Performance and views of examiners in the Applied Knowledge Test for the nMRCGP licensing examination

A Niroshan Siriwardena, Hilton Dixon, Carol Blow, Bill Irish and Paul Milne

ABSTRACT

Background

A new computer-based Applied Knowledge Test (AKT) has been developed for the licensing examination for general practice administered by the Royal College of General Practitioners.

Aim

The aim of this evaluation was to assess the acceptability, feasibility, and validity of the test as well as its transfer to a computerised format at local test centres.

Design of study

Computer-based test and postal questionnaire.

Participants and setting

Panel of examiners, Membership of the Royal College of General Practitioners (MRCGP) examination, UK.

Method

Self-administered postal questionnaires were sent to examiners not involved with the development of the test after completing it. Their performance scores were compared with those of candidates.

Results

The majority of participants (80.9%) were satisfied with the new computer-based test. Responses relating to content and attitudes to the test were also positive overall, but some problems with content were highlighted. Fewer examiners (61.9%) were positive about the physical comfort of the test centre, including seating, heating, and lighting. Examiners had significantly higher scores (mean 83.3%, range 69 to 93%, 95% confidence interval [CI] = 81.9 to 84.7%) than 'real' candidates (mean 75.0%, range 45 to 94%, 95% CI = 74.6 to 75.5%), who subsequently took an identical test.

Conclusion

The new computer-based licensing test (the AKT) was found to be acceptable to the majority of examiners. The pass–fail standard, determined by routine methods including an Angoff procedure, was supported by the higher success rate of examiners compared with candidates. The use of selected groups to assess high-stakes (licensing) examinations can be useful for assessing test validity.

Keywords

assessment; general practice; validity.

INTRODUCTION

In 2007 the new Membership of the Royal College of General Practitioners examination (nMRCGP) replaced the previous system for the licensing of GPs in the UK. The nMRCGP is the single assessment system for UK trained doctors wishing to obtain a Certificate of Completion of Training in General Practice, entry to the GP Register of the General Medical Council, and membership of the Royal College of General Practitioners.[1] This requirement is part of the endpoint assessment of training for GP registrars to certify their fitness for independent practice; before that training consists of 3 years of specialty training, including at least a year in general practice, after two postgraduate years of foundation training.[2] The components of the endpoint assessment are the Applied Knowledge Test (AKT), a clinical skills assessment, and a workplace-based assessment, which, together, assess the curriculum for specialty training for general practice.[3]

The AKT is delivered by the Royal College of General Practitioners (RCGP) to standards agreed by

***AN Siriwardena**, MMedSci, PhD, FRCGP, professor of primary care, School of Health and Social Care, University of Lincoln; **H Dixon**, BSc, DRCOG, EdD, FRCGP, medical director, County Durham and Darlington Primary Care Trusts, County Durham; **C Blow**, FRCP, FRCGP, Dip Derm, DRCOG, Applied Knowledge Test clinical lead, MRCGP examination, Aberdeen; **B Irish**, BSc, MMEd, FRCGP, head, School of General Practice, Frenchay Hospital, Bristol; **P Milne**, PhD, FRCGP, senior lecturer in general practice, School of Public Health and Clinical Sciences, University of Central Lancashire, Colne.*

Address for correspondence
Professor AN Siriwardena, School of Health and Social Care, Faculty of Health, Life and Social Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS.
E-mail: nsiriwardena@lincoln.ac.uk

Submitted: 1 June 2008; **Editor's response:** 22 July 2008; **final acceptance:** 6 October 2008.

This is the full-length article of an abridged version published in print. Cite this article as: *Br J Gen Pract* 2009; DOI: 10.3399/bjgp09X395111.

the Postgraduate Medical Education and Training Board (PMETB), which is the independent regulatory body set up to oversee the content and standards of postgraduate medical education across the UK. The examination is designed to test knowledge and application of knowledge in the context of UK general practice. Questions cover three broad domains:

- medicine related to general practice (80%), comprising general medicine, surgery, medical specialties (such as dermatology, ear, nose and throat, ophthalmology), psychiatry, women's and child health;
- practice administration (10%); and
- critical appraisal and research methodology (10%).

