To consider

Responding to the review of the Professional and Linguistic Assessments Board test

Issue

1. The Working Group has completed the review of the Professional and Linguistic Assessments Board test and its conclusions and recommendations are set out in the Review Report, at Annex A.

Recommendations

2. The Strategy and Policy Board is asked to:

   a. Consider the Working Group’s conclusions and recommendations in relation to its review of the Professional and Linguistic Assessments Board test.

   b. Consider the prioritisation of workstreams arising from the recommendations.

   c. Note the proposed next steps.
Responding to the review of the Professional and Linguistic Assessments Board test

Issue

At its meeting on 7 April 2011, Council agreed terms of reference for the review of the Professional and Linguistic Assessments Board (PLAB) test. It tasked an independently-led Working Group to undertake the review, which has now concluded.

Recommendations

The Working Group has made 23 recommendations which are set out at pages A37-39 of the Report at Annex A. These include:

a Limiting attempts at either part of the PLAB test to four and allowing further attempts only in exceptional circumstances or if candidates have demonstrated remediation.

b Giving candidates more informative feedback on performance.

c Reducing the currency of passes in both parts of the test from three years to two years.

d Changing the name of the test and promoting its purpose to increase stakeholders’ (and in particular employers’) understanding of it.

e Extending the scope of the PLAB test to values and principles in *Good medical practice* that we cannot test because of the test’s current format.

f Retaining the methodologies used to standard set the test but monitoring developments in examination and assessment practice to make sure the PLAB test remains up-to-date.

g The need to increase the reliability of the Part 2 Objective Structured Clinical Examination (OSCE). The Working Group has suggested approaches we might consider to achieve increased reliability.

h The need to continue to improve the gender balance of the OSCE examiner pool.

i The need for further work to investigate the reasons for the differential outcomes in the MRCP (UK) and MRCGP examinations and the ARCP process for PLAB candidates relative to their UK peers.

j Increasing further our International English Language Testing System (IELTS) requirements as evidence of English language capability.
The Working Group concluded that it was premature to decide whether it would be appropriate to use the PLAB test to deliver a national licensing examination. This is because of the implications of any decision to change the timing of full registration for UK graduates (the proposal that it should arose during the review); and the need first to agree the purpose and objectives of a national assessment before deciding how best it should be delivered.

Next steps

Reporting the outcome to Council

Council will consider the Working Group’s Report and recommendations at its meeting on 25 September 2014.

Publication

In the meantime, our Publications team will prepare the Report for publication and we are developing a supporting communications plan. We will disseminate the Report electronically to the organisations and groups who contributed to the Review. We plan to publish the Report on our website following the Council meeting on 25 September 2014.

Implementation

We will develop an implementation plan and engagement strategy to take forward the required work when Council has considered the Report.

We will need to prioritise workstreams when we take forward the Working Group’s recommendations. ‘Quick wins’ might include limiting attempts and reducing the currency of passes, but we will need to consider the scope of any transitional arrangements. Other workstreams, such as extending the test’s scope and improving the reliability of Part 2 OSCE will need time to develop, pilot and implement. The Board’s views on priorities will help formulate the implementation plan.

Development and implementation costs will need to be factored into future budgets.
Supporting information

How this issue relates to the corporate strategy and business plan

11 Strategic aim 1 of the Corporate Strategy: Make best use of intelligence about doctors and the healthcare environment to ensure good standards and identify risks to patients.

How the issues support the principles of better regulation

12 The recommendations focus on steps we can take both to increase the robustness of the test as a means of identifying whether candidates have the knowledge and skills for practice in the UK and ensuring that all key interests have continued confidence in it.

What engagement approach has been used to inform the work (and what further communication and engagement is needed)

13 We called for written evidence to seek key interests’ views on key tasks in the terms of reference. We also held a series of roundtable stakeholder discussion meetings across the UK, two roundtables with members of the public (Cardiff and London), a roundtable with nurses and midwives (Northern Ireland), and a meeting with a group of refugee doctors (Salford). We also sought candidates’ views through a survey. The Working Group took this feedback into account when reaching its conclusions and recommendations. Our implementation plan will include further engagement and consultation required to support the future work programme.

What equality and diversity considerations relate to this issue

14 The purpose of the review was to ensure that the PLAB test continues to command the confidence of all key interest groups by being an objective, fair, non-discriminatory and fit for purpose method of assessing the knowledge and skills of international medical graduates applying for registration. Only International Medical Graduates (IMGs) have to take the PLAB test, and therefore only these individuals would be impacted by recommendations for change. Data shows that the IMG cohort comprises mostly doctors that would be classed as black and minority ethnic (BME). Therefore any changes to the PLAB test will have a differential treatment for BME doctors.

15 We have undertaken an equality analysis to identify anticipated impact for IMGs sharing protected characteristics if Council accepts the Working Group’s recommendations. We will develop an action and monitoring plan when Council has decided which recommendations to accept.

If you have any questions about this paper please contact: Jane Durkin, Assistant Director - Registration, jdurkin@gmc-uk.org, 0161 923 6685.
The Working Group’s Report
Review of the GMC’s Professional and Linguistic Assessments Board (PLAB) test: Final Report
Executive summary

1 The PLAB test is designed to test candidates’ ability to practise medicine at the level expected at the end of the first year of Foundation Programme (F1) training. This is when graduates of UK medical schools are granted full registration.

2 The GMC reviews the PLAB test from time to time to make sure that it remains fit for purpose. The purpose of our review was to ensure that the test continues to command the confidence of the GMC’s stakeholders by being an objective, fair, non-discriminatory and fit for purpose method of assessing the knowledge and skills of international medical graduates applying for registration with the GMC.

3 Our terms of reference fell under four broad themes: ensuring standards, content, confidence and outcomes.

Ensuring standards and content

4 The PLAB test’s blueprint, which is in effect the test’s curriculum, is mapped appropriately against Good medical practice, the outcomes expected at the end of F1 training in The Trainee Doctor, and the Foundation Programme curriculum. However, the two-part format of the test limits the range of values and principles in Good medical practice that can be tested. We therefore recommend that the GMC seeks to extend the scope of the PLAB test to include an assessment of the values and principles that cannot currently be tested. This could be achieved, for example, through a situational judgement test but the GMC will need to explore the most appropriate way of introducing this extended assessment.

5 We were concerned that employers do not fully understand the purpose of the test, what it can and cannot assess, and the level at which it is set. The GMC should therefore give greater prominence to the blueprint and the overarching statement, change the name of the test to more accurately reflect its purpose, and promote the test with employers to increase their understanding of it.

6 Candidates are allowed to take Part 1 of the PLAB test as often as they need to pass but they must pass Part 1 again if they do not pass Part 2 within four attempts. We recommend that the GMC impose a limit of four attempts at both Part 1 and Part 2 to maintain confidence in the test. Further attempts should only be allowed if circumstances beyond candidates’ control affect performance or candidates first demonstrate remediation. The GMC should give candidates more informative feedback to help them prepare better for resits, and should develop a framework for candidates to demonstrate remediation.

7 Some candidates will be working as doctors outside the UK while they progress through the PLAB process and register with the GMC. Some, however, will not
and, given the risk to patient safety of skills decay, we concluded that the GMC should reduce the timeframe for the currency of PLAB passes from three to two years. This two-year timeframe mirrors the UK Foundation Programme Office’s requirement that applicants for F1 training take a clinical assessment if they graduated more than two years previously; and the requirement for GPs who want to join a performers’ list but have been out of general practice for more than two years to take a clinical assessment.

8 The GMC’s Council amended our terms of reference to ask us to consider the feasibility of the PLAB test delivering a national licensing examination for UK graduates towards the end of F1 in the event that they decide that a national assessment is required at that stage of training. We concluded that it was premature to decide whether the PLAB test is the appropriate way of delivering a future national licensing examination. During our review, bringing forward full registration for UK graduates has been proposed. Furthermore, those responsible for medical education and training in the UK have yet to agree the purpose and objectives of a national licensing examination. Both issues must be resolved before a decision on the most appropriate way of delivering a national assessment can be agreed.

Confidence

9 The methodologies used to set the standard and pass marks are widely used in medical education and training around the world and are supported by evidence. The GMC should continue to use them but review evidence of others used in a medical context as it emerges. The GMC should consider using Item Response Theory and statistical equating to support the current standard setting and scoring methodologies.

10 The Part 1 examination has a consistently high reliability and works well. The GMC should, however, seek to improve the reliability of Part 2. We have suggested a range of options, for example by increasing the number and/or length of the OSCE stations. The results of a generalisability study will help the GMC to decide how best to increase reliability of the Part 2. The GMC will, however, have to balance the competing interests of increasing reliability and the logistical feasibility of, for example, increasing the number and/or length of stations.

11 In the interests of patient safety, the GMC should continue to apply 1 standard error of measurement to the Part 2 examination, given the risk to patients of false positive candidates. The GMC should, however, explore options for reducing the possibility of candidates compensating between practical and examination skills stations on the one hand and communication and history taking stations on the other hand.

12 The GMC has a robust, objective and fair procedure for dealing with requests for reasonable adjustments for the purposes of enabling candidates to take the
PLAB test. The GMC should review the procedure and its guidance for candidates in due course to ensure that it reflects relevant learning from the 2012 health and disability review.

