
Making assessment fair in the MRCGP

A Summary of Research and Development related to the Clinical Skills Assessment and Applied Knowledge Test

MeiLing Denney
Research Lead, MRCGP

Introduction

The RCGP provides the licensing assessment –Membership of the Royal College of GPs (MRCGP) that determines whether a doctor is able to work in independent practice as a GP in the UK. The RCGP is committed to fairness in postgraduate assessment and has a Public Sector Equality Duty (PSED) as defined by section 149 of the Equality Act 2010 in respect of the development and delivery of the MRCGP. This ‘duty’ requires the RCGP to have due regard to the need to;

- Eliminate discrimination, harassment, victimisation and any other conduct that is prohibited by or under this Act;
- Advance equality of opportunity between persons who share a relevant protected characteristic and those who do not share it;
- Foster good relations between persons who share a relevant protected characteristic and persons who do not share it.

Protected characteristics include race and gender.

The fairness of the Clinical Skills Assessment (CSA) component of the MRCGP was legally challenged in 2014 with a Judicial Review brought by the British Association of Physicians of Indian Origin (BAPIO) because of persistent differential attainment between UK (UKG) and international medical graduates (IMGs), and between white and black and minority ethnic (BME) UKGs. These issues were summarized in an editorial in the BJGP (Rendel et al, 2015) [45].

These differentials were identified by the RCGP who have routinely analysed and published pass rate data by ethnicity, gender and place of qualification since 2005 in the interests of transparency.

Although BAPIO’s challenge was unsuccessful and the CSA was found to be lawful and fair, differential attainment by IMG and BME candidates still occurs throughout postgraduate and undergraduate examinations. Gender differences in performance also occur, usually favouring female candidates.

The RCGP is committed to investigating the causes of differential attainment in the MRCGP and to working with stakeholders to develop solutions. This paper summarises the rationale for the development of the current MRCGP and the extensive research and reviews that have been commissioned to ensure that the assessments are fair to all candidates

Background to the current MRCGP

Prior to 2007, the MRCGP was a membership examination consisting of a multiple-choice question paper, a written paper, an oral examination, and an assessment of consultation and clinical skills through examination of videoed real consultations submitted by the candidate. As an alternative to the video exam, candidates were able to set a simulated surgery OSCE, provided they could demonstrate that they met the criteria for this (being unable to submit a video). From 2007 onwards,

significant changes were made to the MRCGP to address the requirements of the then regulator the Postgraduate Medical Education and Training Board (PMETB) when re-designing the MRCGP for approval as the compulsory licensing assessment for general practice. PMETB required that the MRCGP should include a test with a much greater focus on clinical skills, and the Clinical Skills Assessment (CSA) was consequently developed following extensive investigation and analysis of the validity, reliability and feasibility of the current range of available methods. One of the conclusions was that the assessment should be administered in a controlled environment to enable standardisation of tasks and objective measurement, while still acting as an authentic replication of patient consultations.

The MRCGP now comprises an integrated assessment system, success in which is a pre-requisite for the issue of a certificate of completion of training (CCT) at the end of specialty training for general practice, after which a doctor is deemed competent to enter independent general practice in the United Kingdom. It is one of a tripos, the other two components being the Applied Knowledge Test (AKT) and Workplace Based Assessment (WPBA), which together cover the knowledge, skills, behaviours and attitudes laid out in the GP Specialty Training curriculum. More information on the MRCGP exams is publicly available on the RCGP website, including an MRCGP exam overview, and within the annual speciality reports submitted to the GMC (GMC Annual Speciality Report). [See 24, 32, 33.]

Before September 2010 the CSA pass mark was determined using a number-to-pass methodology. The change, in September 2010, to a borderline group standard setting methodology, was requested by the regulator and supported by RCGP examiners and governance committees on the basis that it enhanced fairness by enabling allowances to be made for any daily variability in the difficulty of the mix of cases and for compensation between performances on different cases. In 2010 the number of attempts permitted at the AKT and CSA was reduced to four, a figure consistent with the GMC requirements and also guidance subsequently issued by the Academy of Medical Royal Colleges (AoMRC). This approach was endorsed by the relevant governance committees, which include stakeholder and patient representatives.