The examination uses a number of question formats (including single best answer, extended matching questions, and pictorial images) to test the breadth and depth of candidates' knowledge. A test specification is used to ensure adequate coverage of the content as defined in the nMRCGP regulations[4] and the test is blueprinted against a detailed curriculum for general practice, which has been externally validated to reflect contemporary UK general practice.[3,5]

The test is constructed by 10 expert item writers and the performance of individual question items is reviewed using classical test theory, which underpins the evaluation and refinement of poorly performing questions by this expert group.[6] The importance of assessing the validity of postgraduate medical examinations has been emphasised, yet evaluation of assessments is often overlooked. Validity is understood in this context to comprise a number of components of construct validity including content, response process, internal structure, relationship to other variables, and consequences.[7]

Studies of the construct validity of the earlier multiple-choice paper, including candidates' views of the examination content,[8] impact of the assessment on candidate learning,[9] and views and performance of general practice trainers,[10] have been published previously. The new AKT differs in a number of respects from the previous paper:

- the composition is different, with the new test involving a greater proportion of medicine (80% versus 65%), compared with administration (10% versus 15%) and research and statistics (10% versus 20%);
- it is a licensing examination and, therefore, the standard had to be derived independently of the previous test; and
- it is a computer-based, rather than paper-and-pencil test.

How this fits in

Little is known about the perceptions and performance of experienced doctors in licensing tests in the UK. The performance of a group of selected experienced doctors (MRCGP examiners) was compared with examination candidates and the views of examiners were analysed with regard to a computer-based Applied Knowledge Test introduced for licensing of family doctors as part of the new Membership of the Royal College of General Practitioners (nMRCGP) examination. The performance of examiners was significantly better than real candidates, supporting the pass–fail standard derived using routine methods. The majority of participants were positive towards most content areas and most aspects of test administration. The findings support the validity of the current licensing test.

The computer-based test superseded the previous paper-and pencil examination and is administered using local test centres across the UK. It is thought that this is the first UK postgraduate medical examination to employ a computer-based assessment as one of its components. Computer-based testing offers a number of potential advantages including high-quality images, the use of novel questions using multimedia clips, efficient data collection for statistical analysis, the potential for automated assembly of tests, and opportunities for feedback to candidates. GP registrars may sit this component at any time during their GP training but are encouraged to do so after commencing the GP element of their programme.

Given the importance of the new AKT as a licensing assessment and new method of delivery, it was agreed by the MRCGP Examination Board to pilot the new assessment by offering the panel of examiners — all of whom were practising GPs — the opportunity to sit the computer-based test, to provide feedback on the content as well as information on performance.

The aim of this evaluation was to assess the acceptability, feasibility, and validity of the AKT and the transfer to a computerised format at local test centres. This pilot investigated examiners' views of the format and content of the AKT, their experiences of sitting the test, and performance using the new medium of computer-based assessment.

METHOD

An outline of the evaluation was presented to the Panel of Examiners in February 2007 at their annual conference. Examiners are practising GPs, who are expert in assessment and involved in delivery of the nMRCGP. They had achieved a pass mark in the multiple-choice paper as a prerequisite for appointment. Following the presentation, an invitation was sent by email to all 162 examiners to contact the examination office if they wished to participate. Examiners who were invited were not

involved in constructing or setting the AKT, but were involved in other parts of the nMRCGP. They were specifically asked not to prepare for the examination in this initial invitation.

Examiners who expressed an interest in participating were sent details of the examination date and booking procedure, and a unique examination identifier was allocated to them. Some of those expressing an interest had difficulty in finding a suitable venue or time to take the test. Participants completed the test during a 4-week period in June and July 2007.

The examination consisted of a 200-item multiple-choice paper including single best answer, extended-matching, and summary-completion test formats.[11] Participants agreed to maintain confidentiality of the content of the examination. Post-participation evaluation questionnaires were sent with a covering letter, stamped addressed envelope, and instructions for completion. A further reminder was sent 2 weeks later. Feedback results were issued from a third party, by letter, a month later to ensure confidentiality. Their result was expressed as a raw score out of 200.

