

The long case: a case for revival?[†]

Albert Michael,¹ Ranga Rao,² Vishaal Goel³

The Psychiatrist (2013), 37, 377–381, doi: 10.1192/pb.bp.113.043588

¹Norfolk and Suffolk NHS Foundation Trust; ²South London and Maudsley NHS Foundation Trust; ³Northumberland, Tyne and Wear NHS Foundation Trust

Correspondence to Ranga Rao (ranga.rao@slam.nhs.uk)

First received 10 Apr 2013, final revision 9 Aug 2013, accepted 4 Sep 2013

Summary Until 2003 the long case was the clinical component of the two summative examinations for the Membership of the Royal College of Psychiatrists (MRCPsych). This changed to an Objective Structured Clinical Examination (OSCE) format, initially for the Part I examination, and more recently to the Clinical Assessment of Skills and Competencies (CASC) at the end of 3 years of basic training. Although there are distinct advantages to the objective, competency-based assessment, questions remain on its validity. The expectation that formative workplace-based assessments would fill in the void left by the loss of the long case has not materialised. The options for retaining the advantages of the long case within the CASC framework while minimising the shortcomings of the CASC are suggested as a way forward.

Declaration of interest A.M. is a College Examiner and Training Programme Director, General Adult Psychiatry, East of England Deanery. R.R. is a College Examiner and was a member of the Examinations Subcommittee until 2012.

The long case has been the cornerstone of the clinical examination since its inception in 1842 until the introduction of the Objective Structured Clinical Examination (OSCE) in 1964. In 2003 the Royal College of Psychiatrists substituted the long case in the Part I membership examination (MRCPsych) with the OSCE. In 2008 the College replaced the two clinical examinations with a single OSCE and renamed it the Clinical Assessment of Skills and Competencies (CASC), to be taken at the end of 3 years of training. As this new format of examination has been in use for a decade now, it is time to compare the merits of the two systems.

The traditional long case

The old long case required the candidate to spend about 60 min with a real patient, taking a history and performing an examination, while not being observed by the examiners. The candidate would then present and discuss their findings, diagnostic formulation, management options and prognosis to a pair of examiners. The candidate would be asked to interview the patient in front of the examiners and demonstrate and/or clarify aspects of the case. Marking criteria were based on the aforementioned domains.

The long case was evidently excellent at testing the very essence of medical practice, namely history-taking, examination, formulation, differential diagnosis and management planning. Presenting the case to the examiners and defending their findings and conclusions enabled the examiners to get a true sense of the candidates' clinical

skills and competence. The use of real patients, requiring genuine sensitivity, lent authenticity to the case and enabled assessment of the candidate's overall clinical approach. In formative assessments, the long case helps highlight areas for improvement.

The disadvantages of the long case include low reliability, low validity and the inability to generalise from one long case about the candidate's ability in other cases. The single-case format does not allow the breadth of skills to be tested¹ or to sample the curriculum widely,² which results in low reliability. Moreover, there is a lack of standardisation of diagnostic complexity.¹

The use of flexible, subjective and global judgements and the lack of clarity in the marking system lead to poor reliability³ and examiner bias.⁴ Hubbard *et al*⁵ found that the correlation of independent evaluations by two examiners was only 25%. Furthermore, Leichner *et al*⁶ showed that the luck of the draw in selection of examiners and patients played a significant role in the outcome of postgraduate examinations in psychiatry.

The candidate is not observed during the interview with the patient, except for a brief period during the *viva*. This gives the examiners little opportunity to reliably assess the candidate's ability to communicate with the patient,⁷ thereby compromising the validity of the long case. The competence in one long case does not indicate a candidate's ability across a range of other cases and clinical situations.^{8,9} The inability to assess candidate competence through a single case has been termed 'case specificity'.¹⁰ Wilkinson *et al*¹¹ estimated that at least five or six 85-minute long cases (60 min with the patient and 25 min with the examiners) were necessary to achieve 0.8 dependability (a more conservative figure of reliability).

[†]See commentary, pp. 382–383, this issue.

