THINKING SKILLS ASSESSMENTS FOR HIGH-STAKES SELECTION

Paul Crump

Assessment Group Manager, Cambridge Assessment Admissions Testing

crump.p@cambridgeassessment.org.uk

Thomas Iveson-Malcolm

Senior Assessment Manager**,** Cambridge Assessment Admissions Testing

iveson-malcolm.t@cambridgeassessment.org.uk

Abstract

This paper examines the rationale for the use of admissions tests for high-stakes selection in higher education. It discusses the use of critical thinking and problem solving questions in such tests, and sets out Cambridge Assessment Admission Testing's conception of 'thinking skills' with specific reference to the Thinking Skills Assessment (TSA) and Section 1 of the Biomedical Admissions Test (BMAT).The paper concludes by considering how thinking skills content can add value to admissions procedures, helping to ensure fairness and reliability.

Introduction

Cambridge Assessment Admissions Testing is a department within Cambridge Assessment, itself a not-for-profit non-teaching department of the University of Cambridge.

As its name suggests, Cambridge Assessment Admissions Testing produces tests used by a number of institutions in the UK and internationally for admissions purposes. We focus mainly on the production of Mathematics, Science and Thinking Skills content and our tests include the Biomedical Admissions Test (BMAT), the Thinking Skills Assessment (TSA), a suite of admissions tests for our own university, Cambridge, as well as other tests such as STEP. We also produce 'bespoke' assessments for Ministries of Education: most recently we have worked with institutions in Italy, Kazakhstan and Uzbekistan.

Why use an admissions test?

Universities now have no shortage of information about their prospective students on which to base their admissions decisions, and also that the assessment demands placed on a student wanting to go into UK higher education are already rather onerous. Nevertheless, over the last couple of decades a growing number of university departments, especially those in traditionally elite institutions, have made the decision to introduce additional examinations into their admissions requirements, usually taken at the first stage of application. Why is this?

The most straightforward and often-cited reason for using admissions tests is to allow institutions to discriminate between applicants all of whom may be predicted to, and most of whom will, attain top grades in their A levels – in 2018, 26.2% of A levels taken in England resulted in a grade of either A or A* (OFQUAL, 2019). This is a particular issue for the most popular courses at elite institutions – a Cambridge college with a dozen places for Medicine, for example, might typically get between fifty and a hundred applications per year (University of Cambridge, 2018), most of which will be from students with extremely strong academic records and straight-A predictions in their A levels. An additional admissions test may therefore help them to pick out the really outstanding students from among the merely very good.

Another common reason for introducing an admissions test is to allow institutions to have a single point of reference when comparing candidates who may have a range of different qualifications. This is most obviously the case when considering applications from overseas students, which is particularly relevant in the UK where the number of international students has more than doubled in twenty years (UKCISA, 2019). It also applies even when considering only domestic applicants, however: there are a range of alternatives to A levels even in the UK (such as the Scottish Highers and International Baccalaureate), and the equivalency of grades in these assessments are often the subject of some controversy (Gill, 2016).

Even where the qualifications taken and grades predicted are identical, institutions may not feel that they are necessarily the best measure of a student's potential to succeed at university. Predicted grades themselves are notoriously unreliable (Wyness, 2016), and the recent reforms to UK A levels were informed partly by a perception that secondary schools had become 'exam factories' where shallow, 'learn and forget' learning was prioritised over genuine intellectual development (Long, 2017). Admissions tests, which often focus on

students' cognitive or creative abilities rather than the extent of their memorised knowledge, may often give a truer picture of who the most potentially outstanding students are.

Finally, the role of universities in facilitating social mobility now receives considerable scrutiny in the UK. A 2018 report found that pupils from independent fee-paying schools are seven times more likely to get a place at Oxford or Cambridge, and four times more likely to get a place at a high-ranking (Russell Group) university, than students from state schools (Montacute & Cullinane, 2018) This is partly because of A level results, but there are other factors at play: the same report found that, even when comparing schools with very similar results, the pupils from fee-paying schools were more likely to apply, and more likely to be accepted if they did apply. Both in the academic literature and in the press, there is a well-established concern that the admissions processes of Oxford and Cambridge in particular 'further and unjustly increase the cumulative advantages children from high social class origin and privately educated students have accumulated earlier in life' (Zimdars, 2010).