A standard setting exercise involving an Angoff procedure[12] was carried out after the same 'live' test was taken by real candidates. The Angoff procedure involves a small group of 6–10 stakeholders including a practising GP, a GP trainer, a GP assessor, and representatives from assessment bodies including the RCGP and General Medical Council. The group is asked to individually rate the probability of a borderline candidate passing an individual question in the test. The members are fed back the actual pass rate for each question if there is wide variation in responses so that the process is iterative. This is done for a large enough sample of questions to derive an agreed consensus pass mark. Participants were then sent the agreed pass mark.

Experiences relating to test centres, examination content, and attitudes to the examination were presented as summarised data. Template analysis was used to analyse and present qualitative data,[13] and a comparison of examiner and candidate scores was performed using a *t*-test. Quantitative data were analysed using SPSS (version 12.1).

RESULTS

Overall, 61 examiners (37.7%, 61/162) sat the pilot test out of 86 (53.1%) who expressed an interest in participation.

Responses to questions on experience of the test are summarised in Table 1. Most examiners were satisfied or very satisfied with booking arrangements,

Table 1. Experience of the Applied Knowledge Test.

	Positive or strongly positive *n* (%)	Not sure *n* (%)	Negative or strongly negative *n* (%)
Test centre characteristics[a]			
1. Ease of booking your test appointment	47 (74.6)	3 (4.8)	5 (7.9)
2. Choice of test-centre venue	49 (77.8)	3 (4.8)	3 (4.8)
3. Ease of access of the test centre from your home	48 (76.2)	1 (1.6)	6 (9.5)
4. Efficiency of test-centre staff	55 (87.3)	0 (0)	0 (0)
5. Physical comfort of the test centre (seating, heating, and lighting)	39 (61.9)	5 (7.9)	11 (17.4)
6. Ease of completing the AKT online	51 (80.9)	1 (1.6)	3 (4.8)
7. Security of examination materials (AKT paper/questions secure from theft)	45 (71.4)	6 (9.5)	3 (4.8)
Ease of content areas[b]			
8. General medicine	34 (54.0)	19 (30.2)	0 (0)
9. Surgery	40 (63.5)	13 (20.6)	0 (0)
10. Medical specialties (dermatology; ear, nose and throat; ophthalmology)	28 (44.5)	24 (38.1)	1 (1.6)
11. Psychiatry	41 (65.0)	12 (19.0)	0 (0)
12. Women's health	36 (57.1)	16 (25.4)	1 (1.6)
13. Child health	34 (54.0)	18 (28.6)	0 (0)
14. Administration and management	31 (49.2)	22 (34.9)	0 (0)
15. Critical appraisal, statistics, and research	23 (36.5)	21 (33.3)	9 (14.3)
Attitudes to content[c]			
16. Overall the paper assessed my knowledge of common problems in general practice	50 (79.4)	5 (7.9)	0 (0)
17. Overall the paper assessed my knowledge of important problems in general practice	53 (84.2)	1 (1.6)	1 (1.6)
18. Overall the paper assessed my knowledge of problems relevant to general practice	51 (80.9)	2 (3.2)	1 (1.6)
19. Overall the paper assessed my ability to apply my knowledge to problems in general practice	46 (73.0)	7 (11.1)	2 (3.2)
20. Overall the paper assessed my ability to apply my critical appraisal skills to problems in general practice	42 (66.7)	7 (11.1)	5 (9.9)
21. I feel that the questions in the AKT paper tested the range of my knowledge	48 (76.2)	3 (4.8)	4 (6.4)

[a]1 = very satisfied, 5 = very dissatisfied; [b]1 = very easy, 5 = very difficult; [c]1 = strongly agree, 5 = strongly disagree. Missing values account for totals less than 61. AKT = Applied Knowledge Test.

ease of access, efficiency of centre staff, and arrangements for security. The majority (80.9%) were also satisfied with the method of computer-based testing. However fewer (61.9%) were positive about the physical comfort of the test centre, including aspects such as seating, heating, and lighting. Responses relating to ease of content areas and attitudes to the test were also positive overall (Box 1).