13 The GMC also has procedures and processes in place to deal with candidates’ complaints about the PLAB test. Since our review began, the GMC has introduced a formal appeal procedure. We suggest that the GMC regularly reviews appeal outcomes to identify and disseminate trends, lessons learned and development work required.

14 We were concerned by the underrepresentation of women in the Part 2 examiner pool. The GMC has 135 examiners, of whom only 20 are women. The GMC has already undertaken work to understand better the barriers that female doctors face in becoming examiners. We recommend that the GMC seeks to increase further the number of female examiners through the continued use of targeted recruitment campaigns.

15 The GMC provides sufficient information about the PLAB test to help candidates apply and prepare for it; and signposts them to other organisations’ websites for information about job opportunities, the cost of living in the UK and immigration requirements. We suggest that the GMC regularly reviews the information and guidance it provides in the light of candidates’ ongoing feedback.

Outcomes

16 Research commissioned for the review found that doctors who passed the PLAB test underperform in the MRCP (UK), the MRCGP and the ARCP process relative to UK qualified doctors. However, the purpose of the PLAB test is not to identify whether candidates have the potential to achieve equivalent outcomes as UK graduates in postgraduate medical education and training or through medical career pathways. The PLAB test is designed to test candidates’ ability to practise medicine at the level expected at the end of F1 training.

17 The reasons for the differential outcomes are not known. They might reflect the fact that international medical graduates face barriers that UK graduates do not, for example taking longer to find employment and a tendency to follow less structured career pathways. Furthermore, the issue is not unique to the UK. There is evidence of underperformance in postgraduate examinations of a similar magnitude for international medical graduates elsewhere, for example in Canada.

18 We were concerned by the evidence of differential outcomes. However, without an explanation for the reasons, we were unable to recommend a definitive, appropriate and proportionate course of action. We therefore recommend that the GMC and stakeholders investigate further to identify why PLAB candidates’ underperform in these postgraduate examinations and in the ARCP. If and
when the reasons are identified, the GMC should consider any changes required to the purpose and standard of the PLAB test.

19 Before candidates are able to sit the PLAB test they must demonstrate their knowledge of English. They can do this by achieving scores in the academic version of the International English Language Testing System (IELTS) required by the GMC. The research commissioned for the review found that IELTS scores have predictive validity for ARCP outcomes. We therefore recommend that the GMC considers increasing its IELTS requirements.

Conclusion

20 We believe that the combination of our recommendations on extending the scope of the PLAB test, improving the reliability of Part 2, limiting attempts, and reducing the currency of passes in the PLAB test, taken together, should help to ensure continued confidence in the PLAB test, pending a decision on whether to introduce a UK national licensing examination. We suggest that the GMC publishes a progress report on taking forward our recommendations in 12-18 months.
Section 1 - Background

Why candidates take the PLAB test

21 All doctors must be registered and licensed with the GMC if they want to work as doctors in the UK. How doctors are registered and licensed depends, in general terms, on their nationality and where they qualified. Doctors who, broadly speaking, graduate outside the UK and the EEA are known as international medical graduates.

22 Before the GMC will register and license them, international medical graduates must show that:

- They have the necessary knowledge of English language1. Most do this by achieving the GMC's required scores in the academic version of the International English Language Testing System (IELTS). These are at least 7.0 in each domain (reading, writing, speaking and listening) with an overall score of at least 7.52.

- They have the knowledge and skills required for medical practice in the UK3. Most do this by passing the GMC's PLAB test.

The PLAB test's purpose

23 The PLAB test is designed to test international medical graduates' ability to practise medicine at the level expected at the end of the first year of Foundation Programme (F1) training. This is when graduates of UK medical schools are granted full registration. In other words, the GMC expects PLAB candidates to demonstrate that they have the same level of theoretical medical knowledge and practical clinical skills at the same stage of training as UK medical graduates.

The test's format

24 There are two parts to the PLAB test. The first, known as Part 1, is a written test of knowledge applied to the care and treatment of patients. The second, known as Part 2, is an assessment of clinical and communication skills. Only candidates who have passed Part 1 may take Part 2.

Section 2 - Remit of the review

25 The GMC reviews the PLAB test from time to time to make sure that it remains fit for purpose.

---

1 Section 21B(1)(d) of the Medical Act 1983.
2 In June 2014 the GMC increased the overall score required to 7.5 from 7.0.
3 Section 21B(1)(b) of the Medical Act 1983.
The purpose of this review was to ensure that the test continues to command the confidence of the GMC's stakeholders by being an objective, fair, non-discriminatory and fit for purpose method of assessing the knowledge and skills of international medical graduates applying for registration with the GMC. The review's terms of reference are at Annex A4.

Section 3 - Working methods

In April 2011 the GMC established an independently-led working group to undertake the review (membership at Annex A). Our membership included representatives of key stakeholder groups, a licensed doctor who had taken the PLAB test, and two experts to advise the working group.

The working group met nine times between November 2011 and March 2014. Our discussions were informed by a range of evidence, including:

- A literature review5. This examined evidence on the number of times candidates are allowed to sit professional examinations (UK, Europe and elsewhere) and assessments and the periods of validity (in other words the currency) of passes in these tests. It also reviewed examination and assessment methodologies and best practice.
- A review of the PLAB test's blueprint, which is, in effect, the PLAB test's curriculum.
- An analysis6 by Professor Chris McManus7 and Mr Richard Wakeford8 of performance in the PLAB test and the membership examinations of the Royal College of Physicians (MRCP (UK)) and the Royal College of General Practitioners (MRCGP).
- A primary research project9. This analysed data on performance in the PLAB test, the Annual Review of Competence Progression10 (ARCP - see glossary of terms at Annex C) and IELTS. It also examined GMC fitness to practise allegations and outcomes for these doctors. We refer to this as the Durham report.
- An analysis of anonymised Part 2 data.

---

4 In April 2012 the GMC's Council amended the terms of reference to ask us to include consideration of the feasibility of the PLAB test delivering a national licensing examination if one is introduced.
5 Insert hyperlink to publication on website.
6 Insert hyperlink to publication on website.
7 Professor McManus is educational advisor to the MRCP (UK) and a member of the PLAB review working group.
8 Richard Wakeford is psychometric/assessment consultant to the MRCGP.
9 Insert hyperlink to publication on website.
10 ARCP is a process that provides a formal and structured review of evidence to monitor doctors' progress throughout each stage of medical training.
A GMC in-house literature review on the challenges faced by international medical graduates and junior doctors transitioning to the UK workplace.

Feedback from stakeholders on key tasks in our terms of reference. We received this through: a call for written evidence, roundtable discussion groups in the four UK countries with stakeholders, two roundtable discussion groups with members of the public (Cardiff and London), a roundtable discussion group with nurses and midwives (Northern Ireland), a meeting with a group of refugee doctors and a survey of PLAB test candidates. We have referred to feedback relevant to our conclusions and recommendations throughout this report.

A presentation by an expert on the theory and development of situational judgement testing.

A presentation by a former Director of the National Recruitment Office for GP specialty training on the use of situational judgement tests for selection to GP specialty training.

29 The GMC’s main statutory objective is to protect and maintain the health and safety of the public. Our overriding concern has therefore been to ensure that the PLAB test continues to assess the knowledge and skills of international medical graduates to the standard necessary for practising as a doctor in the UK. We have also borne in mind at all times our public sector equality duty under section 149 of the Equality Act 2010.

Section 4 – Conclusions

30 This section sets out our conclusions and recommendations under the four themes in our terms of reference. A full list of our recommendations is at Annex B. A glossary of terms used is at Annex C.

Themes 1 and 2 – Ensuring standards and content

What the PLAB test can and cannot examine – the blueprint

31 International medical graduates who take the PLAB test have no contact with the GMC in an educational sense. Their performance in the test is a snapshot of their performance on the day they take it, rather than a measure of how well they have mastered learning objectives.

32 The blueprint is in effect the PLAB test’s curriculum. It sets out the scope and content of the test in terms of the topics, skills and procedures that a doctor who passes the test needs to know and be able to do. It includes what the PLAB test can and cannot examine.
The blueprint, information for candidates on how to interpret and use it, and links to source documents and reference material are published on the GMC's website.\(^{11}\)

The PLAB test is set at the level of successful completion of F1 training. The blueprint is therefore mapped against the Foundation Programme curriculum, the outcomes expected at the end of F1 training in *The Trainee Doctor*, and the GMC's core guidance, *Good medical practice*.\(^{13}\) It also takes into account UK hospital episode statistics and the NHS Read Codes (so that the presentations in it accurately reflect the workload of a doctor in the UK); and the Royal College of General Practitioners’ training curriculum (so that testing on common, important or acute conditions and the management of long term conditions seen in primary care are included).\(^{14}\)

Questions for each Part 1 and Part 2 examination are chosen using a sampling grid to make sure that the content is consistent.

**Good medical practice and Foundation Year 1**

We approached reviewing the content of the PLAB test in two ways. We considered whether the blueprint is mapped appropriately against *Good medical practice* and is consistent with the outcomes at the end of F1 training set out in *The Trainee Doctor*. We also asked stakeholders whether they think the PLAB test adequately assesses the values and principles in *Good medical practice*.

