To note

PLAB Part 1 Examination – 15 March 2011

Issue

1. The results of the PLAB 1 exam of 15 March 2011 which generated a significant number of complaints.

Recommendation

2. To note the analysis of the PLAB part 1 examination of 15 March 2011 (paragraphs 8-42).

Further information

3. If you require further information about this paper, please contact us by email: gmc@gmc-uk.org or tel. 0161 923 6602.
Background

3. The PLAB 1 exam on 15 March 2011 was taken by 1885 candidates of which 412 (21.9%) passed. The results were delayed for 24 hours by a technical problem which prevented individual marks being released into GMC Online. Since the release of the results on 13 April 2011, the GMC has received 60 complaints. Many of these have been standard templates and all have been on one or more of four themes:

   a. The change in the proportion of Single Best Answer questions (SBA) and Extended Matching Questions (EMQ) in the paper.

   b. Our website guidance about the proportion of SBAs in the exam.

   c. The difference in the average score and pass mark.

   d. The delay in publishing the results.

4. The failed candidates included 722 who cannot book a PLAB 1 place again until they can show that they meet the increased scores in the IELTS exam following changes to the requirements, made in October 2010. Notification of this change and its potential impact was given to all candidates in advance of the exam.

5. Of these, 541 have alternative evidence of their English language proficiency which no longer meets our requirements, following a change on 1 February 2011. From that date we now require at least 75% of any clinical interaction, including personal contact with patients, their families and other healthcare professionals, which takes place as part of a course of study, or work, to be conducted in English. This is in addition to the existing requirement for evidence that the undergraduate course was taught and examined solely in English.

6. A further 181 have IELTS scores which no longer fulfil our criteria following a change on 1 October 2010 when our minimum IELTS requirements were raised to 7.0 in each of the areas tested (speaking, listening, reading and writing), and 7.0 overall.

Discussion

The change in the proportion of SBAs and EMQs

7. SBA is the preferred format for most postgraduate medical examinations, including the most widely taken, the Diploma for the Membership of the Royal Colleges of Physicians (MRCP). It is thought to be a better discriminator and less ambiguous than the EMQ. SBAs were introduced to the Part 1 exam in 2004 and at that time the proportion of SBAs (no more than 30%) in the exam reflected the size of the question bank rather than any preference for the balance of question types in the paper. The website guidance reflected this position.
8. Question writing workshops, held two or three times each year, have focused on increasing the number of SBA questions in the bank, and those in the last 18 months have produced only SBAs. This has been with the intention that the proportion of SBAs would increase and eventually replace EMQs entirely. This reflects best practice in the wider medical community. That said, this change in format has not been explicit and Part 1 Panel minutes, and PLA Board minutes make no reference to this.

9. The volume of SBAs in March differed from earlier papers, as shown in the table below. This trend will continue in the next exam, on 17 June 2011, when there will be 102 SBAs (51%).

<table>
<thead>
<tr>
<th>Exam date</th>
<th>Number of SBAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 2011</td>
<td>118 SBAs (59%)</td>
</tr>
<tr>
<td>Dec 2010</td>
<td>74 SBAs (37%)</td>
</tr>
<tr>
<td>Sept 2010</td>
<td>70 SBAs (35%)</td>
</tr>
<tr>
<td>May 2010</td>
<td>73 SBAs (36.5%)</td>
</tr>
</tbody>
</table>

10. Expert academic opinions have been obtained from Dr John Foulkes (psychometrician, principal of CH Assessment Ltd, the computer bureau that marks our exam papers), Dr Kathy Boursicot (reader in Medical Education and Head of Assessment, St George’s University of London) and Dr David Sales (immediate past Chair of the Part 1 panel, former assessment fellow and medical director of assessment, RCGP) on whether:

   a. SBAs are more challenging than EMQs.

   b. The language ability required to answer an SBA is greater than that required to answer an EMQ.

   c. SBAs take longer to read, assimilate, and answer.

11. These experts, and the PLA Board, are unanimous in their views that SBAs are not harder, they do not require greater language ability and they do not take longer to read, assimilate and answer. There is published research to support these individual views. Further details are in Annex A.

Website guidance about the proportion of SBAs in the exam

12. Following the exam, the GMC received two complaints about the change in the proportion of SBAs and the, then, inaccurate website guidance which stated “The proportion of SBA questions may vary from examination to examination but no more than 30% of the paper is composed of SBA questions.” Following receipt of these complaints, the website guidance was updated (from 29 March 2011) to “The proportion of SBA questions may vary from examination to examination.”
13. There was an oversight in maintaining the currency of the website guidance and this has, legitimately, fuelled some complaints. That said, evidence about the impact on the proportion of SBAs negates any of the points made by the candidates. This minor failure, alongside the recent changes to language requirements and the delay in releasing results drew criticism from some candidates that the GMC was intentionally making their PLAB exam more challenging than necessary and that we were restricting access to the test.