Summarised free-text comments on test administration and test content are highlighted in Boxes 1 and 2 respectively. There were a few comments relating to test administration; these focused on access, comfort, and assistance. The main themes that emerged relating to content were that clinical questions were generally relevant but those on rare conditions, practice administration, and statistics were less so. Some questions referred to knowledge that GPs would look up in practice. Question formats such as images or more complex scenarios also posed problems for some.

Examiners had significantly higher scores (mean 83.3%, standard deviation [SD] 5.5%, 95% confidence interval [CI] = 81.9 to 84.7%) than 'real' candidates (mean 75.0%, SD 8.0%, 95% CI = 74.6

Box 1. Comments on administration at test centres.

- Access
  - Minor problems with booking and access were highlighted:

  'Was unable to attend due to problems with name on College listing being slightly different from passport.'

  - Some participants found the test centre difficult to find:

  'I had difficulty finding a venue on a day I could attend, so had to travel 40 miles. Having said that, once there, the staff were excellent.'

- Comfort

  Several comments were made about lighting, temperature ('too cold' or 'too hot'), seating ('uncomfortable'), the length of time sitting at a computer, and problems with drinking and toilet facilities (candidates' use of such facilities is restricted).

  - There were also expressed difficulties in viewing the computer screens:

  'Monitor was at an uncomfortable position for sitting for hours, and difficult to sit at computer screen for that length of time, [a] *change* [in the] *background colour of screen may help, or compulsory 20-minute break halfway through; harder than a paper exam for that reason.'*

- Assistance
  - Test-centre staff were found to be helpful:

  'No initial appointment available but [test centre] *staff helpful.'*

Box 2. Comments on question content.

- Relevance, ease, and clarity
  - There were positive comments about whether the test questions reflected UK general practice:

  'Generally a good test reflecting general practice.'

  - Some question formats were unclear including more complex criteria:

  'Question about sick notes … it was not clear whether this was a consultation, i.e. seeing the patient, or simple request for discharge letter.'

- Technical problems with graphics
  - Questions with images or graphics were thought by some participants to be less clear or needing more textual support:

  'The two questions with pictures were not good as there was inadequate history to place the appearance in context.'

  'Some of the evidence-based medicine questions had poor images.'

- Retained knowledge versus referral to knowledge sources
  - Some feedback suggested that experienced GPs would know where to look up information rather than retain this in day-to-day practice, or that questions should focus on commonly used data:

  'Some questions one could easily look up, e.g. DVLA reg[ulation]*s or vaccine requirements in* BNF [*British National Formulary*] *so one tends not to commit these to memory.'*

  'When faced with these results — which we do get — I would look up interpretational results or check with [the] *lab. Might have been better to use a more common result, which we deal with more frequently.'*

- Relevance of rare conditions, administration, and statistics
  - Participants commented that some clinical conditions presented were rarely seen in practice:

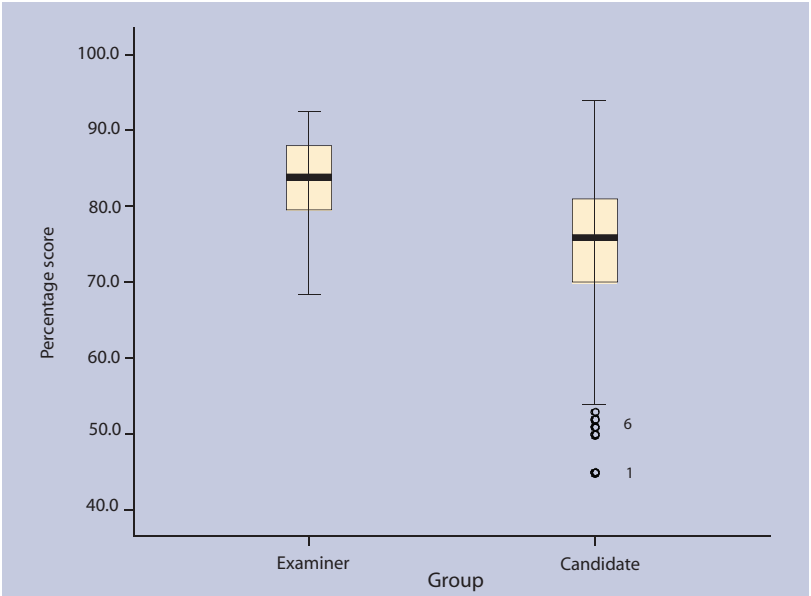
  'Some questions were on really quite uncommon conditions.'