We were satisfied that the blueprint is mapped appropriately against *Good medical practice* and the Foundation Programme curriculum; is consistent with the outcomes expected at the end of F1 training; and identifies the broader professional attributes that successful candidates are expected to demonstrate in clinical practice in the UK. We concluded that the blueprint is therefore fit for purpose. Having said that, we recognised that the PLAB test cannot examine some of the professional and ethical attributes identified in the blueprint because of the limitations of its current two-part format. The blueprint’s overarching statement does, however, identify these attributes and makes it clear that although they might not be assessed during the PLAB test, these are qualities expected of doctors attempting and passing the test, and will be appraised in the workplace.

---


\(^{14}\) The blueprint does not include the advanced duties of a general practitioner.

Professional standards and ethics

38 Stakeholders told us consistently that the PLAB test needs greater emphasis on the wider ethical values and principles that underpin good medical practice in the UK. For example: candidates’ attitudinal attributes; patient autonomy; working within limits of competence and seeking advice; working under pressure; working in multi-disciplinary healthcare teams; raising concerns; probity; and how candidates deal with their own health. The challenge of assessing these kinds of values and attributes is, however, a challenge for any medical assessment: it is not unique to the PLAB test.

39 Revalidation will ensure that all doctors working in the UK engage in regular annual appraisal and develop and maintain their knowledge and skills throughout their career. Furthermore, during the time of our review, the GMC has developed a Welcome to UK practice programme for doctors new to UK practice. This aims to raise awareness of the ethical and professional standards expected of doctors practising in the UK through participative events and online tools (ethical scenarios and a self-assessment tool16). We welcome the fact that the GMC has made these online resources available to PLAB candidates to help them prepare for the PLAB test but the GMC cannot make participation in these events or using the online tools a condition of either booking a place on the PLAB test or of registration.

40 Given the known evidence on barriers faced by international medical graduates transitioning to the UK workplace17, it would be unfair to expect them to have the same in-depth knowledge and understanding of the professional values and principles as UK graduates. However, the values and principles not currently examined through the PLAB test because of the limitations of its current format are fundamental for safe, effective and compassionate healthcare in the UK. We concluded that the PLAB test does not therefore currently adequately assess the full range of the professional values and principles in Good medical practice.

**Recommendation:** The GMC should seek to extend the scope of the PLAB test to include an assessment of the wider ethical values and principles in Good medical practice that the current format cannot test.

**How to bridge this gap?**

41 The literature review, which also identified this weakness in the PLAB test’s current format, suggested the GMC consider introducing a situational judgement test (see glossary of terms – Annex C) to bridge the gap. Some stakeholders also suggested this.

---

17 Insert references.
Situational judgement tests are used in the UK as part of selection to Foundation Programme and specialty GP training to test whether candidates have the professional attributes required at the relevant stage of training. On the other hand, some believe that situational judgement tests identify whether candidates know what they should do rather than what they would do if faced with the scenarios in real practice.

We suggest therefore that the GMC explores with stakeholders how best to extend the scope of the PLAB test to include an assessment of the professional values and principles in *Good medical practice* not currently tested, whether through a situational judgement test or another mechanism.

**Recommendation:** The GMC should explore with its key interests how best to extend the scope of the PLAB test to include an assessment of the professional values and principles in *Good medical practice* not currently tested, for example by including a situational judgement test or another mechanism.

Reflecting real life practice

Some stakeholders told us that, in their view, the PLAB test assesses the ability to recall knowledge rather than to apply it. They said it should test ability to apply knowledge in clinical settings more robustly.

We therefore welcome the work the GMC has already undertaken to introduce more real life scenarios in the Part 2 examination, such as:

- **Couplet stations.** These use information identified at one station to inform action at the next station.
- **High-fidelity simulator stations.** These test candidates’ ability to deal with abnormal signs and acutely-ill patients.
- **Professionalism stations** to test ethical principles.
- **Handover stations** to test candidates’ ability to assimilate information and prioritise patients’ needs.

Employers’ expectations

When we discussed the blueprint we were concerned whether employers understand what the PLAB test does and does not assess in its current format and that it is set at the level of end of F1 training. We were also concerned that the current name of the test does not properly reflect its purpose. This could therefore lead to misunderstanding among stakeholders, particularly employers, about what the PLAB test is designed to do.

Employers are, of course, responsible for ensuring that the doctors they employ have the appropriate knowledge, skills and experience for particular posts. They
must not rely on a pass in the PLAB test as evidence of this. However, we agreed that:

- The GMC needs to give greater prominence to the PLAB test’s overarching statement and blueprint to improve stakeholders’ understanding of what the PLAB test can and cannot currently assess.
- The GMC should change the name of the PLAB test to more accurately reflect its purpose.
- The GMC needs to promote the purpose of the PLAB test through regular engagement with employers to raise awareness that the test is set at the level of entry to F2 training.

These are important, practical steps that the GMC can take to improve stakeholder understanding of the PLAB test’s purpose.

**Recommendation:** The GMC should give greater prominence to the PLAB test’s blueprint and its overarching statement to improve general understanding of the current scope of the PLAB test.

**Recommendation:** The GMC should change the name of the PLAB test to reflect more accurately its purpose, for example the GMC’s Knowledge and Clinical Skills Tests.

**Recommendation:** The GMC should promote the PLAB test’s purpose with employers to increase their understanding of the level at which the test is set.

**Number of attempts**

Candidates can sit Part 1 as many times as they need to pass, provided their evidence of proficiency in English remains current. They are then allowed four attempts to pass Part 2, within three years of the date they passed Part 1. If they do not pass Part 2 on their fourth attempt, or within three years, they must pass Part 1 again (and provide further evidence of proficiency in English, if their current evidence is more than two years old) before being allowed further attempts at Part 2.

Stakeholders told us that there should be a limit on the number of resits allowed. There was no consistent view on what the limit should be but they felt that candidates who need multiple attempts are less competent that those who do not. They said that unlimited attempts undermine confidence in the test itself.

Survey respondents also supported a limit, although they had different views on what the limit should be. However, there was also support to maintain the status quo, particularly among candidates who had not passed the test and those who had needed more than one attempt to pass.
We sought to identify, through the literature review, objective evidence underpinning the number of times candidates may sit comparable professional examinations and assessments within and outside UK. The literature review found\textsuperscript{18} that:

- The evidence base for the number of re-sits allowed in comparable examinations and assessments was at best weak and generally absent.
- The number of attempts allowed is part of an examination’s design and limits interact with training programmes, individual specialty requirements and different national cultures.
- Resitting examinations benefits some candidates, particularly candidates who only just fail.
- There is moderately strong evidence to indicate that there is no further benefit after four attempts. In other words, performance tends to plateau after the fourth attempt.

The literature review also highlighted the rising chance of false positives as the number of re-sits increases. In other words, the risk to patient safety of doctors passing the test by virtue of repeated attempts either fortuitously or due to familiarity with the examination material, when they do not in fact have the basic theoretical knowledge and practical clinical skills to practise safely. There are also costs for candidates retaking examinations, such as time and financial investment. The literature review suggested that a limit of four attempts at Part 1 and four attempts at Part 2 is a reasonable compromise between these competing interests.

The Durham report identified trends relevant to whether to limit the number of times candidates may take the PLAB test. For example, even after controlling for age and gender, performance at the first attempt in both Part 1 and Part 2 predicted a reduced likelihood of censure for fitness to practise concerns. The report also identified possible associations between multiple attempts at both parts of the PLAB test and the likelihood of censure for fitness to practise concerns:

- Borderline significance in the difference in likelihood of censure between candidates who took Part 1 three times and those who passed on first attempt.
- A large difference between candidates who passed Part 1 on first attempt and those who had four or more attempts.
- Candidates who took Part 2 three times or more were far more likely to be censured than those who passed on first sitting.

\textsuperscript{18} Insert reference
There is an argument that a candidate meets the standard through achieving the required scores on the day whether at first or subsequent attempt. However, in the light of the evidence identified in the literature review and the findings in the Durham report, we concluded that the GMC should impose a limit of four attempts at both Part 1 and Part 2 of the PLAB test. In our view, these limits strike the right balance between ensuring patient safety, maintaining confidence in the test and managing candidates' expectations.

Medical royal colleges impose limits on the number of attempts at their national professional examinations. In line with guidance endorsed by the GMC, candidates are allowed a maximum of six attempts but there is flexibility for individual colleges to set more stringent limits. Limiting the number of attempts at Part 1 or Part 2 of the PLAB test to four would therefore be consistent with the spirit of the guidelines on limiting attempts at UK postgraduate examinations.

We also agreed, however, that there might be exceptional circumstances beyond candidates' control that affect performance on the day. In the interests of fairness, we recommend therefore that the GMC takes into account exceptional circumstances if a limit on attempts at the PLAB test is introduced.