How can admissions tests help to mitigate this? Firstly, they can create more of a level playing field for comparing candidates from different backgrounds when compared to the other methods being used. A levels, for example, are known to over-predict the performance of fee-paying school students and under-predict those of state school students (Hoare & Johnston, 2010). Traditional face-to-face interviews have been found to be unreliable (Patterson et al., 2015), and may be subject to homophilic bias on the part of interviewers (Zimdars, ibid). And personal statements from privately educated candidates are known to be markedly different from those of state school candidates with equal educational attainment: they are better written, and draw on a broader range of experience (Jones, 2013)

One factor which is likely to be instrumental in the bias shown in each of the above factors is preparation and coaching – candidates from more advantaged backgrounds will simply have access to greater resources, both inside and outside school, to help them prepare every aspect of their application. Indeed this is likely to be the case for almost *any* hurdle placed in front of students where they have an incentive to do well, and we should not assume that admissions tests in general are free from preparation effects which may work against candidates from less advantaged backgrounds. So it is important when introducing an admissions test to ensure that it does not unfairly advantage candidates who have access to private tuition, intensive preparation classes, and greater parental support.

It is against this backdrop that Cambridge Assessment Admissions Testing developed a suite of admissions tests which focus on candidates' critical thinking and problem solving abilities, known collectively as 'Thinking Skills'. In the next section we will set out what we mean by Thinking Skills, and describe in more detail the subskills we assess. In the final section of this paper we will discuss how Cambridge Assessment Admissions Testing's Thinking Skills assessments help HE institutions to address some of the issues described above and make fair, evidence-based admissions decisions.

What are 'thinking skills'?

Cambridge Assessment Admissions Testing's conception of thinking skills is based on two complementary sub-skills: Critical Thinking and Problem Solving – essentially, 'verbal' and 'numerical' reasoning skills which, according to Fisher (2005) are seen as:

- teachable skills/abilities
- skills which can be developed through special approaches to subject matter or through stand-alone courses
- general, transferable academic skills which are vital to successful university work

Critical Thinking is the analysis and evaluation of arguments. It has been defined as 'the analytical thinking which underlies all rational thought and enquiry', (Black: 2008) and focuses explicitly on the processes involved in being rational. These processes, as assessed in the Thinking Skills Assessment, comprise:

Summarising the main conclusion of an argument
An argument will consist of reasons (or 'premises') which support a conclusion – and in some cases, an argument may contain an intermediate conclusion which in turn supports the main conclusion. The task in this type of question is to answer: 'What is the main message which this passage is trying to get me to accept?'

Drawing a conclusion
A related, but distinct, skill is identifying a conclusion which is not contained within a given argument, but which logically follows from it. Candidates must decide whether the information presented to them in the passage gives good reason to accept a statement or not.

Identifying assumptions
An assumption is something not stated in the argument, but which needs to be taken for granted in order for the conclusion to be logically drawn.

An example of this question type is given in the appendix.

Assessing the impact of additional evidence
This item type requires candidates to grasp the main conclusion of an argument, and then consider which statement (A-E) would most strengthen or weaken it.

Identifying reasoning errors
This requires candidates to understand why the conclusion of an argument does not logically follow from its premises. Candidates need to understand what the argument is trying to establish, and how it tries to establish it, in order to identify the reasoning flaw.

Matching arguments
This item type asks candidates to analyse the logical structure of an argument, and then identify another argument with the same structure.

Applying principles
This requires candidates to identify general recommendation in an argument which, in the passage, will be applied to just one case but which could also be applied to other cases. They must then choose from statements (A-E) which one best illustrates this principle.

Problem Solving has been defined as 'a candidate's ability to analyse numerical and graphical information, which is based on real life situations, and apply the right numerical techniques to find new information or derive solutions'. In Cambridge Assessment Admissions Testing's admissions tests, this is broken down into three subskills as follows:

Relevant Selection
Often a real-world problem will be overloaded with information, much of which is unimportant, so the task with this type of question is to quickly home in on the information required to solve the problem you are presented with. Information will typically be presented in a table, graph, or chart.

Finding Procedures
This item type requires candidates to find a methodology to solve a problem, where no obvious procedure is apparent. The 'problem' will typically involve three or four numbers to be operated on.

An example of this item type is given in the appendix.

