

Evidence-based evolution of the high-stakes postgraduate intensive care examination in Australia and New Zealand

Richard P Lee, Bala Venkatesh and Peter Morley

The Fellowship Examination of the Joint Faculty of Intensive Care Medicine (JFICM) is the “exit” exam for intensive care medicine specialty training in Australia and New Zealand. It developed from the exam that was first set in 1979 as the fellowship examination of the then Section of Intensive Care of the Faculty of Anaesthetists of the Royal Australasian College of Surgeons.

The effect of this exam has been felt throughout Australia and New Zealand, where most intensivists have now passed the exam. The reputation of the training program and the exam has resulted in doctors from around the world coming to train in ICUs in Australia and New Zealand. The impact of the exam has also extended beyond Australia and NZ. Fellows have returned to their own countries worldwide and occupy senior faculty and administrative positions in intensive care medicine (Table 1).

Comprehensive surveys in 1986, 1992 and 1998 confirmed the high level of acceptance of both the exam and the training program.¹⁻³

In this review, we discuss the setting up of the ICU examination, the major periods of change and development, and the evidence-based rationale for those changes, and also outline the ongoing challenges of the assessment process.

The examination

The Fellowship Examination has been used for accreditation, and also as a basis to drive and direct learning within the intensive care training program. Concerns have been expressed that a high-stakes exit examination may not accurately assess clinical competence,⁴⁻⁶ may distract from training, may contain unreliable subjective assessments⁵⁻⁹ with contentious standard setting,¹⁰ and may not necessarily encourage a culture of lifelong learning. These concerns have led to efforts to strengthen the process, to supplement the exam with other assessments during training, and to make it a feasible process to continue into the future.

The exam should be viewed in the context of the training program as a whole. Currently, the program consists of:


- 3 years of basic hospital training and 3 years of advanced hospital training (in addition to a generic postgraduate year 1);
- Success at a basic sciences exam;
- Attendance at a course on communication, grieving and organ donation (ADAPT course);
- Successful submission of a formal research project;

ABSTRACT

The Fellowship Examination for Intensive Care Medicine in Australian and New Zealand, first held in 1979, has undergone four distinct periods of development and change:

- 1979–1996. Initiation and establishment of the exam as a relevant and comprehensive assessment process for a new specialty.
- 1997–2001.
 - Revision to increase breadth of coverage and reliability for a growing number of candidates, and to ensure that each candidate received the same exam.
 - Expansion to incorporate assessment of CanMEDS skills, including communication, procedures and professional qualities.
 - Lengthening to increase the number of exposures, to ensure reliability.
 - Quarantining of candidates to allow the provision of a similar exam for each candidate.
- 2002–2006. Increasing emphasis on examiner training and standard setting, increasing feedback to candidates to improve the educational experience and guide exam preparation, and blueprinting of questions to maintain validity.
- 2007 onwards. Logistic revision to ensure feasibility for a rapidly growing number of candidates, and refinement to apply modern standard setting and quality control.

The exam has been regarded as a “tough but fair” assessment in its 30 years of existence, and the committee overseeing its development has aimed to continually review the process to maintain those qualities, as well as reliability, validity and feasibility. The increasing number of candidates has allowed usable statistics to be accumulated but has tested the feasibility of running such a labour-intensive exam. To date, there have been 800 presentations to the exam, with 498 successful candidates.

 Crit Care Resusc 2009; 11: 144–150

- Acceptable performance in structured in-training assessments; and
- Success at the Fellowship Examination, which is mandated towards the end of the 6-year apprenticeship program.

Table 1. Origin and region of practice of intensive care trainees and fellows

	Percentage
Trainees' country of origin*	
Australia and New Zealand	42%
Overseas	58%
Fellows' region of current practice	
Australia	78.8%
New Zealand	8.5%
Europe	5.5%
Hong Kong	3.1%
North America	2.2%
Asia (other)	1.5%

* Based on trainees presenting at the past four examinations.

Table 2. Numbers of candidate presentations and examiner requirements

	Average no. per exam
Candidates presenting	
1979–1988	3
1989–1993	5
1994–1998	10
1999–2003	15
2003–2007	38
Examiners required	
2002	8
2003	11
2004	11
2005	12
2006	15
2007	21

Changes have been made to the exam through four epochs to maintain or improve its reliability while preserving validity,^{5,6} and to maintain its feasibility in the face of a growing number of candidates (Table 2), and a growing body of evidence on assessment processes.⁶ The exam has been revised regularly, particularly to meet the increasing expectations of the community that the process will help effectively assess the clinical competence of fledgling intensivists.

