education & training

CATHERINE M. THOMPSON

Will the CASC stand the test? A review and critical evaluation of the new MRCPsych clinical examination†

SUMMARY

The aim of this paper is to review and critically evaluate the Clinical Assessment of Skills and Competencies (CASC) examination, introduced in June 2008. The history of clinical membership examinations and rationale for the new examination are considered. The development and implementation of the examination is described by means of a four-step model of assessment. Finally, the paper critically evaluates the assessment with reference to validity, reliability, practicability and effect on learner.

The Clinical Assessment of Skills and Competencies (CASC) is a new clinical examination which was introduced formally as a summative assessment in June 2008. The CASC assesses postgraduate trainees in psychiatry and takes place after the successful completion of three written papers, with College membership (MRCPsych) being dependent on achieving an agreed standard of competency in all four exams. The prime aim of the CASC is to set a standard that determines whether candidates are suitable to progress to higher professional training. In addition, the MRCPsych qualification is considered to be an indicator of professional competence in the clinical practice of psychiatry.

In the light of the General Medical Council (GMC) guide to *Good Medical Practice* (General Medical Council, 2006), the curriculum for basic specialist training has been revised to illustrate the precise knowledge and competencies required to graduate to higher professional training (Royal College of Psychiatrists, 2006) and the content of the examination is tailored to this curriculum. The CASC is a new examination format but it is based largely on the recognised Observed Structured Clinical Examination (OSCE) format introduced in 1975 (Harden & Gleeson, 1979), now widely regarded as a reliable and valid performance-based assessment tool (Hodges *et al*, 1998).

Historically, there have been difficulties with assessing the clinical skills required for the MRCPsych. The traditional method was assessment via a long case at part one MRCPsych level (designed to assess history taking, examination and preliminary diagnostic skills) and then a further long case and a set of patient management problems followed at part two MRCPsych level (designed to assess more complex diagnostic and management skills). The part one long case was replaced by an OSCE format in 2003 and reviews of validity and reliability are generally favourable (Tyrer & Oyebode, 2004). The

revision of the part two clinical examination has been less easy – evidence suggests that OSCE examinations when marked by checklist are not suitable for the assessment of more advanced psychiatric skills (Hodges *et al*, 1999). The long case and PMPs were maintained on the basis that they continued to offer a superior content validity to an OSCE, owing to the complexity of a complete encounter with a real patient and the flexibility allowed to examiners in directing the assessment.

Medicine continues to become increasingly subject to more rigorous scrutiny of its assessment processes which must appear fair and transparent at all times. It is arguable that the traditional long case and patient management problems do not stand up to this. The scrutiny is in part caused by changes in medical training emphasising competency-based assessment, as driven by Modernising Medical Careers and the introduction of the external review body – the Postgraduate Medical Education and Training Board.

It was therefore agreed that the two clinical examinations should be merged into one exam with an aim of assessing those skills assessed in the part one OSCE, coupled with the more complex management skills and diagnostic skills required of advanced trainees, in one objective and structured complete assessment to be known as the Clinical Assessment of Skills and Competencies – CASC.

Assessment method

In order to look in further depth at the chosen assessment method, it is helpful to consider the process as a four-step cycle, as described by Fowell *et al* (1999; Fig. 1).

Step 1: Planning and preparation

The CASC assessment has been designed based on a traditional OSCE format. It was designed to assess
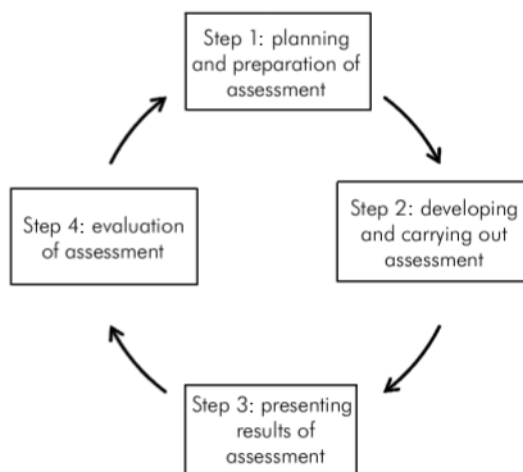
Fig. 1. The assessment cycle. Reproduced with permission from Fowell et al, 1999.

knowledge, skills and attitudes as outlined in the College curriculum based on Bloom's well-established taxonomy (Bloom, 1971) and guided by the GMC principles of good medical practice.