Identifying Similarity
In this type of question, candidates are presented with a set of information, often in the form of a table, chart or graph, and asked to identify if another set of data has a similar structure. Spatial reasoning may also be a feature of this item type: candidates may be presented with a three-dimensional shape, for example, and asked to identify another view of the shape from a different angle.

Why use thinking skills for selection?
Having described the rationale for admissions tests, and set out in some detail the thinking skills tests offered by Cambridge Assessment Admissions Testing, the next question to be addressed is 'Why are thinking skills especially suited for admissions purposes?'. Why not simply construct subject-specific admissions tests that test similar skills and knowledge to that assessed in school-leaving examinations, but at a higher level, or with a different approach? Subject-specific tests certainly have their place, and Cambridge Assessment Admissions Testing offers a number of such tests (e.g. STEP Mathematics, the English Literature Admissions Test, BMAT Section 2), which apply a problem-solving approach to assessment which has a distinctly different emphasis from most standard summative assessments. There are a number of reasons, however, why the inclusion of thinking skills in admissions tests can help in ensuring a good match between course and candidate.

Firstly, Critical Thinking and Problem Solving are thought to be highly important skills for success in higher education across a wide range of academic subjects. They are also seen as important skills to be cultivated while studying at undergraduate-level. The AHELO (Assessment of Higher Education Learning Outcomes) study carried out by the Organisation for Economic Co-operation and Development in 2011 ranked critical thinking and problem solving among the top five most important learning outcomes from university-level study.

A key question for any admissions test is 'Does it accurately predict future performance?' Tests of thinking skills developed by Cambridge Assessment Admissions Testing have been shown to be good predictors of performance in a range of subjects. It should be noted that

there are a number of confounding factors that can make investigating the predictive validity of admissions tests problematic. Admissions test scores are usually correlated with end of first year and/or end of course examination results, but many factors can impact on a student's performance in the long interval between these assessments. In addition, there is the problem of restriction of range: outcome variables are only available for successful applicants, who are a subset of the applicants who completed the selection assessment. These caveats notwithstanding, in a 2009 study (Emery and Bell 2009) BMAT was found to make 'a significant and unique' contribution to predicting performance in medical study. It is perhaps unsurprising that Section 2 ('Scientific Knowledge and Applications') correlated more strongly with performance in pre-clinical courses (it is much more closely aligned with course content). However, Section 1, essentially a thinking skills paper, showed correlations of between 0.2 and 0.5 with performance across a range of first year component courses in Medicine at the University of Cambridge. Research carried out by the University of Cambridge (Partington, 2011) also found significant correlations between performance on problem solving questions and on-course performance in the Cambridge Engineering and Economics triposes.

Another advantage of thinking skills from an assessment perspective is that they are curriculum-agnostic: they do not test a body of knowledge. In an era of increasing internationalisation of the Higher Education sector, this means tests can be more easily be set for cohorts from a wide range of educational backgrounds (and give admissions tutors a single piece of evidence applicable to the entire gathered field of applicants). In addition, thinking skills are not explicitly tested in traditional school-leaving examinations, so tests such as TSA which sit a alongside A levels (or equivalent school-leaving qualifications) provide new information about an applicant.

Thirdly, critical thinking and problem solving are thought to be relatively stable traits (van Gelder, 2005; Fisher, 2005), meaning that, while it is possible to get better at them over time with practice, a candidate's ability is unlikely to change dramatically over a short period of time, for example as a result of doing an intensive course in preparation for an exam. A candidate who is a strong critical thinker and problem solver today will still be one next year, and one who struggles today will not suddenly improve drastically. There is no body of knowledge that can be memorised over a two-week period, regurgitated during the exam, and then forgotten just as quickly.

This has the advantage of making thinking skills a fairly reliable predictor of candidate's future academic performance, but it also means that well-designed thinking skills assessments are much less susceptible to the effects of intensive preparation than most other measures typically used in the admissions process. A 2017 study into the effect of different preparation strategies on scores in Cambridge Assessment Admissions Testing's BMAT exam found that the only strategy associated with a higher score in BMAT Section 1 (which tests Thinking Skills) was the use of practice and past papers under timed conditions (McElwee et al., 2017), which can be done for free by anyone with an internet connection. Although a significant proportion of candidates reported having attended paid-for preparation courses or having help from their school – unsurprisingly, a far higher proportion of private than state school students – neither of these strategies (nor indeed the fact of having attended a private school in itself) was actually associated with a higher test score. This is obviously good news for universities wanting to select

the most promising candidates for their courses whilst at the same time advancing the agenda of widening participation in higher education.