Monitoring for Fairness within the MRCGP

The RCGP has led the field in its collection and publication of data relating to equality and diversity in assessment and research relating to fairness in its examinations (MRCGP Annual Reports). This has taken place systematically since 2002, and from 2005 it has published regular analyses of pass rate data by ethnicity, gender and place of qualification [33]. This transparency and approach to data publication has been commended by the PMETB, the GMC and the BMA in the past, and the RCGP has been at the forefront of collecting and analysing its processes in the manner recommended by the BMA in its surveys of 2006 and 2011 [2, 3].

External Scrutiny and Review by International Assessment Experts.

The RCGP commissioned an external review of the new licensing exam in 2007 from three international experts, Professors Geoff Norman, Cees van der Vleuten, and Chris McManus, to advise the college on its progress in the development of the new exam and the recommended direction of future developments of the Assessment. In broad terms the reviewers saw the RCGP as at the forefront of licensing tests, describing nMRCGP as being 'state-of-the art' in assessment terms and 'in the front line of approaches to specialty certification worldwide' [31].

In May 2009 the RCGP commissioned the Peninsula School of Medicine and Dentistry to review the CSA, carry out a generalisability study (unpublished) and model some of the examination data using different standard setting mechanisms. Their remit was to look at reliability and for sources of error

and they concluded that the variability of case mix from day to day could potentially disadvantage some candidates. The modelling also showed that using a borderline group method was feasible and would address some of the problems associated with the number-to-pass method. As a result, the standard-setting method was changed for the CSA. This generalisability study showed that examiners and cases make only a small contribution (5.4 and 5.9%) to the error variance with the major source of error being the interaction of the candidate and the case.

The GMC commissioned an independent external review in 2013 [20] to determine whether the MRCGP met GMC assessment standards and to identify any issues relevant to the GMC's *"The Trainee Doctor"* standards on equality, diversity and opportunity. A GMC press release at that time stated *"The report by Professor Aneez Esmail found that while there are significant differences in pass rates between different groups of doctors, the way they are assessed in the CSA is not the cause of those difference"*.

The review:

- documented the statistics on differential pass rates but made no findings of discrimination
- concluded that *"the method of the MRCGP assessment is not a reason for the differential outcomes"* and noted that *"the AKT is a machine marked examination testing applied clinical knowledge. There is a differential pass rate for both BME UK graduates and IMG graduates when compared to White UK graduates ... It is difficult to attribute this to bias because of the nature of the test ..."*
- posed some potential alternative explanations for differential pass rates, including the difference in exposure of various groups of medical graduates to general practice, and in the amount and quality of communication skills and reflective training received within medical schools.

A further external review of the AKT and the CSA commissioned by the RCGP was undertaken in 2014 by the international assessment expert Professor John Norcini, Foundation for Advancement of International Medical Education and Research (FAIMER). The review had a particular focus on equality and diversity, and the outcome was that the RCGP should consider increasing the amount of time candidates have available to complete the AKT to reduce the challenge for those candidates who are less proficient in the English language. This recommendation has now been put into action.

Research and development related to fairness generally in the MRCGP

Allied to the monitoring, analysis and publication of statistical information on the performance of MRCGP candidates by ethnicity, gender and place of primary medical qualification has been extensive research into fairness. Differential exam pass rates have been of concern to many organisations concerned with postgraduate medical education and training for many years, and as well as highlighting it in its annual reports, the RCGP has also published on the subject. An article [43] reported on international medical graduates (IMG)'s relative under-performance in these examinations, commenting that

- IMGs form an important and valued part of the British medical workforce, and that 37% of all doctors currently active on British GMC's list of registered medical practitioners (LRMP) are not British graduates.