OSCE/CASC

The OSCE/CASC replaced the long case as it was purported to have better reliability.^{12–14} Currently, the MRCPsych clinical examination consists of 16 stations, split into two circuits; one circuit consists of 8 individual stations of 7 min with a preceding 1 min ‘preparation’ time, and the other circuit consists of 4 pairs of linked stations of 10 min each with an additional 2 min of ‘preparation’ time. The examination lasts 160 min.

The OSCE helps examine a greater breadth of problems than the traditional long case. Having a greater number of examiners reduces the effects of examiner variability. Standardised patients improve reliability and validity. The OSCE also enables the testing of scenarios that might have been distressing to a ‘real’ patient, for example bereavement and terminal illness.

Nevertheless, the OSCE has a number of disadvantages. The face validity and content validity of the OSCE are sufficient for testing the knowledge of junior trainees.^{15–17} However, it has questionable construct validity in assessing senior trainees (at core trainee year 3 (CT3) level), because a checklist approach would be unsuitable for assessing complex knowledge, practical and communication skills and competence of senior trainees, and risks oversimplifying real-life situations.¹⁸ Consequently, the CASC explicitly requires the use of global scores to assign pass/fail decisions. More recently, the College has been exploring domains of competency instead of a single summative judgement of pass or fail. This suggests that global judgement of mastery is more reliable than checklists.^{18,19} In spite of this, the OSCE is not suitable for assessing more complex, yet vital phenomena such as transference or ‘interpersonal connection’.²⁰

Validity is intrinsically linked to context. A number of 10-minute stations could never mimic a thorough, 1-hour clinical assessment, the daily bread of a jobbing psychiatrist.²¹ The current CASC requires candidates to assess a new patient or explain a diagnosis and its management in 7–10 min. This is unlikely to reflect their skills and competencies in completing such tasks in day-to-day clinical practice. Moreover, there is a risk of trainees who are competent in routine clinical work failing the exam, whereas those who may be clinically inept, but have prepared specifically for the CASC exam, may pass it.

Another, albeit anecdotal, observation is that since the introduction of the CASC trainees have become unwilling and/or unable to assess, formulate and present whole cases. They have adapted their learning methods and focus during clinical work to passing the exam, for example undertaking only those tasks that can be completed in 10 min. In the OSCE/CASC, history-taking, examination, formulation and management skills are only partially assessed, if at all, and have led to exam-focused, short-case-competent trainees. Consequently, the new generation of trainees miss out on the vital experience of conceptualising whole cases. This represents a gradual undermining of a holistic, biopsychosocial approach central to the culture of psychiatry.¹

How do the two formats compare?

There is ample evidence suggesting that the long case is at least as reliable as the OSCE.^{22–24} Norman²⁵ suggested that observed multiple long-case examinations may have better reliability than the OSCE. Wass *et al*²⁴ found that the reliability of the long case, when carried out with two pairs of examiners, was not better or worse than that of the OSCE. They estimated that a reliability of 0.8 can be achieved with ten long cases on history-taking, with two examiners observing each long case. However, the time, logistics and cost-effectiveness issues related to running multiple long cases with multiple examiners preclude their use in standard examinations.

As we know from clinical experience, the whole is more than the sum of its parts and artificially breaking down a 1-hour clinical encounter into a number of 7- or 10-minute bites is not sufficient to assess the complex clinical skills and competencies of senior trainees. The Psychiatric Trainees’ Committee of the Royal College of Psychiatrists has been concerned about the disconnection between routine clinical practice and the CASC stations and has requested longer CASC stations (G. A. Lomax, personal communication, 2013).

Perhaps one could conceptualise the difference between the long case and the CASC in terms of competencies and competence. Competencies are a series of discrete skills that are learnt and assessed separately. They are limited to visible behaviour and its measurement. They are necessary, but not in themselves sufficient for safe and effective practice. Competence, in contrast, is a holistic understanding of practice and all-round ability to carry it out. Competence takes into account the subtleties of sensitivity, imagination, wisdom, judgement and moral awareness that are the marks of a wise doctor, and is a better goal than competencies.²⁶ The CASC, as it states in its name, assesses competencies whereas the long case is more attuned to assess competence.