Ensuring fairness and reliability

The specific form of thinking skills test used by Cambridge Assessment Admissions Testing (and mirrored by other major testing organisations) – the use of a large number of short, discrete, single-answer, objectively marked, multiple-choice questions – also brings a number of advantages in terms of the fairness and reliability of the test.

A large number of short, discrete questions helps to ensure the statistical robustness of the tests – Cambridge Assessment Admissions Testing's 50 item TSA, for example, typically reports a test alpha of around 0.8, suggesting a high level of internal consistency of measurement in the items (especially given that the test contains seven distinct item types across two sub-constructs). But it also allows the coverage of the test, in terms of the scenarios presented in the items, to be sufficiently broad that an individual candidate is unlikely to be unfairly advantaged, or disadvantaged, by a particularly familiar or unfamiliar topic, or form of problem, coming up in the test.

The use of objectively-marked multiple-choice items also offers some built-in fairness and reliability, especially in the context of a high-stakes admissions test. Whilst 'constructed response' test items – those which require a candidate to write a paragraph in response to a prompt, for example – may be thought to offer candidates more opportunity to demonstrate individuality and creativity, they also need to be scored. Inevitably, this introduces an element of 'the luck of the draw' - a candidate's score is determined not just by the strength of their performance, but also by which examiner happens to mark their responses, what the examiner's mood is, and the extent to which the content or style of the response match up with the examiner's personal tastes – in this sense, they suffer from the same drawbacks as the use of interviews and personal statements, as mentioned above. Even extensive standardisation training and rigorous double-marking procedures have been shown to have a limited effect in mitigating against these factors (Newton, 1996; Bloxham, 2009; Baird et al., 2004).

Of course, while a multiple-choice test may have some inherent reliability (especially where items are pretested to establish their measurement characteristics, as is the case with Cambridge Assessment Admissions Testing's TSA), steps must still be take to ensure that the test as a whole does not introduce systematic bias on the basis of gender, race, or other construct-irrelevant factors. In some cases this can be done through research based test-design – the use of positively-scored questions in TSA and BMAT is designed to avoid possible gender effects stemming from differences in risk aversion, for example (Devine et al., 2017). In other cases there is natural overlap with good practice generally – fostering an adequately diverse range of backgrounds in the pool of writers who produce test questions, for example. Finally, the use of Differential Item Functioning (DIF) analyses can help to identify questions, and question types, which may introduce bias on the basis of gender, ethnicity, or socio-economic background.

References

Baird, J., Greatorex, J., & Bell, J.F. (2004) What makes marking reliable? Experiments with UK examinations. Assessment in Education: Principles, Policy & Practice, 11:3, pp.331-348

Black, B. (2008). Critical Thinking – a definition and taxonomy for Cambridge Assessment: Supporting validity arguments about Critical Thinking assessments administered by Cambridge Assessment

Bloxham, S. (2009). Marking and moderation in the UK: false assumptions and wasted resources. Assessment & Evaluation in Higher Education, 34:2, pp.209-220

Devine, A., Taylor, L., Cross, B. (2017). 'The biomedical school applicant: Considering the test taker in test development and research', in Cheung, K., McElwee, S., & Emery, J. (eds) Applying the socio-cognitive to the BioMedical Admissions Test, Cambridge: CUP, pp.181-215

Emery, J. L. and Bell, J. F. (2009). The Predictive Validity of the BioMedical Admissions Test for pre-clinical examination performance, Medical Education 46

Fisher, A (2005). 'Thinking Skills' and Admission to Higher Education

Gill, T. (2016). Assessing the equivalencies of the UCAS tariff for different qualifications. Research Matters, 21, pp.16-22. Available online: https://www.cambridgeassessment.org.uk/Images/374776-assessing-the-equivalencies-of-the-ucas-tariff-for-different-qualifications.pdf (accessed 27.8.2019)

Hoare, A. & Johnston, R. (2011). Widening participation through admissions policy – a British case study of school and university performance. Studies in Higher Education, 36:1, pp. 21-41