  'There were several rare diagnoses. These are outside usual GP practice but, on occasion, fitted the clinical scenarios given. Whether they were the most likely, however, given low overall prevalence was arguable — and they are not diagnoses that you'd make in the surgery on the facts given.'

  - Questions on administration and statistics were felt by some to be less relevant:

  'I think some of [the] *administration questions would not need to be known by a starting GP — he would discuss with colleagues if he had any sense!'*

  'I'm not sure how vital statistics are to working GPs.'



The plot shows the median (thick horizontal line), interquartile range (box length), range (values between 1.5 and three box lengths from the upper or lower edge of the box), and extreme cases plotted as circles (with values more than three box lengths from the upper or lower edge of the box).

Figure 1. Box plot comparing examiner and candidate scores.

to 75.5%), who subsequently took the same test. Figure 1 shows a box plot comparing examiners' scores (median 84.0%, interquartile range [IQR] 8.8%, range 69–93%) with those of candidates (median 76.0%, IQR 11.0%, range 45–94%).

DISCUSSION

Summary of the main findings

This is the first such evaluation of a postgraduate licensing examination using examiners to assess the practicability, content, and construct validity. The content and computer-based method of delivery of the nMRCGP licensing examination was found to be acceptable to the majority of examiners. The pass–fail standard, determined by routine methods including an Angoff procedure, was supported by the pass rate of examiners compared with real candidates. Therefore, the validity of test content and cut-off scores was supported.

Strengths and limitations of the study

Participant examiners for the evaluation were volunteers and performance may have been different from non-participants. There may have been biases related to their membership of the panel and, as such, they may not be representative of GPs as a group.

Comparison with existing literature

This is one of very few studies assessing validity of a high-stakes medical examination against the views and performances of practising doctors.[10]

Computer-based testing has been advocated for licensing examinations in the UK[14] following a long history of successful use in the US.[15] Previous research has shown that test scores are comparable whether using pen-and-paper or computer-assisted testing in a variety of settings.[16–18] In this study, experienced GP examiners were positive towards a computer-based method of assessment. A study of Wessex GP registrars also found them to be positive towards a computer-based format for assessment of knowledge.[19] These findings are not surprising as UK general practice universally uses computerised records and computer literacy has become an essential skill.

Implications for future assessment and research

Computer-based testing has now been successfully implemented in a high-stakes professional medical examination[20] in UK general practice and provides lessons for licensing examinations in other specialities. Further evaluation could include views of candidates, such as those with disabilities,[21] and the use of different formats, for example video clips,[22] case simulations,[23] computer-based adaptive testing (in which test content changes depending on the candidate's responses),[24] and other stimulus formats. These might include video, audio, and multimedia clips that could increase validity of the assessment and open up the potential to test in new areas appropriate to the scope of computer-based testing.[25]

Funding body
Royal College of General Practitioners

Ethics committee
University of Lincoln Health and Social Care Research Ethics Committee. The study was not deemed to require National Research Ethics Approval because its primary purpose was as an evaluation

Competing interests
The authors are members of the AKT Development Group. There are no other competing interests

Acknowledgements
We thank the examiners who participated in this study, John Foulkes and members of the AKT development group who advised on this study.