Paterson et al (2011) [26] proposed an agenda to guide future research into the reasons for the substantial differences in outcomes for doctors taking postgraduate examinations as IMGs compared to their UK-trained peers. Using an interdisciplinary approach, they drew on expert inputs from a variety of academic and stakeholder experiences to summarise key issues surrounding fairness in assessment.

- Four key areas to guide further research were presented, ranging from design issues to analysing outcomes in practice; factors other than biased assessment tools should be explored as the major determinant of group differences.
- Evidence suggests that assessments in postgraduate training do not necessarily systematically exhibit bias by underpredicting the performance of minority group members
- In practice, the combination of selection and training placement systems often operate against the interests of the weaker recruits, thereby encouraging a cycle of educational deprivation.

The RCGP has worked with experts in linguistics from Kings College London on a study researching linguistic/cultural features in the CSA [1, 28]. 40 video recordings from the February 2011 CSA exam diet were subject to sociolinguistic analysis. The research team reported on their methods and the inter-collaborative work at the ALAPP (Applied Linguistics and Professional Practice) and COMET (Communication, Medicine and Ethics) conferences in 2011. The work is a sociolinguistic analysis of a key gatekeeping encounter - the licensing of doctors to practise in the UK and is based on a Knowledge Transfer Partnership between King's, Nottingham and Cardiff Universities in partnership with the RCGP. The final report has been published as a book, published by Kings on its website, which

- focussed on socio-linguistic factors that may contribute to the higher failure rate in the MRCGP of International Medical Graduates (IMGs) and graduates from minority backgrounds trained in the UK.
- Recommended that the Interpersonal Skills domain should be reviewed, as well as the role of communication and interpersonal skills in the CSA.
- has identified features of poor performance in the interpersonal skills and data gathering domains of the CSA which may be amenable to training interventions.

The RCGP is currently developing training resources based on this research.

A study was undertaken to assess whether IMGs passing the two examinations set by the Professional and Linguistic Assessments Board of the GMC are equivalent to UK graduates at the end of the first foundation year of medical training (F1), as the GMC requires, and if not, to assess what changes in the PLAB pass marks might produce equivalence [25]. GMC PLAB graduates performance data (PLAB1 and PLAB2, as well as International English Language Testing System (IELTS) scores) were linked with data from UK graduates from the MRCP(UK) Part 1, Part 2, and PACES and the MRCGP AKT and CSA. The study attempted to shed light on the reasons for differential performance between UK graduates and IMGs and found that:

- PLAB1 marks were a valid predictor of MRCP Part 1, MRCP Part 2, and MRCGP AKT, and PLAB2 marks correlated with MRCP PACES and MRCGP CSA.
- PLAB graduates had significantly lower scores on all MRCP and MRCGP assessments, and were more likely to fail and to progress more slowly than UK medical graduates. IELTS scores correlated significantly with later performance
- PLAB is a valid assessment of medical knowledge and clinical skills, correlating well with performance at MRCP and MRCGP, and that PLAB graduates' knowledge and skills at MRCP and MRCGP exams are over one standard deviation below those of UK graduates, although differences in training quality cannot be taken into account.
- Equivalent performance in MRCGP(UK) and MRCGP would occur if the pass marks of PLAB1 and PLAB2 were raised considerably, but that would also reduce the pass rate.

A paper on the linkage of MRCGP/MRCP has just been published in BMC Medical Education (Wakeford et al, 2015) [44]. The authors identified 2,284 candidates who had taken one or more parts of both assessments, MRCP(UK) and MRCGP, and analyzed performance on knowledge-based MCQs (MRCP(UK) Parts 1 and 2 and MRCGP Applied Knowledge Test (AKT)) and clinical examinations (MRCGP Clinical Skills Assessment (CSA) and MRCP(UK) Practical Assessment of Clinical Skills (PACES)).