Workplace-based assessments

Workplace-based assessments (WPBAs) were introduced with the hope that they would preserve some of the advantages of the long case. They were rolled out at the same time as the CASC, as part of the formative assessment of a range of core skills mapped on to the curriculum, with the CASC offering the summative assessment at the end of basic training. One of the WPBAs, the Assessment of Clinical Expertise (ACE), gives supervisors the opportunity to observe trainees in a range of clinical situations and has the potential to assess the more abstract aspects of consultation, akin to the long case. Fitch *et al*²⁷ argue that there is limited evidence base for WPBAs and that they have been designed neither specifically for psychiatry nor for postgraduate education in the UK. Trainee psychiatrists can attempt the CASC only after they have successfully demonstrated their competencies through WPBAs. The CASC pass rate of 39.3%²⁸ suggests that the WPBAs are not assessing what they are supposed to assess. In addition, there are questions about the reliability of the WPBAs, with no systems to add any external quality assurance to the process.

Trainee attitude towards the WPBAs is revealing. Menon *et al*²⁹ found that most trainees, both junior and senior, are unimpressed with WPBAs. They are dissatisfied with the evidence underpinning the assessments, the manner of introduction of the WPBAs and the training of assessors. Furthermore, the majority of trainees did not find the WPBAs benefiting their supervision, training, clinical practice or confidence. They opined that the new system was unacceptable, did not accurately reflect their progress, was no better than the previous system and should not be retained. This is a fairly damning indictment, even from a group of cynical trainees.

The training and support of trainers/assessors is fundamental in maintaining the quality of any assessment system, including the WPBAs. Noel *et al*³⁰ suggested that brief training interventions are insufficient to produce the required accuracy. In this context, the finding that 22% of trainers had received no training whatsoever and that only half of those receiving training felt confident in undertaking the WPBA³¹ is worrying.

The WPBAs were hastily introduced to address the gap caused by the substitution of the long case with the CASC, but have convinced neither the trainees nor the assessors about their utility, let alone led to a reasonable pass rate.

Approaches to improve reliability/validity of examinations

There have been various attempts to meet the urgent need to increase the validity of the final clinical examination while maintaining its reliability.²⁷ Modifications to the long case include incorporating the best aspects of the OSCE, for example structuring the format and the marking scheme; increasing the number of examiners; observing the candidate's behaviour; increasing the number of cases to 4–6; and shorter assessment (e.g. 20–45 min). Attempts to improve the OSCE involved increasing the duration and the examiners directly questioning the candidates.^{9,32–35}

The Royal Australian and New Zealand College of Psychiatrists, for example, employ a hybrid scheme in their final clinical exams, first using an Observed Clinical Interview (OCI), essentially an observed long case, before the trainee progresses to their OSCE, akin to the Royal College of Psychiatrists' CASC.³⁶ The Royal College of Physicians and Surgeons of Canada employ a two-stage clinical examination. The first is a structured OSCE,³⁷ which can be taken at any stage of training. This consists of eight to ten 20-minute stations. There is no contact with 'real' or simulated patients. Much of the assessment takes the form of direct questioning by an examiner, akin to the Royal College of Psychiatrists' now abandoned 'Patient Management Problems'. The second is the Structured Evaluation of Clinical Evaluation Report (STACER),³⁸ which is taken before training is completed, at the stage of a 'junior consultant'. This is similar to the traditional long case but the candidate is observed throughout the assessment.

Gleeson³⁹ introduced the Objective Structured Long Examination Record (OSLER), a 10-item analytical record of the traditional long case, with an examiner-observed history-taking and physical examination, and a criterion-referenced marking scheme to improve the reliability of the

long case. Van der Vleuten & Schuwirth²³ noted the educational value of the OSLER in terms of providing feedback. They opined, however, that reliability could be better improved by increasing the number of cases than by focusing on observing the student during the long case. In a study using observed long cases, Pavlakis & Laurent⁴⁰ established that postgraduate trainees did not pay attention to physical examination skills as these had not previously been observed. They highlighted the value of observation of the long case as it forced the candidates to master clinical assessment skills. They disapproved of the focus on the discussion of patient management in the long case at the expense of the assessment of clinical examination technique.