The validity of the exam has been assessed in terms of how it appears to assess relevant knowledge (face validity), how test items are blueprinted to the objectives of training (content validity), and whether it predicts a sustainable career in intensive care medicine (predictive validity).⁷ The data collected have been regularly analysed for measure-

ment error and bias, and measures have been introduced to minimise these.

1979–1996: establishment

The fellowship exam was first held in 1979, with one candidate and six examiners. To our knowledge, it was the first such exam in the world. The principal architect was Professor G A (Don) Harrison who helped design the exam as part of his Master of Medical Education thesis.¹¹ The guiding principle was the creation of intensive care medicine as a separate specialty, and the aim of the exam was to assess the many relevant skills and areas of knowledge that are required of an independent intensive care specialist. This was a unique concept at the time when most ICUs in Europe and North America were run by part-time intensivists with no formal training in intensive care medicine.

The initial exam comprised:

- Two written papers:
 - an essay paper with two essays, each of 60 minutes' duration; and
 - a short-answer question (SAQ) paper, with 15 questions of 150 minutes' total duration;
- A clinical section:
 - two cross-table vivas, each of 30 minutes' duration; and
 - 60 minutes of clinical cases, comprising 30 minutes of ICU ("hot") cases and 30 minutes of medical ("cold") cases.

The questions were based on the exam committee's concepts of the knowledge and problem-solving skills needed to allow independent practice in intensive care medicine, and the standard was criterion-referenced to that level. The broad coverage, diverse methods of assessment and exposure to multiple examiners were regarded as ideal in the educational environment of the time. Van Der Vleuten and Schwirth stressed that the number and duration of exposures of the candidate to assessors were critical in assessment.¹² Also, Miller stressed that "no single assessment method could provide all the data required for judgement of anything so complex as the delivery of professional services".⁸

In the early years, there was less emphasis on ensuring that each candidate received exactly the same cases or questions in the vivas or clinical section, but it was believed that multiple exposures to examiners and broad testing would maximise the reliability of the exam. There was a focus on problem-solving and patient management. Short-answer questions were chosen rather than multiple choice questions (MCQs) as they provided better, broad-based assessments of knowledge, and they were practical to set and mark for small numbers of candidates.⁷ There was also

evidence that the free-response format was better at discriminating good from poor candidates by avoiding cueing of answers.¹³

In 1982, an “investigations” viva (unstructured) of 30 minutes’ duration was added, which allowed questioning of each candidate by two examiners.

In 1986, a survey of successful exam candidates was undertaken to test the predictive validity of the exam.¹ It found that 81% of respondents were practising predominantly in ICUs, and 51% were working as directors or deputy directors of ICUs. The training and the examination were regarded favourably. The survey was repeated in 1992 and 1998 with similar results.^{2,3}

In 1994 the exam was “blueprinted”¹⁴ to objectives of training, which were published in 1993.¹⁵

1997–2001: review

By 1996, there had been 38 exams and 240 candidate attempts, with an average pass rate of 69%. The number of candidates presenting per year had risen from one in 1979, to 14 in 1989 and 29 in 1997. The Royal College of Physicians and Surgeons of Canada had promulgated the CanMEDS competency framework in 1996, and it clearly delineated the professional qualities expected of a practising specialist, including medical expert, communicator, collaborator, manager, health advocate, scholar/teacher and professional.¹⁶ Further changes were therefore made to the exam because of the increasing cohort size, and the desire to further increase its reliability and to assess the expanding roles and professional skills of the intensivist.

The changes introduced after a review of the CanMEDS framework, the educational literature and the existing exam process, were:

- Addition of an objective structured clinical examination (OSCE) section,¹⁷⁻¹⁹ incorporating investigation, communication and procedure stations;
- Increase in the number of cross-table vivas from two to six separate exposures;
- Allocation of two separate sets of examiners to the clinical section — one pair for hot cases, and another for cold cases;
- Provision of structured questions with marking grids in all sections; and
- Replacement of the two essays with 15 SAQs with an emphasis on the “key features” approach.²⁰

The OSCE contained 10 active stations, and used the most experienced examiners, structured questions and checklist marking. At least two of the stations were interactive, for communication and procedure assessment. Observations of the candidate’s communication skills and ability to perform common procedures were included in the OSCE

to expand the skills assessed, and the candidates were quarantined between sections to allow use of the same questions for all candidates. Questions formulated around case-based problem-solving skills²¹ and practical ward-based topics helped ensure “book learning” was balanced by emphasis on clinical experience.