Jones, S. (2013). "Ensure That You Stand Out from the Crowd": A Corpus-Based Analysis of Personal Statements according to Applicants' School Type. Comparative Education Review, 77:3, pp.397-423

Long, R. (2017). GCSE, AS and A level reform (England). House of Commons briefing paper 06962. Available online: https://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06962#fullreport (accessed 27.8.2019)

McElwee, S., Fyfe, M., & Grant, K. (2017) 'The consequences of biomedical admissions testing on individuals, institutions and society', in Cheung, K., McElwee, S., & Emery, J. (eds) Applying the socio-cognitive to the BioMedical Admissions Test, Cambridge: CUP, pp.181-215

Montacute, R. & Cullinane, C. (2018). Access to Advantage. The influence of schools and place on admissions to top universities. The Sutton Trust. Available online: https://www.suttontrust.com/wp-content/uploads/2018/12/AccesstoAdvantage-2018.pdf (accessed 27.8.2019)

Newton, P.E. (1996). The Reliability of Marking of General Certificate of Secondary Education Scripts: mathematics and English. British Educational Research Journal, 22:4, pp. 405-420

OFQUAL (2019). Guide to AS and A level results for England, 2019. Available online: https://www.gov.uk/government/news/guide-to-as-and-a-level-results-for-england-2019 (accessed 27.8.2019)

Partington, R. (2011). Predictive Effectiveness of Metrics in Admission to the University of Cambridge. Report for Admissions Working Research Party (AWRP).

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousins, F. & Cleland, J. (2015). How effective are selection methods in medical education? A systematic review. Medical Education, 50:1, pp.36-60

UKCISA (2019). International student statistics: UK higher education. Available online: https://www.ukcisa.org.uk/Research--Policy/Statistics/International-student-statistics-UK-higher-education (accessed 27.8.2019)

University of Cambridge (2018). Undergraduate application statistics. Available online: https://www.undergraduate.study.cam.ac.uk/apply/statistics (accessed 27.8.2019)

Wyness, G. (2016). Predicted grades: accuracy and impact. A report for University and College Union. Available online: https://www.ucu.org.uk/media/8409/Predicted-grades-accuracy-and-impact-Dec-16/pdf/Predicted_grades_report_Dec2016.pdf (accessed 27.8.2019)

Zimdars, A. (2010). Fairness and undergraduate admission: a qualitative exploration of admissions choices at the University of Oxford. Oxford Review of Education, 36:3, pp.307-323

APPENDIX

Example item 1: Assessing the Impact of Additional Evidence

Polar bears in captivity frequently engage in obsessive patterns of behaviour, pacing back and forth on the same spot, swinging their heads from side to side, and other signs of stress. They do this even when their living areas are quite spacious. What this shows is that conditions of captivity are not a satisfactory substitute for the natural environment of the polar bear species.

Which of the following, if true, would most weaken the above argument?

A Polar bears are especially ill-suited to a life in captivity.

B Many polar bears in the wild engage in obsessive patterns of behaviour.

C Polar bears in captivity are much better fed than those living in the wild.

D Polar bears in the wild cover many miles a day when they are hunting for food.

The answer is **B**. The conclusion of the argument is that the obsessive behaviour of polar bears in zoos shows that conditions of captivity are not a satisfactory substitute for the polar bear's natural environment. But if **B** is true, that is, if polar bears in the wild behave in the same way as those in captivity, then the behaviour of those in captivity cannot be taken as good evidence that the conditions of captivity are unsatisfactory.

Example item 2: Finding Procedures

Sometimes you will find that even if you have selected all of the relevant information, no solution presents itself. You then have to find a method or procedure which you can use to generate a solution. Typically you will have three or four numbers which have to be operated on. This aspect of Problem Solving is called Finding Procedures.

Three thermometers are each accurate to within 2 degrees above or below the temperature they actually read. One reads 7°, one reads 9° and one reads 10°.

What is the minimum range in which the true temperature lies?

A 5° - 12°

B 7° - 9°

C 8° - 10°

D 8° - 9°

E 7° - 10°

The answer is **D**. The method here is to search for the acceptable highest and lowest temperatures for the conditions to be met, realising that the middle value is irrelevant. As one reads 7°, the temperature cannot be above 9° and, as another reads 10°, the temperature cannot be below 8°. This is given by **D**.