Olsen *et al*³⁵ evaluated a structured question grid for the long case using two examiners, one of whom marked using a structured question grid and the other did not. They found no significant difference in 'the chance of students being assessed as failing' or in the likelihood of a discrepancy between the ratings. Standardising aspects of the case presentation and *viva* improves not only reliability but also the candidates' perception of fairness.

Wass & Jolly⁹ incorporated observation and multiple examiners into the long case. A pair of examiners observed and marked the history-taking and another pair marked the presentation, both using checklists and global scores. They found higher inter-examiner reliability for observation (checklist 0.72; global 0.71) than for the presentation (checklist 0.38; global 0.60). They concluded that observation of history-taking in the long case is a distinct component of clinical competence, which the traditional 'presentation only' format does not measure.

Norcini^{41,42} argued that: (a) case specificity; (b) examiner stringency; and (c) the aspects of competence evaluated contributed to the unreliability of the long case. He proposed: (a) increasing the number of cases; (b) minimising differences among examiners by increasing their number and standardising across examiners by training them; and (c) increasing the number of aspects of competence assessed, providing the examiners with lists of competencies and using examiner observation.

A different approach was adopted by Hamdy *et al*³³ in the Direct Observation Clinical Encounter Examination (DOCEE) and the Integrated Direct Observation Clinical Encounter Examination (IDOCEE).⁴³ In this method, two to three examiners together observed the candidates carrying out history-taking and physical examination of four to six patients. The generalisability coefficient for four cases and two examiners was 0.84 for each case. A similar reliability was achieved by using observed 14-minute history-taking component of the long case followed by a 7-minute interview.²⁴ It was found that if each long case was observed by one unique examiner, at least ten observed history-taking long cases were required to achieve 0.8 reliability. In addition, Luiz *et al*⁴⁴ reported 89% examiner agreement on candidate achievement of clinical skills when each candidate took two structured, standardised, observed long cases, each marked by a different examiner. There are other approaches to increasing the reliability by using more than one long case of different duration.⁴⁵

A proposal for a more reliable and valid examination

Considering the fact that the current summative examination has questionable validity and that the WPBAs have not filled the gap, we propose that the College replace the current CASC with fewer but longer stations. Instead of the current 16 stations in 160 min, we recommend 6–8 stations of 20–30 min duration, from different specialties. The stations could involve combinations of different skills from a blueprint based on the curriculum, for example history-taking, mental state examination, physical examination, formulation, explaining diagnosis, management and prognosis, etc. There could be linked stations where candidates formulate the case they assessed in the previous station and present to the examiners. The marking scheme for these stations could be global judgements, as it is with the current CASC examinations in addition to marking for two to three sub-domains. The logistics of organising such an examination and examiner training need further exploration. Two examiners at each station would increase the reliability.

This format would retain the main benefits of the OSCE/CASC but would also address the problems associated with the shorter stations, namely not assessing whole cases, not assessing the ability to formulate cases and develop comprehensive management plans, etc. It would align the examination more closely with the day-to-day clinical work of psychiatrists. In addition, it would render examination preparation crash courses obsolete, as simply taking a meticulous approach to the routine clinical work would enable passing the examination, which would benefit both patients and trainees. This system is also likely to simplify the process of organising the examination as well as making it cheaper to run, and might ensure that the future generations of psychiatrists are suitably equipped with the complex skills essential to the practice of psychiatry.

