was:

| Ethnic group | % |
|---|---|
| White | 58.1 |
| Asian | 23.7 |
| Black | 4.3 |
| Chinese | 4.2 |
| Mixed | 3.5 |
| Other | 4.7 |
| Not declared | 1.6 |

## Parental Occupation

Candidates were asked during their registration for the test to report their parents' occupations, which we recorded in categories corresponding to those that are used as part of the basis for the National Statistical Socio-economic Classification.

According to these self-reported results, the representation of each occupational group among each of the broad ethnic groupings was as follows *(n=1 8547; figures represent percentage of each ethnic group in each parental occupational category):*

| Occupation | White | Asian | Black | Chinese | Mixed | Other | Not Declared |
|---|---|---|---|---|---|---|---|
| n= | 10783 | 4393 | 800 | 770 | 640 | 862 | 299 |
| Traditional professional | 31.0 | 24.6 | 21.1 | 29.2 | 33.9 | 31.2 | 15.5 |
| Modern professional | 35.0 | 21.3 | 36.0 | 23.8 | 31.3 | 29.4 | 13.5 |
| Senior managers & admins | 10.9 | 8.9 | 6.9 | 14.9 | 10.6 | 11.8 | 2.7 |
| Middle or junior managers | 3.4 | 5.4 | 3.0 | 6.4 | 3.0 | 3.2 | 1.0 |
| Technical & craft occupations | 5.2 | 4.7 | 2.4 | 2.5 | 3.0 | 2.1 | 1.0 |
| Clerical & intermediate occupns | 3.3 | 4.2 | 4.5 | 1.9 | 2.8 | 2.1 | 1.0 |
| Semi-routine manual & service | 1.6 | 5.3 | 3.3 | 3.6 | 1.3 | 1.5 | 1.7 |
| Routine manual & service | 1.0 | 5.9 | 3.1 | 3.2 | 1.9 | 1.4 | 1.0 |
| Do not know | 2.7 | 8.9 | 9.3 | 6.2 | 5.6 | 6.5 | 6.1 |
| Never worked | 0.2 | 1.7 | 3.3 | 0.5 | 0.8 | 2.2 | 0.3 |
| Information withheld | 5.5 | 9.1 | 7.3 | 7.7 | 5.9 | 8.6 | 56.2 |

## 6.2 Overall scores: Medicine and Dentistry

The range of scores of the test overall is shown below. The scores filled most of the anticipated range (300□900, after scaling). Mean scores for candidates for medicine were generally slightly higher than those for dentistry, but there was obviously considerable overlap between the two groups.

*Overall scores for all candidates (n=18540):*

| Section | Mean | SD |
|---|---|---|
| Verbal reasoning | 588 | 95 |
| Quantitative reasoning | 597 | 82 |
| Abstract reasoning | 596 | 86 |
| Decision analysis | 593 | 103 |
| Total scaled score | 2375 | 269 |

*Scores for medicine candidates (n=1 4418):*

| Section | Mean | SD |
|---|---|---|
| Verbal reasoning | 598 | 93 |
| Quantitative reasoning | 604 | 80 |
| Abstract reasoning | 603 | 85 |
| Decision analysis | 602 | 101 |
| Total scaled score | 2407 | 259 |

*Scores for dentistry candidates (n=21 91):*

| Section | Mean | SD |
|---|---|---|
| Verbal reasoning | 557 | 89 |
| Quantitative reasoning | 584 | 77 |
| Abstract reasoning | 584 | 79 |
| Decision analysis | 570 | 98 |
| Total scaled score | 2294 | 239 |

Candidates who applied to more than one type of course (for example, to both medicine and dentistry), or for whom the course information was incomplete at the time they took the test, are excluded from the subject-specific totals. In 2006, the first year of the test, information about applicants' chosen courses (which came from UCAS) was not available for all candidates at the time the analyses were performed; this is not expected to be a problem in future years.

## 6.3 Effects of Sex

The scores for each section of the test, and the total scores, are shown below.