After the introduction of these changes, a survey of candidates (unpublished observations) showed that the new formats and systems were accepted as providing a fairer exam, with maintenance of face validity, and with greater reliability associated with the increased exposures.

2002–2006: refinement and maintenance of quality

From 2002, the process was further refined, with an increased emphasis on blueprinting the curriculum to questions in each section of the exam, and structuring of the individual questions. The practice of holding full-day workshops before each exam was continued, with an increased focus on examiner training to facilitate standard setting. With the larger and growing cohorts, it became feasible to accumulate statistics, to assist decision-making on the validity of the cut-off pass mark and criterion-referenced marking system.

Feedback to candidates was increased to improve the educational experience and guide exam preparation for the following cohorts. This consisted of detailed analysis of the individual’s weaknesses, as well as an increasingly detailed exam report, which included examples of questions asked and sample answers.

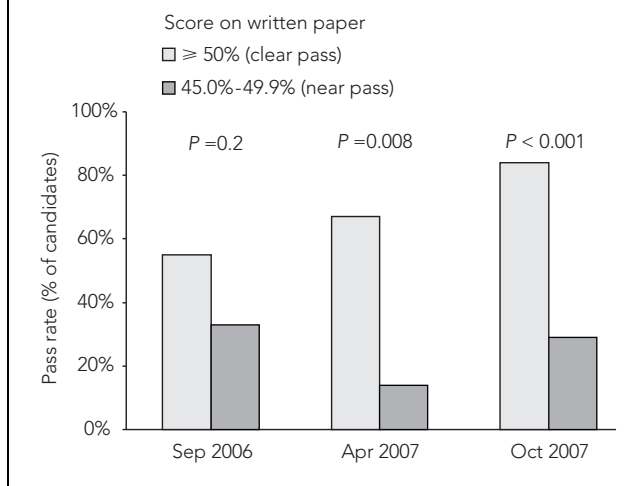
Also, in 2002, separation of passes in the written and clinical sections was introduced, so that a near pass (45%–49.9%) or clear pass $\geq 50\%$ in the written sections was required before the candidate could present to the oral sections. A clear pass could be carried into future exams. The aim was both to reduce the number of futile presentations of candidates who had performed poorly in the written section, and to increase justice by allowing candidates to carry a written-section pass into future exams.

2007 and the future: current status

In 2006, the Board of the JFICM approved several additional changes to the examination process. The drivers for these changes were:

- Observed and forecast increasing numbers of candidates presenting for the exam (Table 2). The number of junior trainees registered with the JFICM had risen exponentially, and therefore the predicted number of candidates who would present to the examination over the ensuing years rose dramatically.

Figure 1. Comparison of examination pass rates between candidates with a clear pass in the written paper and those with a near pass



- Increasing demand on numbers of examiners and examiner time (Table 2). Examiners needed to spend at least a week marking the written papers and another week for the clinical section of the examination.
- Difficulty in finding adequate numbers of suitable cold medical cases for examination purposes.

The changes that came into force included:

- A pass of 50% or more in the written section was required for the candidate to be invited to the oral section. This was introduced primarily to further reduce the likelihood of futile presentations (based on analysis of data from three examinations, which showed a significantly higher failure rate for candidates who presented to the oral component of the exam after receiving a written mark less than 50% [near pass]; Figure 1).
- The written section was changed to two papers of SAQs with an increasing emphasis on data analysis and interpretation, radiology, monitoring, equipment and basic clinical examination.
- The viva section was changed to eight stations, each of 10 minutes' duration, and included a procedure or a simulation station and a communication station.
- The clinical examination section was converted to two ICU cases of 20 minutes each (involving two pairs of examiners), and the cold medical cases were removed. This increasing emphasis on ICU clinical cases was associated with the incorporation of a clearer assessment process for the skills previously examined in cold cases (see below, *Evolution of the clinical cases*).

The distribution of marks from 2007 is:

- written section, 30 marks;
- viva section, 40 marks; and

- clinical section, 30 marks.

The pass criteria are:

- a minimum of 50% or greater total mark; and
- a fail in not more than one section; and
- a score of 40% or greater in the clinical section.