About the authors

Albert Michael MBBS, DPM, MD, FRCPsych is a consultant psychiatrist at Norfolk and Suffolk NHS Foundation Trust and training program director, general psychiatry, East of England Deanery. **Ranga Rao** MBBS, DPM, DMH, DipNBE (Psych), MD, FRCPsych is a consultant psychiatrist, Clinical Director and Joint Leader, Psychological Medicine Clinical Academic Group, South London and Maudsley NHS Foundation Trust, Maudsley Hospital, London. **Vishaal Goel** MBBS, MRCPsych is a consultant liaison psychiatrist at Sunderland Royal Hospital.

References

- Benning T, Broadhurst M. The long case is dead – long live the long case. Loss of the MRCPsych long case and holism in psychiatry. *Psychiatr Bull* 2007; **31**: 441–2.
- Ponnamperuma GG, Karunathilake IM, McAleer S, Davis MH. The long case and its modifications: a literature review. *Med Educ* 2009; **43**: 936–41.
- Wilson GM, Lever R, Harden RM, Robertson JIL. Examination of clinical examiners. *Lancet* 1969; **1**: 37–40.
- Maxim BR, Dielman TE. Dimensionality, internal consistency and inter rater reliability of clinical performance ratings. *Med Educ* 1987; **21**: 130–7.
- Hubbard JP, Levitt EJ, Schumacher CF, Schnabel TG. An objective evaluation of clinical competence. *N Engl J Med* 1965; **272**: 1321–8.
- Leichner P, Sisler GC, Harper D. A study of the reliability of clinical oral examination in psychiatry. *Can J Psychiatry* 1984; **29**: 394–7.
- Gleeson F. Assessment of clinical competence using the Objective, Structured Long Examination Record (OSLER). AMEE Medical Education Guide No. 9. *Med Teach* 1997; **19**: 7–14.
- van der Vleuten CPM. Making the best of the 'long case'. *Lancet* 1996; **347**: 704–5.
- Wass V, Jolly B. Does observation add to the validity of the long case? *Med Educ* 2001; **35**: 729–34.
- Eva KW. On the generality of specificity. *Med Educ* 2003; **37**: 587–8.
- Wilkinson TJ, Campbell PJ, Judd SJ. Reliability of the long case. *Med Educ* 2008; **42**: 887–93.
- Badger LW, deGruy F, Hartman J, Plant MA, Leeper J, Ficken R, et al. Stability of standardised patients' performance in a study of clinical decision making. *Fam Med* 1995; **27**: 126–33.
- Vu NV, Steward DE, Marcy M. An assessment of the consistency and accuracy of standardized patients' simulations. *J Med Educ* 1987; **62**: 1000–2.
- Vu NV, Barrows HS. Use of standardized patients in clinical assessments: recent developments and measurement findings. *Educ Res* 1994; **23**: 23–30.
- Hodges B, Regehr G, Hanson M, McNaughton N. An objective structured clinical examination for psychiatric clinical clerks. *Acad Med* 1997; **72**: 715–21.
- Hodges B, Regehr G, Hanson M, McNaughton N. Validation of an objective structured clinical examination in psychiatry. *Acad Med* 1998; **73**: 910–2.
- Thompson CM. Will the CASC stand the test? A review and critical evaluation of the new MRCPsych clinical examination. *Psychiatr Bull* 2009; **33**: 145–8.
- Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med* 1999; **74**: 1129–34.
- Regehr G, MacRae H, Reznik R, Szalay D. Comparing the psychometric properties of checklists and rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998; **73**: 993–7.
- Hodges B, Hanson M, McNaughton N, Regehr G. What do psychiatry residents think of an Objective Structured Clinical Examination? *Acad Med* 1999; **23**: 198–204.
- Hodges B. Variations of a theme by Harden. *Med Educ* 2003; **37**: 1134–40.
- Marwaha S. Objective Structured Clinical Examinations, psychiatry and the Clinical Assessment of Skills and Competencies; same evidence, different judgement. *BMC Psychiatry* 2011; **11**: 85.
- van der Vleuten CPM, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005; **39**: 309–17.
- Wass V, Jones R, van der Vleuten CPM. Standardised or real patients to test clinical competence? The long case revisited. *Med Educ* 2001; **35**: 321–5.
- Norman G. The long case versus objective structured clinical examinations. *BMJ* 2002; **324**: 748–9.
- Fish D, de Cossart L. *Developing the Wise Doctor: A Resource Book for Trainers and Trainees in MMC*. RSM Press, 2007.
- Fitch C, Malik A, Lelliott P, Bhugra D, Andiappan M. Assessing psychiatric competencies: what does the literature tell us about methods of workplace-based assessment? *Adv Psychiatr Treat* 2008; **14**: 122–30.
- Bateman A. *MRCPsych Examinations Cumulative Results 2008–2010*. Royal College of Psychiatrists, 2011.
- Menon S, Winston M, Sullivan G. Workplace-based assessment: survey of psychiatric trainees in Wales. *Psychiatr Bull* 2009; **33**: 468–74.
- Noel GL, Herbers JE, Capow MP, Cooper GS, Pangaro LN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? *Ann Intern Med* 1992; **117**: 757–65.