*Overall scores for all candidates (n=1 0474 female; 8066 male):*

| Section | Female Mean SD | | Male Mean SD | |
|---|---|---|---|---|
| Verbal reasoning | 582 | 95 | 596 | 94 |
| Quantitative reasoning | 585 | 81 | 614 | 81 |
| Abstract reasoning | 598 | 87 | 593 | 86 |
| Decision analysis | 591 | 102 | 595 | 104 |
| Total scaled score | 2357 | 268 | 2398 | 268 |

Differences between male and female were significant: $p < 0.001$ for all scores except for Decision Analysis, for which $p < 0.003$ (Mann-Whitney U test). All differences were within a limit of 2% except for the effect in quantitative reasoning reported in other similar tests, where the average score for males was almost 5% higher than for females.

## 6.4 Effects of Age

The overwhelming majority of candidates for the test were aged 19 or below, and the numbers in higher age groups are relatively small, making formal comparisons difficult. However, the scores for each section of the test, and the total scores, are shown below. Because the numbers in the higher age groups are so small, we have not tried to separate out graduate entrants from mature students who are not graduates, nor to investigate any difference in the performance of those with higher academic degrees or professional qualifications.

*Overall scores for all candidates (n=18542):*

| Section | n= | Age ≤19 15051 Mean SD | | Age 20□34 2472 Mean SD | | Age 25□34 809 Mean SD | | Age≥35 143 Mean SD | |
|---|---|---|---|---|---|---|---|---|---|
| Verbal reasoning | | 588 | 92 | 590 | 104 | 587 | 115 | 577 | 124 |
| Quantitative reasoning | | 602 | 80 | 583 | 85 | 571 | 99 | 550 | 104 |
| Abstract reasoning | | 599 | 84 | 590 | 89 | 568 | 97 | 539 | 105 |
| Decision analysis | | 596 | 102 | 589 | 103 | 569 | 108 | 533 | 124 |
| Total scaled score | | 2384 | 259 | 2351 | 287 | 2295 | 329 | 2198 | 372 |

*** – the fall in mean score (and in median score, not shown in the table) with age is highly significant ($p < 0.001$, Kruskal-Wallis test: the data are not normally distributed).

28

## 6.5 Ethnic Background

We investigated the possibility that the test might be more difficult for candidates of particular ethnic or cultural background. The total scores on each section were compared with the candidates' ethnic origin (self-reported, on the UKCAT registration form). Although there are small differences in the total test score between ethnic groups, we have no data to allow us to compare these differences with the candidates' other academic achievements. In addition, as a separate exercise (reported on success rates for each question in the test were correlated with candidates' ethnic origin. Although it is difficult to be absolutely certain in the absence of other data on candidates' academic ability, neither of our investigations suggested that individual ethnic groups would be disadvantaged in the test.

The total scores varied between broad ethnic groups, as follows:

| Broad Ethnic Group | number | Mean Total Score | SD |
|---|---|---|---|
| Asian | 4393 | 2286.0 | 267.4 |
| Black | 800 | 2163.1 | 282.5 |
| Chinese | 770 | 2471.9 | 241.9 |
| Mixed | 640 | 2398.3 | 280.2 |
| White | 10780 | 2428.9 | 244.5 |
| Other | 862 | 2245.1 | 287.1 |
| Not declared | 297 | 2352.9 | 299.7 |

## 6.6 Socio-Economic Correlations

Candidates were asked to report their parents' occupations as part of the registration process for the test. The total mean scores for each broad grouping of parental occupation are shown below: the occupations are those of the highest-scoring (on the National Statistics Socio-economic Classification) or only parent.

| Parental Occupation | number | % | Total mean | SD |
|---|---|---|---|---|
| Traditional professional | 5348 | 28.8 | 2417 | 258 |
| Modern professional | 5678 | 30.6 | 2396 | 263 |
| Senior managers & administrators | 1916 | 10.3 | 2379 | 253 |
| Middle or junior managers | 726 | 3.9 | 2355 | 259 |
| Technical & craft occupations | 848 | 4.6 | 2364 | 271 |
| Clerical & intermediate occupations | 634 | 3.4 | 2346 | 280 |
| Semi-routine manual & service occup'ns | 490 | 2.6 | 2334 | 254 |
| Routine manual & service occupations | 440 | 2.4 | 2261 | 281 |
| Do not know | 917 | 4.9 | 2255 | 279 |
| Never worked | 154 | 0.8 | 2237 | 286 |
| Information withheld | 1391 | 7.5 | 2143 | 298 |
| *Total* | *18542* | *100.0* | | |