- Correlations between MRCGP and MRCP(UK) were high, and these support the validity of each, suggesting they assess knowledge cognate to both assessments.
- BME candidates performed less well on all of the assessments - although White candidates out-perform BME candidates, the differences are largely mirrored across the two examinations.
- The similarity of the effects in independent knowledge and clinical examinations suggests the differences are unlikely to result from specific features of either assessment and most likely represent true differences in ability.

Fairness research relating to CSA Examiners

MRCGP examiners are carefully selected, trained and performance-monitored [8, 21].

- There is a robust process of selection and training for CSA examiners, with subsequent ongoing training, quality assurance and analyses of performance for all examiners on a regular basis.
- The efforts invested in ensuring robust selection processes are in place are a key factor in exam reliability and fairness for candidates.
- Examiner selection days are rigorous
- Inability to demonstrate adequate problem solving and team working skills were the two commonest reasons cited for failure of applicants on the selection days.

The RCGP collects equality and diversity data on its examiners as well as its candidates. A study of examiners' grades by their own and their candidates' gender, ethnicity and background looked at whether examiner sub-groups grade candidate sub-groups differentially [7]. The analyses were based on the 4,000 candidates taking the CSA in 2011-2012, and the 52,000 cases consequently adjudged by examiners that year.

- Candidate-examiner interaction effects are inconsistent in direction and slight in the calculated impact. As candidates' case scores are concerned, the only substantial sources of variance in this examination relate to their own (12% of variance), rather than the examiners' (0.2% of variance) demographic characteristics.
- No substantial effect is apparent in the CSA of examiners favouring candidates of the same gender, ethnicity or degree source.

CSA examiners' performance is statistically monitored. A study looked at three years' of examiner (n=187) marks of individual consultations, rank ordering examiners on all three continua in each year. We calculated the mean between-year correlation for each continuum [42]. The study concluded that:

- Examiners vary quite consistently regarding the level and spread of the marks they award
- The extent of their agreement with other examiners is less consistent. This may be due to examiner 'case specificity of performance'.

Fairness research relating to CSA Role-players

All role players the CSA are recruited from a single agency and are trained for their role by the agency and the RCGP. They participate in calibration sessions with examiners during each CSA day to optimise consistency across the three CSA circuits. Inadequate role player performance could compromise fairness for candidates [35].

- There is a clear process for the initial selection and training of the actors, and how they are prepared on the day of the exam.
- Quality assurance processes are designed to ensure that the exam has excellent role players performing as the simulated patients.

An observational study of roleplayer performance in the CSA using a semi-structured observation tool was undertaken in 2013. 470 video recordings of CSA consultations were analysed in detail by groups of assessors to establish whether there had been a genuine difference in roleplayer performance between candidates, whether it affected the subsequent roleplayer- candidate interaction and the overall challenge of the case [22]. This study :

- demonstrated a very high degree of RP consistency between candidates in the CSA
- showed no evidence of systematic bias on the part of CSA RPs towards any subgroup of candidates
- confirmed a definite difference in RP performance in 4.9% of consultations, with the commonest type of difference appearing to be some form of saving behaviour on the part of the role player.

Role players' views, which give the patient perspective, have also been collected to inform candidates in their preparation for the CSA. Focus-group discussions and questionnaires of RPs attending the RCGP's Simulated Surgery examinations were conducted in 2006/2007, followed by in-depth, one-to-one, semi-structured interviews with nine RPs attending the MRCGP Clinical Skills Assessment (CSA) in 2010 [34].

- RPs had clear views as to what made 'good' or 'poor' candidates, and suggested some new approaches to improving performance.
- Owing to the clinical nature of the examination, they did not feel able to mark or give feedback to candidates.

Research into the possibility of role-player bias in the CSA examined 92,989 case scores from the CSA using univariate analysis to inspect the data and then multifaceted analysis to quantify the extent of case variance attributable to interaction between candidate ethnicity and role player [6]. The results showed that:

- roleplayers do not constitute a significant source of error variance or unfair bias by candidate ethnicity
- RPs' case score difference by candidate ethnicity (or apparent bias) decreased markedly with increased candidate exposure (univariate analysis)
- candidate ethnicity by RP interaction accounted for a non-significant 2% of overall variance (multifaceted differential function analysis).