- 31 Babu KS, Htike MM, Cleak VE. Workplace-based assessments in Wessex: the first 6 months. *Psychiatr Bull* 2009; **33**: 474–8.
- 32 McKinley RK, Fraser RC, van der Vleuten C, Hastings AM. Formative assessment of the consultation performance of medical students in the setting of general practice using a modified version of the Leicester Assessment Package. *Med Educ* 2000; **34**: 573–9.
- 33 Hamdy H, Prasad K, Williams R, Salih FA. Reliability and validity of the direct observation clinical encounter examination (DOCEE). *Med Educ* 2003; **37**: 205–12.
- 34 Norcini JJ, Blank LL, Duffy D, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med* 2003; **138**: 476–81.
- 35 Olsen LG, Coughlan J, Rolfe I, Hensley MJ. The effect of a Structured Question Grid on the validity and perceived fairness of a medical long case assessment. *Med Educ* 2000; **34**: 46–52.
- 36 Royal Australian and New Zealand College of Psychiatrists. *Regulations: Basic Training and Advanced Training for Fellowship*. RANZCP, 2003 (<http://psychtraining.org/Regulations.pdf>).
- 37 Royal College of Physicians and Surgeons of Canada. Specific information regarding the Royal College oral and clinical component examination in psychiatry. RCPSC, 2013 (<http://www.royalcollege.ca/cs/groups/public/documents/document/y2vk/mday/-edisp/tztest3rcpsced002222.pdf>).
- 38 Royal College of Physicians and Surgeons of Canada. *Structured Assessment of Clinical Evaluation Report (STACER): Psychiatry Clinical Evaluation for Senior Residents (PGY-4 and PGY-5)*. RCPSC, 2011 (<http://www.royalcollege.ca/cs/groups/public/documents/document/y2vk/mdaw/-edisp/tztest3rcpsced000734.pdf>).
- 39 Gleeson F. Defects in postgraduate clinical skills as revealed by the objective structured long examination record (OSLER). *Ir Med J* 1992; **85**: 11–4.
- 40 Pavlakis N, Laurent R. Role of the observed long case in postgraduate medical training. *Intern Med J* 2001; **31**: 523–8.
- 41 Norcini JJ. The validity of long cases. *Med Educ* 2001; **35**: 720–1.
- 42 Norcini JJ. The death of the long case? *BMJ* 2002; **324**: 408–9.
- 43 Abouna GM, Hamdy H. The integrated direct observation clinical encounter examination (IDOCEE) – an objective assessment of students' clinical competence in a problem-based learning curriculum. *Med Teach* 1999; **21**: 67–72.
- 44 Luiz EAT, Roberto OD, Fernando CF, Eduardo F, Lio CM, Ana LCM, et al. A standardised, structured long case examination of clinical competence of senior medical students. *Med Teach* 2000; **22**: 380–5.
- 45 Kroboth FJ, Hanusa BH, Parker S, Coulehan JL, Kapoor WN, Brown FH, et al. The inter-rater reliability and internal consistency of a clinical evaluation exercise. *J Gen Intern Med* 1992; **7**: 174–9.