## 6.7 Differential Item Function (DIF) Analysis

Pearson Vue undertook a differential item function (DIF) analysis for us, to ensure that there were no questions that showed evidence of particularly disadvantaging candidates of a particular age, sex or ethnicity. Eleven (3.6%) of the scored questions (i.e., of those that

contributed to the test score, and were not unscored "pretest" questions) were found to produce a score that correlated with particular ethnic groups or age groups. As has already been noted, there were differences between ethnic, social and gender groups in the performance of whole sections of the test, but allowing for this background no other questions (apart from the eleven that were identified) showed differences that stood out from the performance of the section as a whole. The bias among the eleven questions pointed in different directions, and it is unlikely that any candidate was disadvantaged by the very small degree of bias represented by these questions.

| Comparison | Verbal | Quant've | Abstract | Decision |
|---|---|---|---|---|
| Female/Male | 0 | 0 | 0 | 0 |
| Age: <20/>34 | 3 | 0 | 0 | 0 |
| White/Black | 1 | 3 | 0 | 2 |
| White/Mixed | 0 | 0 | 0 | 0 |
| White/Other | 0 | 0 | 0 | 1 |
| White/Asian | 0 | 0 | 0 | 1 |
| White/"Withheld" | 0 | 0 | 0 | 0 |

## 6.8 Reliability data

Reliability, or internal consistency (expressed as Cronbach's alpha), varied slightly between the four sections of the test. Not surprisingly, the reliability was highest in those sections composed of questions that had previously been tried in other tests (or that were based on

designs from other tests). The lowest reliability came from the "decision analysis" section, which was developed specifically for the UKCAT, and had not been tried previously. The Test Development Group has noted the lower reliability of this section and, based on the lessons that we learnt from the 2006 test, is working to improve the reliability for future years.

*Reliability data from the 2006 test: Cronbach 's alpha for each section:*

| Section | Reliability |
|---|---|
| Verbal reasoning | .74 |
| Quantitative reasoning | .71 |
| Abstract reasoning | .86 |
| Decision analysis | .58 |

## Correlation of scores between sections of the test

We looked at candidates' scores across the four sections of the test to see how well they correlated with one another – that is, to see whether a candidate who performed well in one section of the test was likely to perform well in another section. The results suggest that there is some correlation, as might be expected, but that a high performance in one section of the test is not automatically associated with a high performance in the other sections, i.e., that some candidates have particular strengths in particular areas, which are not mirrored in the other areas examined by the test.

We originally recommended that the test results should be interpreted as four individual scores: but in practice relatively few schools seem to have used the scores separately except in cases where a wide disparity between scores in different sections was used to call attention to a particular candidate. As experience with the test builds up, it might be sensible for schools to begin to look more closely at the scores for individual sections, rather than considering only the total score for the test.

*Pearson correlation coefficients for marks in the four sections of the test (whole cohort, n=18542):*

| Section | Verbal | Quant've | Abstract |
|---|---|---|---|
| Quantitative | .356*** | | |
| Abstract | .450*** | .371*** | |
| Decision analysis | .382*** | .385*** | .351*** |

*** denotes that the correlation is significant ($p < 0.001$)

## 6.10 Effect of the Test Subtype on the Scores

As noted above, two alternative sets of questions were prepared for each of the four sections of the test, and the question sets were selected at random when a candidate took the test. This produced sixteen possible combinations of questions in the test. Each combination was taken by a similar number of candidates (ranging from 1085 to 1220, mean 1189), and the minimum and maximum scores were:

| | *Mean total score* |
|---|---|
| Highest-scoring combination | 2398 |
| Lowest-scoring combination | 2352‡ |

‡ – One-way analysis of variance showed a significant difference between groups (F=2.87, $p < 0.001$).