CSA case mix/palette selection

The RCGP has a cadre of skilled and trained case writers for its summative exams. Detailed guidance is used for case palette selection to ensure a representative mix of CSA cases on any one test day. To ensure an even mix of difficulty between exam days, cases are assigned a difficulty score both by the case writers and through on-going analysis of the performance of candidates in CSA consultations. The relative difficulty of all the cases in its case bank is designated by a numerical descriptor.

Work has been done to look at the variation in the challenge and reliability of CSA cases, using grounded qualitative analysis to look at the factors that affect the performance of a case [4]. A range of cases was used, and trained assessors looked at random samples to produce key word descriptors of the cases to explain their performance. These responses were categorized.

- There are common characteristics for easy cases, hard cases, and cases with high reliability.
- This case analysis demonstrated a way of improving fairness to candidates through the identification of characteristics of less helpful CSA case types, and providing information on case difficulty to optimise consistency between the examination days of the CSA.

Candidates' behaviour and views

A study has been undertaken to consider differences in performance between male and female candidates in the CSA [27]. Routinely collected data were analysed enabling detailed comparison of gender performance. 92,989 consecutive encounters were examined by candidate sex for the assessment overall, by assessment domain, and according to the curriculum area being assessed: the significance of differences was calculated by analysis of variance. The results showed that:

- female GP trainees outperform their male peers in the CSA overall, by assessment domain and by every curriculum statement
- the difference in performance is most marked in the areas of Women's Health and Sexual Health and least marked in Cardiovascular Problems and Rheumatology and Musculoskeletal.

To investigate the extent to which candidate age is a determinant of success, 7,680 CSA attempts between 2007 and 2010 were analysed [11]. Success was correlated with the variables (gender; time since qualification as a surrogate for age; ethnicity; UK or International Medical Graduate) first independently, then by multivariate analysis.

- Each of the individual predictors was highly significantly associated with success, but stepwise multiple regression showed that variance in success was explained as follows: country of primary medical qualification 30%; ethnicity 3%; gender 2%; and time since qualification 1%.
- Despite an apparently powerful negative relationship between candidates' age and examination performance, its extent is trivial when other confounding variables are accommodated.

The RCGP provides feedback to candidates - examiners then have 16 feedback specifics which they can then tick to describe the reasons for failure. In 2011 the RCGP explored an approach to combine fine-grain assessment data across candidates and cases towards providing candidates feedback on the pattern of their behaviour, without imposing a pre-conceived structure of clinical performance [13]. 8,352 of 36,296 candidate-case encounters in the CSA in 2009 were failed. These were subjected to exploratory factor analysis to ascertain patterns.

- Four factors emerged: poor data-collection/diagnosis, but management shared; disorganised/unsystematic, generally; doctor-centred, but management alright; and case focus- and risk-blind.
- Factor scores showed differences amongst candidate sub-groups.
- These factors are more complex than imposed domains - analysis may help develop more sophisticated feedback than that based on pre-conceived performance domains.

MRCGP Applied Knowledge Test (AKT)

When the AKT was developed for the licensing examination for general practice it was piloted on examiners and evaluated to assess the acceptability, feasibility, and validity of the test as well as its transfer to a computerised format at local test centres. The new computer-based licensing test (the AKT) was found to be acceptable to the majority of those involved in this initial evaluation [38].

Pass rates for the AKT mirror those for the CSA and reflect similar differentials in examination performance in knowledge tests in Higher Education and other medical examinations. The RCGP has also researched a number of areas relating to the AKT and differential performance.

One paper's aim was to compare relative performance of male and female candidates for gender-related items tested in the AKT [37]. Data were included from 3627 candidates. After adjusting for sex-neutral score, age, time since qualification, year of speciality training, ethnicity, and country of primary medical qualification, there were differences in performance in sex-specific questions.