Evolution of the clinical cases²²

Over the past 30 years, the literature on the validity and reliability of assessment using clinical cases has increased. These cases are traditionally used to assess candidates' competence in undergraduate and postgraduate examinations,²³ and varying formats have been described including:

- A single long case, where the candidate is allowed to spend about 45–60 minutes with the patient, unsupervised by examiners.
- Several short cases with an observed encounter between the candidate and patient of about 10–20 minutes.

Both formats are used to test the candidate's ability to examine a patient, elicit clinical signs, formulate a differential diagnosis, and interpret investigations.

The traditional long case has been intensely criticised for unreliability.²⁴ This is due mainly to the potential for erroneous pass or fail decisions based on a single case (case specificity), and variable interpretation by examiners (examiner specificity). The addition of extra short cases and observed case performance during clinical training partly overcomes this limitation,²⁵ and candidates now have to examine patients under supervision during training before being approved to sit the exam.

An intermediate case format has been used in our examination. The duration is longer than the conventional short case, and the discussion spans a wider perspective than merely clinical methods. The exposure to two separate teams, with two examiners in each team, and an increasingly structured approach to questioning and marking also improves the reliability. For logistic reasons candidates cannot see the same patients.

Quality control of the examination process

The evolution of the examination is summarised in Table 3. Several measures have been introduced to ensure fairness and maintain quality for the various stages of the exam including:

- Inclusion of multiple formats — written, vivas and clinical. Each component carries a significant part of the mark, thus providing an opportunity for candidates to make up for case specificity and deficiencies in one section.
- Exposure of all candidates to the same questions in the written exam and viva sections, to ensure objectivity and fairness.

Table 3. Evolution of the Fellowship Examination

Component	1979–1998	1999–2005	2005–2006	2007
Written	Written (essay paper + SAQ paper) (270 min)	Written (essay paper + SAQ paper) (270 min)	Written (2 SAQ papers) (300 min)	Written (2 SAQ papers) (300 min)
OSCE	Investigation vivas (30 min)	OSCE (100 min)	OSCE (100 min)	
Vivas	2 cross-table vivas (60 min)	6 cross-table vivas (60 min)	6 cross-table vivas (60 min)	Cross-table vivas (80 min)
Clinical	Medical cases (30 min) + ICU cases (30 min)	Medical cases (30 min) + ICU cases (30 min)	Medical cases (30 min) + ICU cases (30 min)	ICU cases (40 min)
Total duration	7 h	8 h 10 min	8 h 40 min	7 h

SAQ = short-answer questions. OSCE = objective structured clinical examination.

- Inclusion of multiple stations, with exposures to multiple examiners in the oral (vivas and clinical) sections. The examiners are blinded to the candidate's marks in the various stations and sections of the examination. The dependability coefficient of the hot case is currently being evaluated.
- Allocation of two examiners, who mark the candidate independently, to each question in the written section, the hot case assessment and some stations in the viva examination.
- Mandatory attendance of examiners at regular examiner calibration workshops:
 - At the annual written examination workshop, each examiner is required to submit SAQs with a structured marking grid, and these questions and answers are discussed at a face-to-face meeting of all examiners. "Hobby horses" and esoteric questions are removed. Examiners are also required to mark SAQs, and the degree of variability is continually assessed. The concordance (a qualitative measure of agreement between the examiners about whether candidates pass or fail) is of the order of 75%, and the intra-class correlation coefficient is about 0.5.
 - At a pre-oral examination workshop, all viva questions are reviewed, practiced between a "surrogate candidate" (usually an examiner) and an examiner, and the questions and timing are further refined if necessary. This face-to-face meeting of all examiners is organised twice a year, the day before the commencement of the oral examination.
 - Video calibration of examiners: all examiners are required to mark a video of a hot case and a viva, and the variability is evaluated and discussed.
- Examiners and candidates are assigned to each other in a random, blinded fashion.
- An independent assessor reviews the performance of examiners on a regular basis while they are examining.

- The conduct of the exam is frequently assessed by an independent external assessor.
- Appointments to the panel of examiners are based not only on qualifications, experience and competence, but also on referee reports that testify to the examiner's interrogation skills and ability to examine without bias.