Miller's (1990) hierarchy of competencies informs on the choice of the CASC as an assessment method through defining assessment of competency as four hierarchical concepts: knows, knows how, shows how, and does. Traditionally, OSCEs are perceived to assess at the 'show-how' level, with preceding levels of 'knows-how' and 'knows' being assessed by written examinations. The CASC aims to assess at all these levels, combining the assessment of cognition and behaviour in more complex situations than a purely skills-based OSCE. As we cannot assess the 'does' part except '*in vivo*', it would appear that a clinical structured examination is one appropriate choice of assessment tool and the CASC aims to be more reliable and valid than its predecessors, the long case and the patient management problems.

Step 2: Developing and carrying out assessments

The CASC comprises two circuits – the first circuit consists of eight individual stations of 7 min each with 1 min preparation time, and the second circuit consists of four pairs of linked stations, each station lasting 10 min with 2 min of preparation time. (As part of the phased introduction of the new examination, candidates in June 2008 were required to complete only one circuit which consisted of two linked pairs). The College has only published one example of a linked station scenario but it has been assumed by some that the first station is equivalent to the old OSCE part one exam and the second related station assesses skills at a higher level needed to obtain full membership. It remains to be seen as to whether this assumption is correct.

The stations consist of several elements: there is a construct against which the examiner assesses candidates' performance, instructions to candidates, instructions to the role player and mark sheets (examples are available from the College at www.rcpsych.ac.uk/exams/about/mrcpsychcasc.aspx). The College has also produced a blueprint for the CASC, which attempts to match core objectives to be assessed (each with a weighted percentage) with content relating to five major subspecialties (Table 1).

Step 3: Presenting the results

The pass mark for the summative CASC is based on achieving a predetermined competency to progress to higher level psychiatric training and as such it is criterion-referenced. There should be two examiners for each station and it is clear that there will be no crossover between the patient and marking examiner roles, as had occurred previously. A global marking scheme is used, in preference to a checklist, with pass, borderline pass,

Table 1. The Royal College of Psychiatrists' blueprint for the CASC exam 2008 (www.rcpsych.ac.uk/pdf/MRCPsych%20CASC%20Blueprint%202.pdf)							
	General adult psychiatry[4]	Old age psychiatry	CAMHS	Learning disability psychiatry	Psychotherapy	Forensic psychiatry	%[5]
History[1]	x	x	x	x	x	x	30–40
Mental State Examination	x	x	x	x		x	30–40
Risk assessment[2]	x	x	x	x		x	15–30
Cognitive examination	x	x					10–20
Physical examination	x	x					5–15
Case discussion[3]	x	x	x	x	x	x	15–30
Difficult communication	x	x	x	x	x	x	5–15

CAMHS, child and adolescent mental health services.

1. May include obtaining collateral history as well as taking a history directly from the 'patient'.

2. May include assessment of capacity.

3. Case discussion can be with medical and paramedical professionals as well as lay people including the patient, family, carers and others professionals. It also includes skills related to psychotherapy formulation (cognitive and dynamic) and discussion of treatment.

4. Includes all subspecialties.

5. There will be minor variation in the percentage of stations for each skill area according to how skills are represented in each station. This is due to some station constructs assessing more than one skill (e.g. risk assessment requiring a mental state examination or difficult communication occurring within the context of a case discussion).

146

borderline fail or fail options. For each station, comments are expected from examiners for each given construct to justify their overall mark.

Step 4: Evaluation of the assessment

Using the information from the previous stages of assessment, an evaluation can be made on the basis of the following criteria: validity of assessment, its reliability, practicability and effect on the learner.