- Males performed worse than females on female-specific questions (-4.2%, 95% confidence interval [CI] = -5.7 to -2.6) but did not perform significantly better than females on male-specific questions (0.3%, 95% CI = -2.6 to 3.2%).
- There was evidence of better performance by females in female-specific questions but this was small relative to the size of the test. Differential performance of males and females in sex-specific questions in a licensing examination may have implications for vocational and post-qualification general practice training.

A candidate questionnaire administered immediately after the exam sought candidates' opinions about the content and difficulty of the test, and this quantitative and qualitative paper [14] has been accepted for publication in *Education in Primary Care*. Key points are that:

- Candidates thought the AKT was valid and fair
- There was limited evidence of insufficient time but a small minority did not complete the paper
- Only 3% of respondents said English was not their preferred language for reading and writing

A recent paper [16] on non-pre-tested AKT items in order to test new and emerging knowledge shows that, if the question writing process is rigorous:

- New items testing new and emerging knowledge perform in terms of reliability as well as pre-tested questions.
- Their use enhances the validity through the inclusion of questions testing new clinical knowledge and guidelines.

In June 2013 an AKT 'fairness project' was conducted by AKT core group members, the concept being to recruit doctors trained overseas and of non-white ethnic origins from as wide a background as possible, and show them AKT questions to see if they could identify any bias [30]. They would then be shown questions where there was a wide gap in performance to see if they could see any reasons why the white UK-trained doctors were doing better. The eleven participants had all taken and passed the AKT within the past 5 years and were all working in general practice in a range of roles. This summary of results was:

- That overall the AKT was seen as fair
- That IMGs need to identify their areas of need and rectify them to practise medicine in a different country
- That research and statistics are areas where IMGs may need specific support in training
- The questions with the greatest differential performance were not identified by the group.

In 2014, the AKT Core Group also commissioned a piece of work by Chris Sinclair, Department of Modern Languages, University of Southampton, looking at cultural and linguistic features of the test items [36]. This consultant report revealed no linguistic features that could be considered likely to either advantage or disadvantage any particular individual or group of test-takers.

- No linguistic difficulties were observed in any of the items with the biggest differential score between test takers with a primary medical qualification in the UK compared to those with a primary medical qualification from outside the UK. These items were judged to be at an IELTS level of 6.5-7.5.
- The reading load was considered reasonable for test-takers with the expected level of English language proficiency and was not considered excessive for the time allowed.

* * *

Note: Fairness-related research on MRCGP Examinations prior to the new MRCGP

Research into making assessment fair did not begin with the new MRCGP, but has taken place across all modules of the previous examinations. A number of papers were published concerning the previous version of the MRCGP [5, 9, 10, 12, 15, 17, 18, 19, 39, 40, 41].

* * *

References

(in alphabetical order)

1. Atkins S, Roberts C and Hawthorne K. Sociolinguistic factors affecting the performance in the MRCGP Clinical Skills Assessment - are poor candidates that easy to spot? Poster, AMEE Milan 2014.
2. British Medical Association, Examining equality- A survey of royal college examinations. BMA Science and Education department and the Equality and Diversity Committee, May 2006 http://bmaopac.hosted.exlibrisgroup.com/exlibris/aleph/a21_1/apache_media/1DGXLP73Y5LNL6FFLIGRBAH2LIEJSN.pdf
3. British Medical Association,. Examining equality- A survey of royal college examinations Progress review. BMA Science and Education department and the Equality and Diversity Committee, January 2014

http://bmaopac.hosted.exlibrisgroup.com/exlibris/aleph/a21_1/apache_media/AI4U4LYMLK12CDB353ID8PLBAS6XXD.pdf