We also concluded that, in the interests of fairness, candidates who do not pass within the limits should not automatically be precluded from taking the test again. However, we agreed that the GMC should allow a further attempt only if candidates are able first to demonstrate on the basis of robust, objective and verifiable evidence that they have remediated the deficiencies in their performance in the PLAB test over an appropriate period. This is consistent with the GMC's guidance to the medical royal colleges where candidates have exceeded the maximum number of attempts at any component of national professional examinations.

We acknowledged that many of these unsuccessful candidates will be working or undertaking postgraduate training in healthcare systems elsewhere in the world. They might not, therefore, have access to the same level of educational support available to doctors in either professional practice or postgraduate medical education and training in the UK. However, we agreed nonetheless that they should demonstrate that they have engaged in additional learning and professional development over an acceptable period of time before being allowed a further attempt at the PLAB test. This will help to improve the chances of success for candidates who have not passed within the limits we propose and about whom there must inevitably be a question as to whether they are capable of practising safely at the level of entry to F2 training.

The literature review suggested two years as an appropriate timeframe for a period of remediation before allowing a further attempt, although there is no evidence to support this timeframe. We acknowledged the challenge for the
GMC of assessing the objectivity and robustness of evidence that candidates might submit. We also acknowledged that the opportunity to give meaningful feedback in the context of a summative examination where candidates do not have educational supervisors to help them reflect will inevitably be limited. We suggest that, in conjunction with stakeholders, the GMC develop a framework within which candidates can realistically, but objectively and robustly, demonstrate sufficient remediation to allow a further attempt at Part 1 or Part 2.

**Recommendation:** The GMC should impose a limit of four attempts at both Part 1 and Part 2 of the PLAB test. Further attempts should only be allowed if circumstances beyond candidates’ control have affected performance or on the basis of demonstrable remediation over a period acceptable to the GMC.

**Feedback on candidate performance**

61 Providing feedback on performance enables candidates in any examination or assessment (both successful and unsuccessful) to reflect on performance and identify future learning needs. The GMC currently provides PLAB candidates with their Part 1 scores and, for unsuccessful Part 2 candidates, a breakdown of grades awarded in the OSCE stations.

62 The majority of survey respondents and a group of refugee doctors told us, however, that they would like better feedback on their performance in the PLAB test. This is because it would help them to prepare better for resits and to identify areas for continuing professional development in the workplace. There is therefore a desire for a greater level of feedback among candidates.

63 We concluded that the GMC needs to provide more meaningful feedback on candidate performance to help candidates identify future learning needs. This may also help mitigate the impact of a limit on the number of attempts for the minority of candidates who need more than four attempts to pass either part of the PLAB test.

**Recommendation:** The GMC should provide more meaningful feedback on performance to PLAB test candidates.

**Timeframes**

64 Successful Part 1 candidates must pass Part 2 within three years of passing Part 1. If they do not, they must re-sit and pass Part 1 again, as well as providing further evidence of their proficiency in English if their current evidence has expired. And, generally speaking, successful Part 2 candidates must apply for registration and a licence to practise within three years of passing Part 2.

65 It is not clear why some candidates do not apply for registration and a licence to practise in the UK immediately, or relatively soon, after passing Part 2, given
that the PLAB test is designed to assess knowledge and skills for the purposes of registration.

66 Stakeholders told us that the current timeframes are too long. They were concerned about the currency of the knowledge and skills of some PLAB candidates when they enter the UK workplace. There was, however, no consensus on what the timeframes should be.

67 Some survey respondents also supported shorter timeframes but, again, there was no consensus on appropriate timeframes. A significant minority said, however, that there should be no time restrictions. Respondents who were not registered and licensed with the GMC expressed a stronger preference for no time limit than those who were.

68 The literature review sought to identify evidence underpinning the currency of passes in comparable examinations and assessments in the UK and elsewhere. It found that currency is most often tied to individual examination structures and design, and can also be linked to training structures. This makes it difficult to generalise on good practice. It also found extensive evidence on skills decay, which depends on a range of variables. For example: skills required for complex tasks decay quicker than those required for simpler tasks; theoretical knowledge might decay more slowly than practical skills; the confidence and experience of the individual; the nature of initial training and feedback; and acquiring new knowledge and relinquishing previously learned knowledge.

69 Some candidates will be engaged in medical practice while progressing through the PLAB process and during the time between passing Part 2 and registering with the GMC. Some, however, will not and we were particularly concerned about the risk of the currency of their knowledge and skills for patient safety. But the issues are clearly complex; and the literature review concluded that, notwithstanding the evidence on skills decay, there is a lack of significant evidence for deciding appropriate periods of validity of passes in the PLAB test.

70 There is, however, precedent for assessing the clinical competence of doctors who have taken time away from medical education and professional practice in the UK. The UKFPO requires applicants for F1 training to take a clinical assessment if they graduated more than two years previously; and GPs who want to join a performers’ list but have been out of clinical general practice for more than two years must also take a clinical assessment.

71 Given the risk to patient safety of skills decay, we concluded that it is unrealistic for candidates to expect that passes in either part of the PLAB test will not be time-bound. We recommend that the GMC reduces the timeframes for passes to two years. This is because of the precedents identified and the stakeholder support for shorter timeframes.
**Recommendation:** The GMC should require candidates to pass Part 2 of the PLAB test within two years of passing Part 1; and to apply for GMC registration and a licence to practise within two years of having passed Part 2.

**Could the PLAB test deliver a national licensing examination?**

72 The prospect of using existing examinations and assessments to deliver any future national licensing examination is attractive. Using the PLAB test would obviate the need to develop new examinations and assessments, which would have considerable development and implementation costs. We concluded that there are significant questions that need to be answered before deciding whether it would be appropriate to use the PLAB test to deliver a future national licensing examination.

73 The most fundamental question is the future shape of education and training for UK medical students. Proposals to bring forward the point of full registration for UK graduates are under consideration. If full registration is brought forward in this way, the outcomes that the GMC will require medical students to demonstrate at the end of their undergraduate education and training might change. Bringing forward full registration without a period of provisional registration might therefore have implications for the blueprint of a national licensing examination and the standard at which it is set.

74 Secondly, it is premature to decide how to deliver a future licensing examination until all those responsible for medical education and training in the UK have agreed its purpose and objectives. For example, whether the examination is a summative assessment of educational achievement following an undergraduate medical programme; or whether it should assess suitability for entry to postgraduate education and training. Only once these principles are agreed can discussions begin on how to achieve this.

75 It is also important to note that the PLAB test will no longer exist if a national licensing examination is introduced for UK graduates. The PLAB test governance framework would not therefore be appropriate for a UK national licensing examination, although the GMC’s experience of running the PLAB test could inform the development of a national licensing examination. Medical schools, foundation schools, local education and training boards, deaneries, and employers will need to be jointly responsible, and accountable, for any future national licensing examination. The experience of professional regulators and licensing boards in other jurisdictions (for example in the US and Canada) would also helpfully inform the development phase.

76 Having said that and on the assumption that no change is made to the point when UK graduates are granted full registration, the PLAB test could, in theory
at least, deliver a national licensing examination for UK graduates towards the end of F1 training:

- The blueprint is fit for purpose in terms of standard and content. It is mapped against Good medical practice, the Foundation Programme curriculum and the outcomes expected at the end of F1 training.
- The standard setting methods for both parts of the PLAB test are well studied in the literature and widely used to set the standards of medical examinations and assessments around the world.
- The methods used to deliver the test (multiple choice examination and objective structured clinical examination (OSCE) – see glossary of terms at Annex C) are widely used in examinations in medical education and training.

77 We agreed, however, that a detailed capacity study would be needed to help understand the pressure on existing GMC resources and to model proposals (and likely costs – capital and operational) of the GMC delivering the OSCE component of a national licensing examination.

78 In 2013 just over 7,500 medical students graduated from UK medical schools. It would not be difficult for the GMC to deliver a multiple choice examination to such a large cohort locally through computer-based testing. However, an initial review of the GMC’s Clinical Assessment Centre’s capacity to meet the additional demand of around 7,500 UK graduates each year suggests that there would be a shortfall in capacity. Furthermore, if demand is compressed into a short period (for example a hypothetical two-month period), logistical problems inevitably become greater. Releasing some 7,500 doctors from their posts over a short period could affect employers’ ability to deliver high quality healthcare, if a national licensing examination is taken towards the end of F1 training.

79 Furthermore, adequate safeguards would have to be developed to ensure consistent standards and to mitigate any risk to patient safety:

- Running the OSCE component of a national licensing examination locally across the UK would introduce variables that could adversely impact on standards, such as inconsistent marking by examiners across different sites, and quality of venue and equipment.
- The GMC and partner organisations would need to develop, expand and maintain the existing Part 1 and Part 2 question banks to ensure that the integrity of both examinations is not compromised. While challenging, this is not insurmountable and could be achieved, at least partly, through sharing and adapting other organisations’ question banks.
- The challenges of recruiting and training sufficient OSCE examiners, and maintaining reliability, should not be underestimated.

80 There will also be principles of fairness and equality of opportunity to consider:
If a national licensing examination took place towards the end of F1 training, candidates sitting it too early during F1 would be disadvantaged: they will not have been afforded the same opportunities for learning and gaining experience, compared to candidates taking it at a much later stage. This would be avoided through all candidates taking the assessment on the same day. But as already noted this would affect delivery of healthcare.