Post-examination feedback

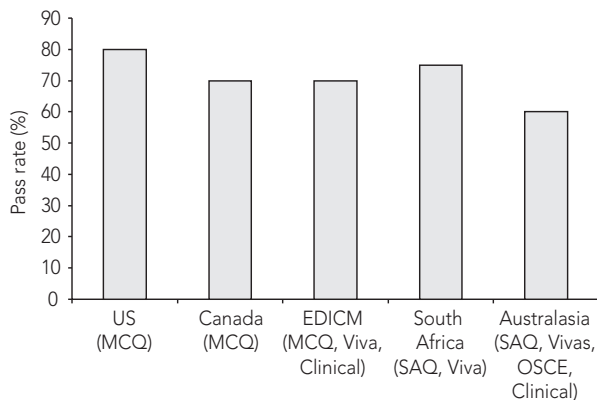
After completion of the examination, a detailed report is circulated to all the candidates, trainees, specialists, and supervisors of training in intensive care. This report consists of the answer templates for the written paper and viva section, and detailed statistics for performance in each section of the examination.

All unsuccessful candidates also receive a letter from the JFICM detailing their performance in various sections of the examination. The Chairman of Examinations also communicates with the candidates and/or their supervisors of training detailing areas of weakness, to help candidates prepare for future examinations.

Effect of the examination on ICU practice in Australia and New Zealand

There is indirect evidence that the exam may be partly responsible for the high standard of intensive care practice in Australia and New Zealand. Currently, 82% of specialists working in ICUs in Australia are intensivists certified by completing the JFICM requirements.²⁶ Data from the Critical Care Resources Survey for the Australian and New Zealand Intensive Care Society Research Centre also confirm the high standard of practice. For instance, in 64 186 episodes of care in Australian ICUs, there was a mean APACHE II score of 15.56 and standardised mortality ratio (SMR) of 58.38% (95% CI, 55.76%–58.08%) of predicted, and a mean APACHE III-J score of 50.79 and SMR of 82.14% (95% CI, 80.51%–83.80%) of predicted.⁴ Both SMRs were

Figure 2. Pass rates and formats of intensive care certification examinations in different countries*



US = United States. MCQ = multiple choice questions.
 EDICM = European Diploma in Intensive Care Medicine.
 SAQ = short answer questions.
 OSCE = objective structured clinical examination.

* Data as of 2007.

of an acceptable standard and had narrow confidence intervals, tending to indicate a homogeneously acceptable standard of care.

Conclusion

The Fellowship Examination of the JFICM commenced as a rigorous assessment of a new model of specialist, the intensivist, to complement the other elements of a new training program. The Examination Committee of the JFICM has reviewed and revised the process to align it with current evidence. At the time of writing, 59 Fellowship Examinations had been administered since 1979, with 800 presentations to the exam and 498 successful candidates. The size of the cohort presenting for the exams has continued to increase, in contrast to the American experience.²⁷ The pass rates are comparable with those of other assessment systems worldwide (Figure 2).²⁸

Although examinations are valuable for certification and can be useful drivers of learning, several challenges remain. Examinations require regular review and data analysis. Moreover, standard setting and determination of a valid pass mark remain elusive goals. Examinations are not an assessment panacea, as important elements of clinical competence are not practically or reliably assessed under exam conditions (eg, teamwork, ethical behaviour and communication), and performance at an exam does not always reflect behaviour at work.^{29,30} The examination will continue to undergo changes and evolve as knowledge of assessment methods improves.

Acknowledgements

The Editors of *Critical Care and Resuscitation* and *Anaesthesia and Intensive Care* have agreed to simultaneous, dual publication of this article because of the nature of the topic and the need for wide dissemination. We believe this article provides an important and useful background to the examination process for examiners, supervisors of training, fellows and trainees.

Author details

Richard P Lee, Past Chairman, Fellowship Examination in Intensive Care,¹ and Clinical Associate Professor²

Bala Venkatesh, Chairman of Examinations,¹ and Professor of Intensive Care³

Peter Morley, Past Chairman, Fellowship Examination in Intensive Care,¹ and Associate Professor⁴

1 Joint Faculty of Intensive Care Medicine, Australian and New Zealand College of Anaesthetists and Royal Australasian College of Physicians, Melbourne, VIC.

2 Northern Clinical School, University of Sydney, Sydney, NSW.

3 Division of Anaesthesiology and Critical Care, Princess Alexandra and Wesley Hospitals, University of Queensland, Brisbane, QLD.