Validity

The face validity of the CASC appears to be good and the OSCE-type examinations are a well-accepted measure of clinical skills in medicine (Hodges *et al*, 1998). The content validity also appears to hold up and in its favour the blueprint samples a broad spectrum of the curriculum and as such can be expected to provide a more valid representation of breadth of knowledge than a long-case scenario; the CASC appears to be appropriate for testing the knowledge required of a part one trainee. It is when considering the depth of knowledge required by a more advanced trainee that the content validity is called into question – can a 10-minute station assess the complexity of knowledge and skills required in practice? In an attempt to address this issue, the College has modified the traditional OSCE structure into the CASC, with an emphasis on assessing knowledge and skills further by a paired station system which combines element of history taking with more complex management and diagnostic scenarios. They have recognised that the checklist approach is not appropriate for assessing higher-level trainees (Hodges *et al*, 1999) and have employed a global marking scheme to address this. Global marking schemes allow expert examiners to give some flexibility in their marking approach and emphasise key tasks, there is some room for subjectivity but this is not increased compared with long-case marking techniques. The global marking scheme may improve construct validity but it should be noted that the 'halo effect' is increased by a global marking scale.

Concurrent and predictive validity are more difficult to comment on due to the novelty of the CASC. One suggestion would be to compare candidates' mock-CASC results with other assessments and in particular with workplace-based assessments. Does the CASC really measure the theoretical construct it is designed to measure? All examiners are provided with a construct for each station. Its purpose is 'to define what the station is set out to assess in such a way that the examiner is clear as to what constitutes a competent performance. These have a standardised format with elements in common between stations of a similar type' (www.rcpsych. ac.uk/pdf/CASC%20Guide%20for%20Candidates%20 UPDATED%206%20August%203.pdf). The construct is open to interpretation and it is stressed that where a checklist is not used as in the CASC, examiners must be expert with CASC-specific training in order to promote construct validity.

Reliability

The reliability of the CASC is dependent on many variables which can interplay to affect it, e.g. the setting, the scenario, the patient, the examiner and the candidate. As the CASC is a new exam, its test–retest reliability is yet to be established. However, the similar OSCE format has proven to be repeatable with similar results and is known to be reliable overall (Hodges *et al*, 1997).

Increasing the assessment testing time increases reliability, as does assessing candidates over frequent periods of time (Newble & Swanson, 1988). It could be questioned as to whether assessment over twelve 10-minute stations is sufficient to give reliable results and whether condensing three separate clinical examinations undertaken over 3 years into one 2-hour exam may further decrease the reliability. On the other hand, reliability is vastly increased by the standardised approach, with everyone being subjected to the same exam content and conditions, in contrast to the long case and patient management problems. It has been estimated that it would be necessary for each candidate to interview ten long cases to achieve the test reliability required for such a high-stakes examination (Wass *et al*, 2001).

The argument that simulated patients are not reliable probably does not hold up. If the 'patients' are appropriately trained, the evidence suggests they ensure standardisation of the procedure and excellent reliability (Hodges *et al*, 1998). Still, it could be argued that complex aspects of psychopathology, such as thought disorder, are impossible to simulate, thus decreasing the validity of simulated patients.

Global marking schemes are less reliable than a checklist scheme but this can be countered by the provision of two independent marking examiners for each station. Alternatively, an increase in the number of stations could be considered (Newble & Swanson, 1988).

Practicability

The CASC is not without practical difficulties; considered preparation is needed to ensure there is sufficient number of examiners and simulated patients trained specifically for the CASC. The practical benefits of assessing a large cohort at one time are clear and the standardisation of scenarios and use of simulated patients avoids the need for finding large numbers of suitable real patients, as in the long case.

Effect on learner

The CASC appears to promote learning of a broad curriculum in a fair and relatively transparent manner, allowing little room for discrimination of candidates. It is clear that this is a tool designed to assess competency rather than excellence and critics could argue that the old clinical exam system had enough flexibility to test the highest performing candidates further and assess excellence, which a pass/fail criterion-based system cannot and was not designed to do. Promotion of excellence in clinical

education & training

147

education &
training

skills and another more appropriate form of assessment could be developed as a further option for interested candidates.