What are the implications for F1 trainees who fail? The GMC and stakeholders would need to consider very carefully the scope for re-sits and an appeal procedure.

Additionally, the law will need to change if full registration for UK graduates is dependent on success in a national licensing examination. There is currently no statutory provision to require UK graduates to pass a national licensing examination before the GMC registers and licenses them.

To conclude, we agreed that it is too early to determine whether the PLAB test could, or should, deliver any future national licensing examination in the UK. This can only be decided once both the future shape of education and training for medical students and newly qualified doctors and the purpose and objectives of any future licensing examination are settled. If a decision is then made that the PLAB test is the appropriate mechanism to deliver a national licensing examination, the GMC and key stakeholders will need to resolve the caveats we have identified with regard to capacity, governance, maintaining standards, and equality and fairness.

**Recommendation:** The GMC and those responsible for medical education and training in the UK should agree the most appropriate way of delivering a national licensing examination when the future shape of education and training for medical students and newly qualified doctors and the purpose and objectives of the examination are settled.

**Theme 3 - Confidence**

This section sets out our conclusions on the scoring and standard setting methods for both parts of the PLAB test, which the terms of reference asked us to review.

**Standard setting**

The literature review concluded that that ‘there is no single universally optimum method of standard setting... The current methods used in PLAB are well studied in the literature, show acceptable properties, and there is no consistently better method under all circumstances’[19]. We therefore agreed that

[19] Insert hyperlink. Page 56 of report
the GMC should retain the Angoff and borderline group scoring methods for setting the standard of the Part 1 examination and the Part 2 assessment respectively.

The literature review highlighted other standard setting methods. It also identified other statistical approaches, such as Item Response Theory, that can support standard setters and help develop a deeper understanding of how items and candidates perform. We therefore agreed that the GMC should consider using Item Response Theory to support the current standard setting and scoring methodologies for the PLAB test. We also agreed that the GMC should regularly review how it sets the standard of the PLAB test as evidence of other methods in a medical context emerges.

**Recommendation:** The GMC should retain the Angoff and borderline group scoring methods for setting the standard of the Part 1 examination and the Part 2 assessment.

**Recommendation:** The GMC should consider using Item Response Theory and statistical equating to support the current standard setting and scoring methodologies for the PLAB test.

**Recommendation:** The GMC should regularly review the PLAB test’s standard setting and scoring methodologies to ensure that it remains up-to-date and in line with evidence of other methods in a medical context.

**Reliability**

We also reviewed the reliability of the PLAB test.

**The Part 1 examination**

Part 1 is a written test of medical knowledge designed to assess the application of clinical knowledge to the care of patients. It is a three-hour multiple choice examination comprising 200 single best answer questions (see glossary of terms at Annex C). It may be taken in the UK or at locations overseas.

This examination method is very widely used in undergraduate and postgraduate medical education and training around the world. The reliability of the Part 1 examination is consistently high at 0.9 and above. The Part 1 examination therefore works well.

**Computer-based support for delivering the PLAB test**

Part 1 candidates mark their answers on answer sheets which are then scanned using an optical mark reader. The questions at any Part 1 examination may
already include the use of images, electrocardiograms and x-rays. However, the literature review noted that computer based testing enables a wider range of assessment approaches\textsuperscript{20}. For example, the use of video or audio clips. They also noted that computer based testing enables much quicker feedback for candidates and a greater range of data analysis. Against these benefits, however, they highlighted that it might not always be feasible to use computer based testing if large numbers of computer stations are required in a wide variety of locations.

We concluded that the GMC should explore the feasibility of introducing computer based testing for the Part 1 examination, given the potential benefits. The wider range of assessment approaches available through computer based testing will introduce even more realistic examination questions. If it is feasible, the GMC will need to consider any impact for candidates with protected characteristics.

We also concluded that the GMC should explore the feasibility of introducing electronic marking for both parts of the PLAB test. This will allow the GMC to provide quicker feedback on performance to candidates and will enable real-time understanding of candidates' performance on Part 1 and Part 2 days.

**Recommendation:** The GMC should explore the feasibility of introducing computer-based testing for the Part 1 examination and electronic marking for both parts of the PLAB test. This would enable the GMC to provide quicker feedback for candidates, use a wider and even more realistic range of assessment techniques, and gather additional intelligence for analysis.

*The Part 2 examination*

Part 2 is an objective structured clinical examination (OSCE) comprising 14 five-minute stations. It is an assessment of practical clinical and communication skills in four domains: communication, examination, history taking and practical skills. It takes place in the GMC's Clinical Assessment Centre and only candidates who have passed Part 1 may take Part 2.

This examination format is well recognised and widely used in both undergraduate medical education and training around the world. We support its use for the Part 2 examination. Examiners are carefully selected and well trained and are committed to the examination. The facilities at the GMC’s purpose-built Clinical Assessment Centre are excellent.
Setting the pass mark

Each station in each Part 2 examination has between three and six objectives. Each objective is weighted, with the total weightings for each station adding up to 100%. Examiners award candidates a grade between A and E for each objective but are not aware of the objectives' weightings. Grades awarded are then converted to a scale of 0-4 and calculated to give an overall station score. The examiner also gives a separate overall judgment of each candidate’s overall performance (pass, borderline or fail). They base this on their professional experience using the notional ‘minimally competent’ doctor at the level of entry to F2 training as a benchmark.

The passing score for each station is then calculated from the mean scores of previous candidates to whom examiners have given borderline judgements. The scores for each station are added, plus one standard error of measurement (SEM), to determine the total score for each Part 2 examination.

The risk of false positives

It is easy to identify good or poor candidates in any examination or assessment. The key challenge is to differentiate between the just passing and just failing candidate. No assessment method will entirely remove false positives – those who pass when they should not. Many examining boards attempt to deal with this, however, by adding a margin to the pass mark so those who might barely pass will fail.

The GMC does not apply any SEM to the Part 1 pass mark: candidates pass or fail based on their raw scores. However, the GMC adds 1 SEM to the Part 2 pass mark to minimise the chances of false positive candidates. This is imperative for patient safety, given that passing the PLAB test is a route to registration and a licence to practise, and thereby access to patients. Removing this margin would undoubtedly lead to candidates passing who should not and therefore put patients at risk of harm.

Applying any SEM may be seen as unfair to borderline candidates. However, we agreed that it is nonetheless in the wider public interest. We concluded that the addition of 1 SEM is a reasonable compromise between the risk of false positives for patients and the consequences of failure for the slightly raised number of false negatives.

**Recommendation:** In the interests of patient safety, the GMC should continue to apply 1 standard error of measurement to the pass mark for the Part 2 examination.
Reducing the risk of compensation

The stations in the Part 2 examination are designed to examine different skills areas: practical and examination skills and communication and history taking skills. To pass, candidates must meet or exceed the total passing score, as well as achieve the passing score in a minimum of nine (of 14) stations. This requirement was introduced to minimise the chances of candidates compensating poor performance in some stations with a very high performance in others.

An analysis of anonymised Part 2 data, however, found that some candidates had compensated poor performance in communication and history skills stations with better performance in the practical and examination skills stations. We were concerned by this, given the importance of communication skills for safe and effective patient-centred care. We agreed that candidates should demonstrate competence across the skills areas and should not be able to compensate overly poor performance in some skill areas with better performance in others.

We concluded therefore that the GMC should explore how to reduce the possibility of candidates compensating between practical and examination skills stations on the one hand and communication and history taking stations on the other hand.

**Recommendation:** The GMC should explore how to reduce the possibility of candidates compensating between practical and examination skills stations on the one hand and communication and history taking stations on the other hand. For example, the GMC should consider whether to require a passing standard of performance in the practical and examination skills stations and also in the communication and history taking stations.

Reliability of Part 2

The reliability of the Part 2 examination is generally around 0.65, which is typical for an OSCE of this length, but nevertheless falls below the widely-accepted minimum of 0.8. This is linked to the number and length of the OSCE stations. Furthermore, it is also recognised that reliability is not the sole measure of an assessment: other considerations such as feasibility and cost-effectiveness are also important.

Improving the reliability of Part 2

We considered ways that might increase the reliability of the Part 2 examination, for example by increasing the number of stations in each OSCE, increasing the length of the OSCE stations or using two examiners in each station rather than one.
Some stakeholders told us that they were concerned that Part 2 candidates need only pass nine of the 14 OSCE stations and that five minutes was too short a time to assess competence in a real-world clinical skill. Some said that candidates should pass all stations but, as the literature review concluded, this is unrealistic.

Increasing the number of stations in each Part 2 examination and the length of each station could increase the reliability of the Part 2 examination. And the literature review sought to identify any evidence on the optimum number of stations that should be included in the Part 2 OSCE. However, the literature review suggested that, to help answer both these questions, the GMC undertake a generalisability study, followed by a decision study (see glossary of terms at Annex C).

We also discussed whether other factors might affect the reliability of the Part 2 examination, such as:

- Whether there is sufficient link between an examiner’s mark for overall judgement and whether that particular candidate passes or fails.
- The fact that examiners do not know the weightings given to each objective or the borderline pass marks.
- The fact that examiners do not know whether the candidates they have assessed pass or fail or the other examiners thought of the candidates on the day.