Discuss this article
Contribute and read comments about this article on the Discussion Forum: http://www.rcgp.org.uk/bjgp-discuss

REFERENCES

1. Royal College of General Practitioners. *nMRCGP: the CCT and new membership assessment.* London: RCGP, 2008. http://www.rcgp-curriculum.org.uk/nmrgcp.aspx (accessed 1 Nov 2008).
2. Wass V. A guide to the new MRCGP exam. *BMJ Careers* 2005; **331:** 158–159. http://careers.bmj.com/careers/advice/bmj.331.7520.s158.xml (accessed 20 Dec 2008).
3. Riley B, Haynes J, Field S. *The condensed curriculum guide for GP training and the new MRCGP.* London: RCGP, 2007.
4. Royal College of General Practitioners. *nMRCGP regulations and related documents.* London: RCGP, 2008. http://www.rcgp-curriculum.org.uk/nmrcgp/regulations_and_documents.aspx (accessed 1 Nov 2008).
5. Royal College of General Practitioners. *RCGP curriculum map.*

London: RCGP, 2008. http://www.rcgp-curriculum.org.uk/extras/curriculum/index.aspx (accessed 1 Nov 2008).

6. Kline TJB. *Psychological testing: a practical approach to design and evaluation.* Thousand Oaks, CA: Sage Publications, 2005.
7. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ* 2003; **37(9):** 830–837.
8. Dixon H. Candidates' views of the MRCGP examination and its effects upon approaches to learning: a questionnaire study in the Northern Deanery. *Education for Primary Care* 2003; **14(2):** 146–157.
9. Dixon H. The multiple-choice paper of the MRCGP examination: a study of candidates' views of its content and effect on learning. *Education for Primary Care* 2005; **16(6):** 655–662.
10. Dixon H, Blow C, Irish B, *et al.* Evaluation of a postgraduate examination for primary care: perceptions and performance of general practitioner trainers in the multiple choice paper of the Membership Examination of the Royal College of General Practitioners. *Education for Primary Care* 2007; **18(2):** 165–172.
11. Case S, Swanson D. *Constructing written test questions for the basic and clinical sciences.* Philadelphia: NBME, 2001. http://www.nbme.org/PDF/ItemWriting_2003/2003IWGwhole.pdf (accessed 1 Nov 2008).
12. Norcini JJ. Setting standards on educational tests. *Med Educ* 2003; **37(5):** 464–469.
13. Crabtree BF, Miller WL (eds). *Doing qualitative research.* London: Sage, 1992.
14. Cantillon P, Irish B, Sales D. Using computers for assessment in medicine. *BMJ* 2004; **329(7466):** 606–609.
15. Clyman SG, Orr NA. Status report on the NBME's computer-based testing. *Acad Med* 1990; **65(4):** 235–241.
16. Russell M, Haney W. Testing writing on computers: an experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives* 1997; **5(3).** http://epaa.asu.edu/epaa/v5n3.html (accessed 20 Dec 2008).
17. Vrabel M. Computerized versus paper-and-pencil testing methods for a nursing certification examination: a review of the literature. *Comput Inform Nurs* 2004; **22(2):** 94–98.
18. Lee G, Weerakoon P. The role of computer-aided assessment in health professional education: a comparison of student performance in computer-based and paper-and-pen multiple-choice tests. *Med Teach* 2001; **23(2):** 152–157.
19. Irish B. The potential use of computer-based assessment for knowledge testing of general practice registrars. *Education for Primary Care* 2006; **17(1):** 24–31.
20. Irish B, Sales D. Does computer based testing (CBT) have a future in the assessment of UK general practitioners? *Education for Primary Care* 2006; **17:** 1–9.
21. Yocom CJ. Computer based testing: implications for testing handicapped/disabled examinees. *Comput Nurs* 1991; **9(4):** 145–148.
22. Lieberman SA, Frye AW, Litwins SD, *et al.* Introduction of patient video clips into computer-based testing: effects on item statistics and reliability estimates. *Acad Med* 2003; **78(10 Suppl):** S48–51.
23. Hagen MD, Sumner W, Roussel G, *et al.* Computer-based testing in family practice certification and recertification. *J Am Board Fam Pract* 2003; **16(3):** 227–232.
24. Roex A, Degryse J. A computerized adaptive knowledge test as an assessment tool in general practice: a pilot study. *Med Teach* 2004; **26(2):** 178–183.
25. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* 2004; **38(9):** 974–979.