4. Cobb J. Qualitative analysis of case challenge and reliability in the Clinical Skills Assessment of the MRCGP. Poster presentation, AMEE Annual Conference, Glasgow, September 2010.
5. Coombe M, Freeman A, Dixon H, Wakeford R. The consulting skills module of the MRCGP examination: an evaluation of candidates' views. *Education for Primary Care* 2008; 19, 24-31.
6. Denney ML & Wakeford R, Do individual roleplayers systematically bias the outcome of a high-stakes postgraduate OSCE, in terms of candidate ethnicity? Paper presented at the Ottawa conference, 2014
7. Denney ML, Freeman Adrian, Wakeford R. MRCGP CSA: Are the examiners biased, favouring their own by sex, ethnicity, and degree source? *BJGP*, November 2013, pp. e718-e725(8).
8. Denney ML, Rughani A, Foreman P. Selecting examiners for a national family medicine examination (MRCGP): Matching tools to required examiner competencies, averting deselection of inadequate examiners. Poster presentation Ottawa Conference, Kuala Lumpur, March 2012
9. Denney ML, Wakeford R, Hawthorne K et al. Experiences of simulated patients and candidates in the MRCGP Simulated Surgery examination. *Education for Primary Care* 2006: 17; 354-361
10. Denney ML, Wakeford R, Whitehorn M. Quality Enhancement of Oral Examinations in Specialty Accreditation Paper presented at the AMEE Annual Conference, Malaga, Spain, August 2009
11. Denney ML, Wakeford R. Are older trainees wiser? Relating MRCGP candidates' OSCE performance to the time since their first medical qualification. Paper presented at AMEE 2011, Vienna.
12. Denney ML, Wakeford R. Does performance on MCQ and free response written tests vary differentially by candidate gender and ethnicity? So can choice of 'test modality' be discriminatory? Paper presented at the Annual Ottawa Conference on Assessment, Miami, May 2010)
13. Denney ML, Wakeford R. The Structure of Failure - an analysis of the patterns of consulting behaviour amongst failing candidates in a high-stakes postgraduate OSCE towards enhancing the quality of candidate feedback (Research Paper given to the AMEE Annual Conference, Glasgow, Scotland, September 2010)
14. Dixon H, Blow, Milne P, Siriwardena N, Milne H, Elfes C. Quality Assurance of the Applied Knowledge Test (AKT) of the MRCGP examination. An immediate post-test questionnaire evaluation of the candidates' views. *Education for Primary Care*, in press.
15. Dixon H, Blow C, Irish W, Milne P, Siriwardena AN. 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; 165-72.
16. Dixon H, Blow C, Milne P, Siriwardena N. Reliability of non-prettested versus prettested questions in the applied knowledge test (AKT) of the MRCGP: evidence of quality assurance. *Education for Primary Care* (2014) 25: 149-54
17. Dixon H. Candidates' views of the MRCGP examination and its effect upon approaches to learning. A questionnaire study in the Northern Deanery *Education for Primary Care* 2003: 14; 146-157
18. 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* 2006: 16(6); 655-662.
19. Dixon H. The Purpose and Process of Assessing General Practitioner Registrars. A Qualitative Study of Stakeholders' Views. Ed. D Thesis, University of Newcastle, 2006