Introducing a feedback session directly after each Part 2 examination at which examiners see the overall marks given by all examiners to all candidates and discuss discrepancies in marks awarded might help to increase reliability. However, we recognised that the benefit of introducing such an approach needs to be weighed carefully against the risk of perceived collusion or ‘group think’ among examiners. The individual objective judgement of each examiner is key to any OSCE format.

The GMC has commissioned a generalisability study but until the results are known, we cannot recommend the most appropriate approach, or combination of approaches, that the GMC should take to increase the reliability of Part 2.

We therefore recommend that the GMC considers how best to increase the reliability of the Part 2 examination in the light of the results of the generalisability study. In making this recommendation, we recognise that the GMC will have to balance the competing interests of increasing the reliability of the Part 2 examination against the feasibility of administering it if the number, or length, of the OSCE stations were to be increased.

**Recommendation:** The GMC should seek to increase the reliability of the Part 2 examination. Options include increasing the number and/or
length of the OSCE stations, introducing a feedback session after each OSCE for examiners to discuss candidates’ performance, and using two examiners to assess performance. However, the GMC will need to take into account the results of the generalisability study and feasibility when determining the most appropriate and proportionate way of increasing the Part 2 reliability.

**Equality of opportunity**

110 The third theme in the review’s terms of reference also asked us to review whether the PLAB test and its administration, including complaint-handling and associate recruitment, remain compliant with the Equality Act 2010. Given the breadth of this key task, we focussed on equality of opportunity. In particular we considered:

- How the GMC deals with candidates’ requests for reasonable adjustments
- How candidates may seek redress if they think something has gone wrong
- How the GMC recruits PLAB associates and
- Candidates’ access to information about the PLAB test.

**Reasonable adjustments**

111 The Equality Act 2010 requires the GMC to make reasonable adjustments to either part of the PLAB test to ensure that candidates are not disadvantaged by a disability. The examination regulations include provisions for requests for reasonable adjustments as well as for review of a decision not to make a reasonable adjustment.

112 The GMC’s guidance to candidates explains how to request a reasonable adjustment and gives examples of adjustments that it has made for candidates. This ensures that applicants from countries where reasonable adjustments are not routinely made know that they can ask for an adjustment.

113 GMC staff follow a written procedure for dealing with requests for reasonable adjustments, which includes taking into account, where necessary, a report from an appropriate specialist recommending the extent of adjustments required; and provision for the candidate to agree the extent of proposed adjustments.

114 We treated with caution the data on reasonable adjustments made as the number of adjustments is very small when compared with the total candidate population. However, the GMC has not refused a request for a reasonable adjustment supported by an appropriate specialist; and the PLA Board chair has never had to consider a referral from a dissatisfied candidate.

115 The GMC’s approach to dealing with requests for reasonable adjustments is consistent with the approach agreed by the medical royal colleges’ Academy
We concluded that the GMC’s procedures for dealing with candidates’ requests for reasonable adjustments are objective, robust and fair in the context of enabling candidates with disabilities to take the PLAB test.

In 2012 the GMC established the Health and Disability in Medical Education and Training Group to develop a comprehensive picture of the issues disabled medical students and trainees in the UK face during their education and training. In 2013 the GMC took forward work based on the Group’s findings. We welcome the GMC’s commitment to ensure that learning from the health and disability review relevant to the PLAB test is reflected in procedure and candidate guidance on reasonable adjustments.

Recommendation: The GMC should review its procedure for dealing with candidates’ requests for reasonable adjustments and its candidate guidance in due course to ensure that it reflects relevant learning from the 2012 health and disability review.

Seeking redress

Each GMC directorate deals with complaints about the services it provides, with the emphasis on addressing them promptly and effectively, and on making timely adjustments to procedures, where necessary. A dedicated team in the Registration and Revalidation Directorate deals with all complaints about the PLAB test in line with a written procedure. As well as responding to individual complaints, this team analyses and reports on complaint trends to make sure that any lessons learned are fed into service improvement projects.

The examination regulations covering both parts of the PLAB test stipulate that candidates who want to complain should do so in writing within 28 days of their examination. In practice, however, the GMC accepts complaints in alternative formats, for example over the telephone.

The number of complaints received is very low when compared to the number of candidates who take the PLAB test. Only 293 doctors complained between 2009 and 2013.

Candidates may also ask for a clerical check of their final marks. This is an administrative check of whether marks on performance in the written examination or the practical assessment have been transcribed correctly from the completed scripts or the examiners’ score sheets. As with complaints received, the number of requests for clerical checks is very low when compared to the number of candidates. The GMC received just 122 requests for clerical checks between 2009 and 2013. No wrong marks have ever been found.
The GMC therefore has procedures and processes in place to deal with the low number of candidates' complaints and to check that marks have been correctly transcribed. However, we were concerned that neither amounted to a fair process for candidates to appeal against their marks. In the light of our concern, the GMC has since implemented an appeal procedure. We welcome this development and suggest that the GMC regularly reviews outcomes to identify and disseminate any trends in appeals, lessons learned and test development work required.

**Recommendation:** The GMC should regularly review appeal outcomes to identify and disseminate any trends, lessons learned and development work required.

**Associate recruitment**

The GMC uses associates to deliver a wide range of its functions, including the PLAB test. Associates are not GMC employees but are contracted to provide services in areas of their expertise.

The demographic profiles of the PLA Board and the Part 1 and Part 2 Panels (see glossary of terms) gave us no cause for concern. We were concerned, however, that women are underrepresented among the largest group of PLAB associates – the Part 2 examiners. There are 135 examiners, of whom the majority are men (115). There are only 20 female examiners.

Historically associate recruitment was managed locally by individual teams within the GMC. Since 2011, however, the GMC has recruited and selected all PLAB associates against defined competencies in the same way as all other GMC associates through recruitment campaigns overseen by the GMC’s corporate HR function.

The gender balance has improved during our review through better targeting of potential female examiners, for example through raising awareness of recruitment with the Women’s Medical Federation. But the GMC recognises the need for further improvement and has already undertaken work to identify barriers faced by female doctors in becoming examiners. This will help the GMC understand the underlying reasons for female underrepresentation and enable the GMC to put in place an action plan to improve the gender balance of the examiner pool. We therefore recommend that the GMC continues to seek to improve female representation in the examiner pool.

Younger doctors are also underrepresented in the examiner pool, although to a much lesser extent than female doctors. However, we accept that there is a

---

balance to be struck here. Examiners must have experience of assessing performance but can only gain this as their professional experience increases. Appointing examiners who do not have sufficient assessment experience would be self-defeating and put patients at risk.

**Recommendation:** The GMC should seek to increase further the number of female examiners through the continued use of targeted recruitment campaigns.

**Help for candidates**

128 The GMC provides a wealth of information about the PLAB test and advice on how to apply and prepare for it. This includes:

- What the test includes, its format and how it is marked.
- The blueprint together with information on how candidates can use and interpret it, and links to source documents.
- Example Part 1 questions and Part 2 OSCE scenarios.
- A video on the Part 2 OSCE so that candidates know what to expect on the day.
- Links to the GMC’s ethical guidance and interactive learning materials.

129 To help them consider whether life in the UK is right for them, the GMC also signposts candidates to other organisations’ websites which have information about job opportunities, the cost of living in the UK and immigration requirements.

130 Despite this, survey respondents and a group of refugee doctors told us that they would like better and more realistic information about job and training opportunities in the UK, training career pathways and other information about living and working in the UK, including the cost of living. Some also told us that they would like better revision materials, including more example questions and scenarios, and a list of medical text books.

131 We recognise the significant challenges for anyone deciding to move to live and work in another country. It is, however, up to individuals to research fully the implications for them of moving to the UK. And doctors are responsible for their own learning to ensure that their knowledge and skills remain up-to-date. The blueprint sets out the scope of the PLAB test and is available for candidates to use as a guide to the topics on which they might be examined. And the GMC provides information for candidates on how to interpret the blueprint. As in any other summative assessment, it is, however, up to candidates to satisfy the GMC that they have the broad range of knowledge and skills across all areas identified in the blueprint.
We concluded that the GMC provides sufficient information to help candidates prepare for the PLAB test and signposts them appropriately to other organisations which are best place to provide information on living and working in the UK. However, the GMC should regularly review the information and guidance it provides, taking into account ongoing feedback from candidates.

**Recommendation:** The GMC provides sufficient information to help candidates prepare for the PLAB test and signposts them appropriately to other organisations. However, the GMC should regularly review candidates’ feedback to make sure that the information and guidance it provides adequately meets their needs.

**Theme 4 - Outcomes**

The fourth theme in our terms of reference asked us to consider whether international medical graduates granted full registration following a pass in the PLAB test are more or less likely than other cohorts of doctors to experience difficulties in UK medical practice.

We considered this through research into the performance of PLAB candidates relative to UK graduates in specialty examinations and in ARCP, as well as data on fitness to practise cases. The findings of Professor McManus’s and Mr Wakeford’s study and the Durham report informed our discussion.