4 Clinical School, Royal Melbourne and Western Hospitals, University of Melbourne, Melbourne, VIC.

Correspondence: bala_venkatesh@health.qld.gov.au

References

- 1 Byth PL, Harrison GA. A survey of successful FFARACS candidates in Australasia. *Crit Care Med* 1986; 14: 583-6.
- 2 Harrison GA, Byth PL. A survey of Fellows of the Faculty of Anaesthetists of the Royal Australasian College of Surgeons endorsed in intensive care by examination in the first 10 years of final examinations in intensive care. *Anaesth Intensive Care* 1992; 20: 203-10.
- 3 Harrison GA, Byth PL, D'Este CA. The third five-year survey of fellows (by examination) of the Faculty of Intensive Care, Australian and New Zealand College of Anaesthetists. *Anaesth Intensive Care* 1998; 26: 401-10.
- 4 Dauphinee WD. Assessing clinical performance. Where do we stand and what might we expect [editorial]? *JAMA* 1995; 274: 741-3.
- 5 Wass V, Van der Vleuton CPM, Shatzer J, Jones R. Medical education: assessment of clinical competence. *Lancet* 2001; 357: 945-9.
- 6 Epstein RM. Assessment in medical education. *N Engl J Med* 2007; 356: 387-96.
- 7 Jones RW. Medical specialist examinations: item format types and minimising error. *Anaesth Intensive Care* 2007; 35: 80-5.
- 8 Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65 (9 Suppl): S63-7.
- 9 Norcini JJ. Current perspectives in assessment: the assessment of performance at work. *Med Educ* 2005; 39: 880-9.
- 10 Cusimano MD. Standard setting in medical education. *Acad Med* 1996; 71: S112-20.
- 11 Harrison GA, Clarke GM. The training/examination programme in intensive care, Australian and New Zealand College of Anaesthetists. 2. Examination. *Anaesth Intensive Care* 1993; 21: 854-60.
- 12 Van der Vleuten CPM, Schwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ* 2005; 39: 309-17.

MEDICAL EDUCATION

- 13 Newble DI, Baxter A, Elmslie RG. A comparison of multiple-choice tests and free-response tests in examinations of clinical competence. *Med Educ* 1979; 13: 263-8.
- 14 Bridge PD, Musial J, Frank R, et al. Measurement practices: methods for developing content-valid student examinations. *Med Teach* 2003; 25: 414-21.
- 15 Clarke GM, Harrison GA. The training/examination programme in intensive care, Australian and New Zealand College of Anaesthetists. 1. Training. *Anaesth Intensive Care* 1993; 21: 848-53.
- 16 Royal College of Physicians and Surgeons of Canada. The CanMEDS Physician Competency Framework. Available at: <http://rcpsc.medical.org/canmeds/> (accessed Oct 2008).
- 17 Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ* 1975; 1: 447-51.
- 18 Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical exam. *Med Educ* 1988; 22: 325-34.
- 19 Newble DI. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ* 2004; 38: 199-203.
- 20 Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 2005; 39: 1188-94.
- 21 Schuwirth LWT, Verheggen MM, van der Vleuten CPM, et al. Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ* 2001; 35: 348-56.
- 22 Van der Vleuten CPM. Making the best of the "long case" [commentary]. *Lancet* 1996; 347: 704-5.
- 23 Wass V, Van der Vleuten CPM. The long case. *Med Educ* 2004; 38: 1176-80.
- 24 Norcini JJ. The death of the long case? *BMJ* 2002; 324: 408-9.
- 25 Wilkinson TJ, Campbell PJ, Judd SJ. Reliability of the long case. *Med Educ* 2008; 42: 887-93.
- 26 Drennan K, Hart G, Hicks P. Intensive care resources and activity: Australia and New Zealand 2006/2007. Melbourne: Australian and New Zealand Intensive Care Society Centre for Outcome and Resource Evaluation, 2008: 1-40.
- 27 Reshetar RA, Norcini JJ, Mills LE, et al. The first decade of the American Board of Internal Medicine certification in critical care medicine: an overview of examinees and certificate holders from 1987 through 1996. *Crit Care Med* 2000; 28: 1191-5.
- 28 Venkatesh B, Morley P, Lee RP, van Heerden PV. Examinations in intensive care medicine: an international perspective. *Crit Care Resusc* 2007; 9: 259-63.
- 29 Sharp LK, Bashook PG, Lipsky MS, et al. Specialty board certification and clinical outcomes: the missing link. *Acad Med* 2002; 77: 534-42.
- 30 Grosch EN. Does specialty board certification influence clinical outcomes? *J Eval Clin Pract* 2006; 12: 473-81. □