20. Esmail A, Roberts C. Independent Review of the Membership of the Royal College of General Practitioners (MRCGP) Examination, GMC 27 September 2013 http://www.gmc-uk.org/education/education_news/23425.asp
21. Foreman P, Denney ML. 'Potential New Examiners (PNE) for the Clinical Skills Assessment Exam- where do they go wrong'. Poster, AMEE Conference, Glasgow, 2010
22. Foreman P. The effects of roleplayer-candidate interactions on fairness in the Clinical Skills Assessment. Poster, AMEE Conference, Prague, 2013.
23. General Medical Council. The state of medical education and practice in the UK 2011. General Medical Council (2011). Available from: <http://www.gmc-uk.org/publications/10586.asp>
24. General Medical Council. Annual Specialty Report - RCGP 2010/11. Available from: http://www.gmc-uk.org/RCGP_ASR.pdf_48265538.pdf
25. McManus I, Wakeford R. PLAB and UK graduates' performance on MRCP(UK) and MRCGP examinations: data linkage study. *BMJ* 2014; 348.
26. Patterson F, Denney M, Wakeford R, Good D. Fair and equal assessment in postgraduate training? A future research agenda *Br J Gen Pract* 2011 December; 61(593): 712–713.
27. Pope L, Hawkrigde A, Simpson R. Performance in the MRCGP CSA by candidates' sex: differences according to curriculum area. *Educ Prim Care*. 2014 Jul;25(4):186-93.
28. Roberts C, Atkins S & Hawthorne K. Performance Features in Clinical Skills Assessment: Linguistic and Cultural Factors in the MRCGP exam. Centre for Language, Discourse & Communication King's College London, 2014.
29. Roberts C, Sarangi S, Southgate L, Wakeford R, Wass V. Oral examinations—equal opportunities, ethnicity, and fairness in the MRCGP. *British Medical Journal* 2000: 320; 370-375
30. Royal College of General Practitioners. AKT "fairness" project. Internal publication MRCGP AKT Core Group, 2013.
31. Royal College of General Practitioners. External Review Report nMRCGP. Dr Neil Munro, Medical Director of Assessment. RCGP (2008, unpublished).
32. Royal College of General Practitioners. MRCGP exam overview. Available from: <http://www.rcgp.org.uk/gp-training-and-exams/mrcgp-exam-overview.aspx>
33. Royal College of General Practitioners. MRCGP Annual Reports. Available from: <http://www.rcgp.org.uk/gp-training-and-exams/mrcgp-exam-overview/mrcgp-annual-reports.aspx>
34. Russell D, Etherington C, Hawthorne K. How can simulated patients' experiences suggest ways to improve candidate performance in the MRCGP clinical assessments? *Education for Primary Care* 2012 23(6):391-8
35. Russell D, Simpson R, Rendel S. Standardisation of role players for the Clinical Skills Assessment of the MRCGP. *Education for Primary Care* 2011: 22; 166–70
36. Sinclair C. Investigation into potential sources of linguistic or cultural bias in the AKT, University of Southampton, April 2014
37. Siriwardena AN, Irish B, Asghar ZB, Dixon H, Milne P, Neden C, Richardson J, Blow C. Comparing performance among male and female candidates in sex-specific clinical knowledge in the MRCGP. *Br J Gen Pract* 2012 Jun; 62(599): e446-50. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3361125>
38. Siriwardena N, Dixon H, Blow C, Irish B, Milne P. Performance and views of examiners in the Applied Knowledge Test for the nMRCGP licensing examination. *Br J Gen Pract*. Feb 1, 2009; 59(559): e38–e43.

39. Wakeford R, Farooqi A, Rashid A, Southgate, L. Does the MRCGP examination discriminate against Asian doctors? *British Medical Journal* 1992: 305; 92-94.
40. Wakeford R, Foulkes J, McManus C et al. MRCGP pass rate by medical school and region of postgraduate training. *British Medical Journal* 1993: 307; 542-543.
41. Wakeford R, Southgate L, Wass V. Improving oral examinations: selecting, training, and monitoring examiners for the MRCGP. *British Medical Journal* 1995: 311; 931-935.
42. Wakeford R, Theatrical dove or hawkish wassock? The consistency of examiner behaviour in a national postgraduate clinical examination. Paper presentation, Ottawa conference, April 2014
43. Wakeford, R. International medical graduates' relative under-performance in the MRCGP AKT and CSA examinations. *Education for Primary Care* May 2012 23(3):148-152(5).
44. Wakeford R, Denney ML, Ludka-Stempien K, Dacre J and McManus IC. Cross-comparison of MRCGP & MRCP(UK) in a database linkage study of 2,284 candidates taking both examinations: assessment of validity and differential performance by ethnicity. *BMC Medical Education* 2015, 15:1
45. Rendel S, Foreman P, Freeman A. Licensing exams and judicial review: the closing of one door and opening of others? *Br J Gen Pract.* 2015 Jan; 65(630):8-9.

* * *