Professor McManus and Richard Wakeford found that:

- Performance in the PLAB test is a valid predictor of performance in the MRCP (UK) and the MRCGP.
- PLAB candidates did not show outcome equivalence at the MRCP (UK) or the MRCGP at first attempt when compared with the median UK graduates’ performance.
- PLAB candidates underperformed in comparison to UK graduates at the knowledge tests by approximately 1 standard deviation; and at the clinical assessments by approximately 1.5 standard deviations.

The Durham report found that:

- PLAB candidates performed less well at ARCP than UK graduates.
- Based on an analysis of PLAB candidate performance in IELTS, requiring a minimum score of 8.5 in IELTS would achieve bare equivalence in ARCP between PLAB candidates and UK graduates.
The over-representation of PLAB candidates censured in fitness to practise procedures was reduced to an insignificant level when controlled for age and gender differences.

The probability of fitness to practise outcomes is not easily predicted from PLAB scores but might be associated with the number of attempts and, to an extent, overall performance in the PLAB test.

International medical graduates have contributed over the years to the success of the NHS; and they work at all levels of practice in the UK health sector. The evidence of barriers they face on entering the UK workplace is well known\(^25\). The differential outcomes for PLAB candidates might reflect the fact that international medical graduates face barriers that UK graduates do not. For example, taking longer to find employment and a tendency to follow less structured career pathways.

This is not, however, a problem unique to the UK. A Canadian study published in 2011 and data in the Medical Council of Canada’s 2012 Annual Report showed that international medical graduates in Canada (including Canadian nationals who had trained outside Canada) underperformed in the College of Family Physicians of Canada’s and the Royal College of Physicians and Surgeons of Canada’s postgraduate examinations. This underperformance is of a similar magnitude to PLAB candidates’ underperformance in the MRCP (UK) and the MRCGP compared with UK qualified doctors. It is despite entry equivalence through the Medical Council of Canada Qualifying Examination which both Canadian graduates and doctors who trained outside Canada must pass to be able to practise in Canada. The 2011 study identified the need for better orientation and support for international medical graduates in Canada as well as better training for those who train international medical graduates. These messages chime with findings on the need for better support for doctors new to UK practice in the GMC’s State of Medical Education and Practice Report 2011\(^26\).

The purpose of the PLAB test is not to identify whether candidates have the clinical ability and professional attributes required either for postgraduate medical education and training or progression in the workplace. It does not aim to identify whether candidates have the potential to achieve the equivalent outcomes as UK graduates at this higher level or through medical career pathways. The purpose of the PLAB test is to identify whether candidates have at least the same level of knowledge and skills as the notionally minimally competent UK graduate who has completed F1 training. In other words, the

\(^{24}\) Not all PLAB candidates demonstrate English language ability through achieving the scores the GMC requirements in IELTS (academic version).

\(^{25}\) Non UK qualified doctors and Good Medical Practice; The experience of working within a different professional framework, Slowther et al, 2009.

GMC aims through the PLAB test to identify whether candidates demonstrate equivalent outcomes at the point of entry to the register as UK graduates. The standard is set at this level because the GMC considers that it would be unfair to require candidates to demonstrate a higher level of knowledge and skills than UK graduates for the purposes of full registration.

140 The researchers’ findings on differential outcomes for PLAB candidates in specialty examinations and postgraduate training do not of themselves amount to evidence either that successful PLAB candidates are not safe to practise in the UK at the level of entry to F2 training or that the pass marks for both Part 1 or Part 2 are set too low. And the Canadian evidence suggests that entry equivalence is not in any event the solution to solving differential outcomes for international medical graduates in postgraduate examinations and training.

141 We were concerned by the evidence of disparities in outcomes for PLAB candidates compared with UK graduates but the reasons are far from clear. The researchers suggest significant increases to the pass mark for both parts of the PLAB test with a view to ensuring equivalence in outcomes in the MRCP (UK), the MRCGP and the ARCP process.

142 The research findings cannot be ignored but the issues are both complex and not unique to the UK. We were unable to recommend a definitive, appropriate and proportionate course of action as the reasons for PLAB candidates’ underperformance in the MRCP (UK), the MRCGP and ARCP relative to their UK peers appear to be complex. We agreed therefore that the GMC and stakeholders need to investigate further and the GMC should explore with stakeholders any changes that might be required to the purpose and standard of the PLAB test.

**Recommendation:** The research commissioned for the review found that PLAB candidates’ underperform in the MRCP (UK), the MRCGP and ARCP relative to UK graduates. The reasons for this are not entirely clear and appear complex. The GMC and its key interests should investigate further and the GMC should consider any changes that might be required to the purpose and standard of the PLAB test.

**IELTS**

143 In June 2014 the GMC increased its IELTS overall score requirement from 7.0 to 7.5. Given the Durham report’s findings on the predictive validity of IELTS for ARCP outcomes, we agreed that the GMC should consider a further increase in its IELTS requirements.

**Recommendation:** The GMC should consider further increasing its IELTS requirements in the light of the Durham report’s findings on the predictive validity of IELTS for ARCP outcomes.
Conclusion

144 The combination of our recommendations on extending the scope of the PLAB test, improving the reliability of the Part 2 examination, limiting attempts at the PLAB test, and reducing the currency of passes in the PLAB test, taken together, should help to ensure continued confidence in the PLAB test, pending a decision on whether to introduce a UK national licensing examination.

145 The GMC should publish a progress report on taking forward work arising from our recommendations in 12-18 months.

Recommendation: The GMC should publish a progress report on taking forward work arising from our recommendations in 12-18 months.
Terms of reference and membership

Purpose

1 The purpose of the review will be to ensure that the PLAB test continues to command the confidence of all key interest groups by being an objective, fair, non-discriminatory and fit for purpose method of assessing the knowledge and skills of international medical graduates applying for registration with the GMC.

Key tasks

2 The key tasks of the review are to consider and make recommendations to Council in relation to the following broad themes:

Theme 1: Ensuring standards

3 To review whether the knowledge and skills demonstrated by a pass in both parts of the PLAB test continue to be equivalent to those of doctors who have successfully completed the first year of Foundation Programme training.

4 To examine whether there should be a limit on the number of attempts at passing the test in order to ensure public confidence in the standards of practice in the UK.

5 To review whether the periods of validity of passes in both parts of the PLAB test are appropriate.

6 To consider the feasibility of using the PLAB test for the purpose of a national examination towards the end of F1, in the event that Council were in the future to decide that a national assessment was required at that stage of training.

Theme 2: Content

7 To review whether the blueprint (which sets out the scope and content of the PLAB test in terms of the topics, skills and procedures) is consistent with the outcomes for provisionally registered doctors described in The Trainee Doctor (which replaces The New Doctor).

8 To consider the extent to which the PLAB test could (and should) further assess knowledge and application of the values and principles of Good Medical Practice.

Theme 3: Confidence

9 To review the reliability and validity of the PLAB test.

10 This will include examining whether:
a The scoring systems and standard setting reflect best practice.

b The Part 1 written examination and Part 2 objective structured clinical examination stations are fit for purpose and reflect current examination and assessment best practice.

c The test reliably and accurately differentiates between candidates who meet the required standards of both parts of the test and those who do not.

11 To review whether the PLAB test and its administration remain compliant with the requirements of the Equality Act 2010. This will include the processes for appointing panel members and examiners and for dealing with complaints by candidates.

Theme 4: Outcomes

12 To examine whether international medical graduates granted full registration following a successful pass in the PLAB test are more or less likely than other cohorts of doctors to experience difficulties in medical practice in the UK.

13 This will include:

a Examining existing research on the workplace experiences of graduates making the transition to the workplace.

b Examining any evidence of disparity between the success rates of UK medical graduates and international medical graduates in postgraduate examinations and assessments.

c Examining the implications of any further research in respect of 12a-b required as part of the review.

Membership

14 Prof Kathy Boursicot – Assistant Dean for Assessment and Medical Education Research, Lee Kong Chian School of Medicine, Singapore

15 Professor Kenneth Cochran – Chair of the PLA Board

16 Professor Ian Cumming OBE (Chair) – Chief Executive, Health Education England

17 Professor Jane Dacre – GMC Council member (medical)

27 Professor Cochran joined the working group when GMC Council members’ membership ended on 31 December 2012.
18 Dr Sue Davison - GMC Council member (lay)⁲⁹

19 Professor Derek Gallen - Postgraduate Dean for the Wales Deanery, National Director of the UK Foundation Programme Office, Chair of the Conference of Postgraduate Medical Deans

20 Dr Alison Graham – Executive Medical Director, NHS Ayrshire and Arran

21 Ms Roswyn Hakesley-Brown CBE – former Chair of Trustees, Patients Association

22 Dr Abrar Hussain - consultant psychiatrist⁳⁰

23 Professor Chris McManus - Professor of Psychology and Medical Education, University College London

²⁸ Professor Dacre’s membership of the working group ended when her term of office as GMC Council member ceased on 31 December 2012.
²⁹ Dr Davison’s membership of the working group ended when her term of office as GMC Council member ceased on 31 December 2012.
³⁰ Dr Hussain passed the PLAB test in 2004 and has been registered with the GMC since 2005.
Annex B

Recommendations:

- The GMC should seek to extend the scope of the PLAB test to include an assessment of the wider ethical values and principles in *Good medical practice* that the current format cannot test (paragraphs 11-20).