Conclusions

It is hard to argue that the CASC will be any less valid or reliable than its predecessors, the long case and patient management problems. The validity, reliability and successful implementation of the OSCE lend some support to the CASC as its related successor. There are some remaining doubts with regard to the validity of the CASC in assessing higher level clinical skills. Also, reliable assessment of competence may be compromised by the single examination format. The global marking scheme has better validity than the checklist, but only in the hands of a trained expert, and paired examiners will optimise reliability. The process is relatively transparent and open to scrutiny, making it a fairer assessment than its predecessors but it may not promote excellence.

It remains to be seen as to whether the CASC stands up on its own merits as a valid and reliable test of higher level competencies. It is hoped that essential skills in holistic assessment and management are not lost by losing the realities and complexities that the long case and patient management problems provided. The CASC's success will also depend on the choice of outcome measures used – comparing competencies between the CASC doctors and those of us who have been assessed in the old system is an extremely complex task and the attribution of any change in competency due to the introduction of the CASC may be impossible. With reference to Miller (1990), it should not be forgotten that being able to 'show how' is not the same as 'do' and a demonstration of competence does not necessarily predict future performance (Rethans *et al*, 1991).

The CASC appears to be a useful tool but it needs to be placed in the context of other assessments and workplace-based assessments, and appraisal processes are particularly vital in further informing on clinical competence. One thing is certain, while we continue to

be guided by the current GMC code of *Good Medical Practice* (General Medical Council, 2006), competency-based assessments are unlikely to be replaced.

Declaration of interest

None.

References

BLOOM, B. (1971) *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain.* David Mackay.

FOWELL, S. L., SOUTHGATE, L. J. & BLIGH, J. G. (1999) Evaluating assessment: the missing link? *Medical Education*, **33**, 276–281.

GENERAL MEDICAL COUNCIL (2006). *Good Medical Practice.* GMC.

HARDEN, R. M. & GLEESON, F. A. (1979) Assessment of clinical competence using an Objective Structured Clinical Examination (OSCE). *Medical Education*, **13**, 41–54.

HODGES, B., REGEHER, G., HANSON, M. *et al* (1997) An objective examination for evaluating psychiatric clinical clerks. *Academic Medicine*, **72**, 715–721.

HODGES, B., REGEHR, G., HANSON, M., *et al* (1998) Validation of an objective structured clinical examination in psychiatry. *Academic Medicine*, **73**, 910–912.

HODGES, B., REGEHER, G., McNAUGHTON, N., *et al* (1999) OSCE checklists do not capture increasing levels of expertise. *Academic Medicine*, **74**, 1129–1134.

MILLER, G. (1990) The assessment of clinical skills/competence/performance. *Academic Medicine*, **65** (suppl.), s63–s67.

NEWBLE, D. I. & SWANSON, D. B. (1988) Psychometric characteristics of the objective structured clinical examination. *Medical Education*, **22**, 325–334.

RETHANS, J., STURMANS, F., DROP, R., *et al* (1991) Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice. *BMJ*, **303**, 1377–1380.

ROYAL COLLEGE OF PSYCHIATRISTS (2006). *A Competency Based Curriculum for Specialist Training in Psychiatry.* Royal College of Psychiatrists.

TYRER, S. & OYEBODE, F. (2004) Why does the MRCPsych examination need to change? *British Journal of Psychiatry*, **184**, 197–199.

WASS, V., JONES, R. & Van DER VIEUTSON, C. (2001) Standardised or real patients to test clinical competence? The long case revisited. *Medical Education*, **35**, 321–325.

Catherine M. Thompson   Specialist Registrar in Liaison Psychiatry, Harplands Hospital, North Staffordshire Combined Healthcare Trust, Hilton Road, Stoke on Trent ST4 6TH, email: catherine.thompson@northstaffs.nhs.uk