## 6.11 Effect of the UKCAT on Other Schools

In a self-reported survey (undertaken by Drs Jane Adam and Lyndon Cabot on behalf of the Medical Schools' Council and the UKCAT board), many schools in the UKCAT consortium reported a small fall in numbers of applicants in the 2006 cohort (i.e., those applying in 2006 for entry in 2007 or later: the first year in which the UKCAT test was used), compared with the numbers in the previous year, and some of those schools that were not using any admissions test reported small increases in applicant numbers – in one case, a substantial increase.

In 2006, total applications to medicine across the UK were down by 3.8% compared with the previous year; applications to dentistry were up by 3.1% (figures from UCAS).

# 7. Future Developments

## 7.1 Section : Non-cognitive testing

UKCAT was designed to not only increase discrimination between highly qualified candidates but to also select students on the basis of the most appropriate characteristics. To this end we included as a trial four behavioural tests in the 2007 testing for the constructs of empathy, integrity and/or robustness. The purpose of this trialling period is to evaluate the psychometric characteristics of each test and collect data for future validity studies before one or all tests are used to select for entry to medical or dental school.

Currently the planned four instruments are:

- Interpersonal traits (ITQ 100) – Narcissism, Aloofness, Confidence and Empathy

- Interpersonal valves (IVQ49) – Measure of ethical orientation

- Combined and abridged ITQ50 and IVQ33)

- MEARS (Managing Emotions and Resilience Scales) – Three subscales: cognitive, behavioural, emotional

## 7.2 Research Programme

The key tasks in the initial phase of the research programme are:

1. to establish links with UCAS and the Universities

2. to relate UKCAT results to UCAS data, academic and non-academic

3. to assess the impact of UKCAT on admissions processes

4. to relate UKCAT results to progress at medical or dental school

5. to relate UKCAT results to progress as a postgraduate

# 8. Management of the UKCAT

## 8.1 Committee Structure

The Consortium is formally constituted as a limited company, with the participating medical and dental schools as its shareholders. It is managed by a Board of Directors, which has formed three subcommittees to supervise various aspects of the UKCAT: test development,

test delivery and research. Each subcommittee is chaired by a member of the Board; the subcommittees are assisted by external advisers who attend meetings to provide expert advice on the subcommittee's area of interest.

## 8.2 Consortium

Each participating medical school has nominated one representative to the UKCAT consortium: in most cases, this is the school's admissions dean or an academic with experience of the admissions process. The Consortium meets together twice in each year.

As of March 2008 the members of the consortium are:

University of Aberdeen

Barts and The London School of Medicine and Dentistry

Brighton and Sussex Medical School

Cardiff University

University of Dundee

University of Durham

University of East Anglia

University of Edinburgh

University of Glasgow

Hull York Medical School

Imperial College London*

Keele University

King's College London

University of Leeds

University of Leicester

University of Manchester

University of Newcastle

University of Nottingham

University of Oxford*

Peninsula Medical School

Queen's University Belfast

University of Sheffield

University of Southampton

University of St Andrew's

St George's, University of London

University of Warwick

* – these schools require the UKCAT for their graduate-entry courses only, and use another entrance test (the BMAT, in both cases) for their undergraduate courses.

## 8.3 Board

In 2006, the Board consisted of eight members elected by the Consortium from among their own number, plus four members appointed by the Council of Heads and Deans of Dental Schools and by the Medical Schools Council. Each place on the Board is open for election every three years (the elections are staggered, to avoid a whole-sale change in the Board at one time). The current membership of the Board is:

Professor Ian Johnson, University of Nottingham (Chairman)

Dr Jane Adam, Hull York Medical School

Mr Martyn Annis, King's College London

Professor Barbara Chadwick, University of Wales* *(from June 2007)*

Dr Paul Dennis, University of Oxford

Professor Mary Ann Lumsden, University of Glasgow

Dr Sandra Nicholson, Barts and The London School of Medicine and Dentistry

Dr Katie Petty-Saphon, Medical Schools Council**

Mr Nigel Siesage, University of Leicester *(from December 2006)*

Dr Chris Stephens, University of Southampton

Professor Sir John Tooke, Peninsula Medical School**

Professor Tony Weetman, University of Sheffield**

\* – Nominated by the Council of Heads and Deans of Dental Schools

\** – Nominated by the Medical Schools Council (formerly the Council of Heads of Medical Schools)