- The GMC should explore with its key interests how best to extend the scope of the PLAB test to include an assessment of the professional values and principles in *Good medical practice* not currently tested, for example by including a situational judgement test or another mechanism (paragraphs 21-23).

- The GMC should give greater prominence to the PLAB test blueprint and overarching statement to improve general understanding of the current scope of the PLAB test (paragraphs 26-28).

- The GMC should change the name of the PLAB test to reflect more accurately its purpose, for example the GMC’s Knowledge and Clinical Skills Tests (paragraphs 26-28).

- The GMC should promote the PLAB test’s purpose with employers to increase their understanding of the level at which the test is set (paragraphs 26-28).

- The GMC should impose a limit of four attempts at both Part 1 and Part 2 of the PLAB test. Further attempts should only be allowed if circumstances beyond candidates’ control have affected performance or on the basis of demonstrable remediation over a period acceptable to the GMC (paragraphs 29-40).

- The GMC should provide more meaningful feedback on performance to PLAB test candidates (paragraphs 41-43).

- The GMC should require candidates to pass Part 2 of the PLAB test within two years of passing Part 1; and to apply for GMC registration and a licence to practise within two years of having passed Part 2 (paragraphs 44-51).

- The GMC and those responsible for medical education and training in the UK should agree the most appropriate way of delivering a national licensing examination when the future shape of education and training for medical students and newly qualified doctors and the purpose and objectives of the examination are settled (paragraphs 52-62).

- The GMC should retain the Angoff and borderline group scoring methods for setting the standard of the Part 1 examination and the Part 2 assessment (paragraphs 64-65).

- The GMC should consider using Item Response Theory and statistical equating to support the current standard setting and scoring methodologies for the PLAB test (paragraphs 64-65).

- The GMC should regularly review the PLAB test’s standard setting and scoring methodologies to ensure it remains up-to-date and in line with evidence of other methods in a medical context (paragraphs 64-65).
The GMC should explore the feasibility of introducing computer-based testing for the Part 1 examination and electronic marking for both parts of the PLAB test. This would enable the GMC to provide quicker feedback for candidates, use a wider and even more realistic range of assessment techniques, and gather additional intelligence for analysis (paragraphs 69-71).

In the interests of patient safety, the GMC should continue to apply 1 standard error of measurement to the pass mark for the Part 2 examination (paragraphs 76-78).

The GMC should explore how to reduce the possibility of candidates compensating between practical and examination skills stations on the one hand and communication and history taking stations on the other hand. For example, the GMC should consider whether to require a passing standard of performance in the practical and examination skills stations and also in the communication and history taking stations (paragraphs 79-81).

The GMC should seek to increase the reliability of the Part 2 examination. Options include increasing the number and/or length, of the OSCE stations, introducing a feedback session after each OSCE for examiners to discuss candidates’ performance, and using two examiners to assess performance. However, the GMC will need to take into account the results of the generalisability study and feasibility when determining the most appropriate and proportionate way of increasing the Part 2 reliability (paragraphs 82-89).

The GMC should review its procedure for dealing with candidates’ requests for reasonable adjustments and its candidate guidance in due course to ensure that it reflects relevant learning from the 2012 health and disability review (paragraphs 91-97).

The GMC should regularly review appeal outcomes to identify and disseminate any trends, lessons learned and development work required (paragraphs 98-102).

The GMC should seek to increase further the number of female examiners through the continued use of targeted recruitment campaigns (paragraphs 103-107).

The GMC provides sufficient information to help candidates prepare for the PLAB test and signposts them appropriately to other organisations. However, the GMC should regularly review candidates’ feedback to make sure that the information and guidance it provides adequately meets their needs (paragraphs 108-112).

The research commissioned for the review found that PLAB candidates’ underperform in the MRCP (UK), the MRCGP and ARCP relative to UK graduates. The reasons for this are not entirely clear and are complex. The GMC and its key interests should therefore investigate further and the GMC should consider any changes that might be required to the purpose and standard of the PLAB test (paragraphs 113-122).
The GMC should consider further increasing its IELTS requirements in the light of the Durham report's findings on the predictive validity of IELTS for ARCP outcomes (paragraph 123).

The GMC should publish a progress report on taking forward work arising from our recommendations in 12-18 months (paragraphs 124-125).
Annex C

Glossary of terms

Angoff
A method of standard setting based on group judgments about the performance of hypothetical borderline ('just passing') candidates. The GMC uses Angoff to set the pass mark for each Part 1 examination.

ARCP
Annual Review of Competence Progression (ARCP) is the process which scrutinises the suitability of doctors in training to progress to the next stage of, or to complete, a training programme. It is usually held annually, but some specialties have more frequent reviews in the early years of training. Foundation Programmes have incorporated an ARCP since 2013.

Borderline group scoring
The GMC uses the borderline group scoring method to set the pass mark for each station in the Part 2 examination. Examiners make an overall judgement as to whether candidate performance in each stations rates as pass, borderline or fail. The overall judgement is used to determine the pass mark for future candidates. The GMC takes the mean scores of previous candidates judged borderline in each station to work out the pass mark.

Decision study
A decision study uses the results of a generalisability study to model the effects of changing aspects of the OSCE. For example, it will help the GMC to answer questions such as:

What would be the reliability of the OSCE if it had \( n \) stations instead of 14?

Would two examiners and \( n \) stations attain a particular reliability coefficient?

How is the SEM affected by decreasing the number of stations?

What would be the effect of increasing the length of the OSCE stations?
Generalisability study
An extension of classical reliability theory and methodology which indicates the magnitude of errors from various specified sources, such as number of items in the assessment, the number of assessors etc. The analysis is used both to indicate the reliability of the test and to evaluate the generalisability beyond the specific sample of items, persons and observational conditions that were studied. A decision study uses the results of a generalisability study to model the effects of changing aspects of the OSCE.

OSCE
Objective Structured Clinical Examination – a multi-station clinical examination (typically having 15 to 25 stations). Candidates spend a designated time (usually 5 to 10 minutes) at each station demonstrating a clinical skill or competency at each. Stations frequently feature real or (more often) simulated patients. Artefacts such as radiographs, lab reports and photographs are also commonly used.

Part 2 of the PLAB test is an OSCE which has 14 five-minute stations and a rest station. It may also include a pilot station, for example to test a new scenario, but pilot stations do not count towards candidates' results. All candidates take the same stations during each Part 2 examination and their performance is marked by an examiner in each station.

Part 1 Panel
The Part 1 Panel is responsible for overseeing and developing the Part 1 examination, and maintaining the question bank. The Part 1 Panel has 14 members, all of whom are medically-qualified associates. They are appointed for their knowledge and expertise in question writing, piloting and standard setting. The Panel reports to the PLA Board.

Part 2 Panel
The Part 2 Panel is responsible for overseeing and developing the Part 2 examination and maintaining the bank of OSCE scenarios. The Part 2 Panel has 13 members, who are medically-qualified associates. They are appointed for their knowledge and expertise in running OSCEs and in question writing, piloting and standard setting. The Panel also has a communications consultant. The Panel reports to the PLA Board.

PLA Board
The PLA Board has overall responsibility for overseeing and developing the PLAB test. The
Board has 11 members, all of whom are associates. Nine are medically-qualified and have a range of experience in assessment practices, as well as undergraduate and Foundation Programme education and training. They include an international medical graduate who has passed the PLAB test. Two are not medically-qualified.

Reliability

Reliability expresses a trust in accuracy or the provision of the correct results. In the case of assessments, it is an expression of internal consistency and reproducibility. This quality is usually calculated statistically and reported as coefficient alpha, which is a measure of a test's internal consistency. Generalisability theory is becoming the preferred alternative because, although it is considerably more complicated to calculate, it provides much richer information.

A perfectly reproducible test would have a coefficient of 1.0. That is, all candidates would achieve the same score on retesting. In reality, tests are affected by many variables and combinations of factors, so a figure of 1.0 is unrealistic. A reliability coefficient of greater than 0.8 is frequently quoted as an appropriate figure for high stakes assessments\(^\text{31}\) (although it is acknowledged that reliability measures are difficult or even impossible for small cohorts).

Single best answer

An examination format in which each question consists of a stem and a set of possible answers from which the candidate must choose the best answer. The format allows facts to be placed in a clinical context so that the application of knowledge and problem solving can be tested.

Situational judgement test

Situational judgement tests present realistic hypothetical scenarios and ask candidates to identify an appropriate response, generally in multiple choice format.

**Standard deviation**

Standard deviation is a widely used measure of variability, showing how much variation or dispersion there is from the average (mean) or expected value. A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values. The standard deviation is one component of the equation to calculate the standard error of measurement (SEM).

**Standard error of measurement**

The standard error of measurement (SEM) gives the confidence intervals for marks awarded to candidates. It is important in identifying borderline candidates.

**Validity**

In the case of assessment, validity refers to the degree to which a measurement instrument truly measures what it is supposed to measure. It is concerned with whether the right things are being assessed, in the right way, and with a positive influence on learning.