## 8.4 Subcommittees of the Board

The Board has overall charge of the UKCAT, but chooses to delegate some work in specific areas to three subcommittees: the Test Delivery Group, which manages the physical process of delivering the test to the candidates (including registration, the web site, management of the bursaries and the distribution of results to the participating medical and dental schools); the Test Development Group, which manages the development of new questions for the test, the development and publication of practice questions, and the marking of the test; and the Research Working Group, which has an overview of research strategy and is charged with managing the database of test results and student progress through medical and dental schools.

### Test Delivery

The Test Delivery Group is responsible for the logistics of delivering the test: ensuring that the test is ready by the due date, that the process of administering the test (including the process of registration) is satisfactory, and that there are sufficient places available for candidates who wish to take the test. This group also has an overview of the UKCAT web site and of the management of the bursary scheme. Distribution of test results to medical and dental schools also falls within the remit of this committee.

The membership of the test delivery group in 2006 was:

Mr Martyn Annis, King's College London (Chairman)

Dr Jane Adam, Hull York Medical School

Ms Caroline Persaud, St George's Hospital Medical School

Mr Nigel Siesage, University of Leicester

and members of the PearsonVue team

### Test Development

The Test Development Group is responsible for the strategic direction of the test: the form and direction of the questions, the development of the bank of questions used in the test, and the mark schemes for the test.

The membership of the test delivery group in 2006 was:

Dr Sandra Nicholson, Barts and The London School of Medicine and Dentistry (Chairman)

Professor Mary Ann Lumsden, University of Glasgow

Dr David Heylings, University of East Anglia

Dr Lyndon Cabot, King's College London

and members of the PearsonVue team

## Research Working Group

The Research Working Group is responsible for the co-ordination of follow-up studies on the new test: design of studies to monitor students as they pass through their medical courses, and to establish the predictive value of the UKCAT.

The membership of the group in 2006 was:

Professor Mary Ann Lumsden, University of Glasgow (Chairman)

Dr Jane Adam, Hull York Medical School

Professor Martin Bland, Hull York Medical School

Dr Lyndon Cabot, King's College London Dental Institute

Professor Barbara Chadwick, University of Wales

Dr Paul Dennis, University of Oxford

Dr Jon Dowell, University of Dundee

Professor David James, University of Nottingham

Professor Ian Johnson, University of Nottingham

Professor Chris McManus, University College London

Dr Sandra Nicholson, Barts and The London School of Medicine and Dentistry

In March 2008 the Board took the decision to restructure the management of the research programme; details of the new management group will be made available to Consortium members as soon as they are agreed by the Board.

# 9. Financial Statement

The statement below represents a summary of the cash flow of the UKCAT Company from inception until the close of business on 31 March 2007. Figures are rounded to the nearest £1000.

|  |  | £,000 |
|---|---|---|
| **The Test** | | |
| Testing fees | | |
| UK and rest of EU | 1,026 | |
| Rest of world | 105 | |
| **Turnover** | **1,132** | |
| | | |
| Testing provider's charges: | | |
| Testing | (1,006) | |
| Other | (44) | |
| **Cost of sales** | (1,050) | |
| | | |
| Surplus from testing | | 82 |
| | | |
| Contributions from members | | 130 |
| | | |
| **Administration:** | | |
| Office and administration charges | (22) | |
| Hotels, travel and subsistence | (6) | |
| Printing and stationery | (2) | |
| Legal fees | (43) | |
| Accountancy and taxation fees | (3) | |
| Sundry expenses | (0) | |
| Insurance | (2) | |
| Bank charges | (0) | |
| Admin expenses | | (79) |
| **Interest receivable:** | | |
| Bank interest | 3 | |
| Other interest | 13 | |
| Total interest receivable | | 15 |
| | | |
| **SURPLUS BEFORE TAX** | | **148** |
| | | |
| Tax | | (4) |
| **SURPLUS AFTER TAX** | | 144 |

The surplus will be used to pay for the research activities of the UKCAT, and to support the operation of the bursary scheme.