The difference between average score and pass mark

14. The pass mark is set using the Angoff method. Our website describes this as:

“…a panel of trained and experienced clinicians decide what percentage of minimally competent doctors at the appropriate stage of training would answer each question correctly. The Angoff method of standard setting is internationally recognised and it ensures that examinations are of a consistent standard over time.”

15. The Part 1 Panel conducts the Angoff standard setting exercise for new questions after every Part 1 exam, in terms of the likely probability that the candidate would know the correct answer to the question as it appears in the examination paper as opposed to whether the doctor could deal with the equivalent scenario in an authentic work situation. This approach takes into account whether the presentation or wording of a question affects the likelihood of the candidate answering correctly. See Annex A for details of research about Angoff standard setting.

16. The Part 1 Panel represent a range of specialties and are familiar with the process, in PLAB and elsewhere.

17. The focus is on discussion of questions. Having undertaken the Angoff exercise, the Panel reviews the paper in the light of the performance statistics and may allow extra answers or exclude items that this process identifies as flawed.

18. A psychometrician is present. A psychometrician is a mathematical, scientific, psychological and educational expert who is able to analyse examination results.

19. Prior to attending the meeting, the psychometrician will have processed the candidate marksheets and analysed performance against the correct answers. They will lead the exam review, drawing the panel’s attention to items which have performed unexpectedly or unusually based on the candidate performance statistics.

20. Unexpected or unusual performance would be where:-

a. There is negative discrimination on the correct answer but positive discrimination on another answer; this would indicate that more able candidates have identified the correct answer but it is incorrect in our answer key. Usually in this situation the psychometrician would simply change the answer key before re-running the results.
b. The majority of candidates appear to prefer an answer other than the one we are using as the correct answer but there is positive discrimination on both; if the panel decides the question is ambiguous, they would allow credit for both answers: the one we are using as the correct answer plus the one the candidates prefer. The question would need to be re-edited before reusing, to exclude the double answer. Alternatively the panel may decide the candidates are simply wrong but the question itself is not at fault, and do nothing (this is a common situation and often alerts the panel to a weak cohort before they see the pass rate figures).

c. The candidates’ answers are scattered - there is no majority going for the correct option; the question may be "too difficult" for this group of candidates (either written at too high a level, or the question is fine but the candidates are weak) and the candidates are guessing. If the panel agrees the question is essentially too difficult for PLAB they would usually allow a mark for the answer (so the candidates who got it right gain credit) but agree to edit the question before it is used again. If they decide the question is so flawed it is unfair to expect candidates to get the answer right, they would suppress the answer from the total score so it would not count towards the total pass mark. Or, as in the situation above, it is simply a weak group of candidates and the question itself is not at fault, in which case the panel would do nothing, and accept only the original answer.

21. The aim is that no candidate is disadvantaged due either to a flaw in our questions or a mistake in our answer key. The panel will give credit wherever possible. If a question is suppressed, the mean score for that item is deleted from the standard-setting spreadsheet, thus the overall score required to pass the paper goes down.

22. For this exam, four questions were suppressed which is consistent with previous exams.

23. The psychometrician will also outline their comparison of facility, by skill set and content. Facility values are simply a reflection, shown on a scale of 0.0 to 1 (where 1 is the full population) of how many candidates at a given sitting answered the questions correctly. This enables the panel to judge whether candidates are particularly strong in one skill set over another (for example, management of emergencies as compared with palliative care) and also whether there is evidence of speciality-specific ability. Whilst this is largely for the panel’s interest, it also demonstrates how important it is to maintain a consistent weighting of skill and topic areas from exam to exam.

24. The average facility is 0.546 (54.6%) for SBA and 0.549 (54.9%) for EMQ for the March 2011 questions. This means that candidates found each type of question equally difficult on this occasion.
25. There is no evidence to suggest the proportion of SBAs has any bearing on individual performance. For this exam, the reliability index (a statistical measure of consistency between 0 and 1) has remained high and in line with previous exams, with a value of 0.92.

26. There have been complaints that the average score is lower than usual and that the gap between the average score and pass mark must mean that the pass mark is incorrect. The two are unrelated. The Panel does not refer to the average score when setting the pass mark and it is an indication only of the performance of candidates. This is not a formative process. Candidates are at all stages of their career and there is no expected pass rate or mean score. We do not adjust the pass mark to determine the pass rate.

27. Another measure of candidate performance is the number running out of time, and this can be measured by the number of questions not answered. The following table shows what happened in this exam. Of 1885 candidates, 1737 (92%) attempted question 200.

<table>
<thead>
<tr>
<th>Question</th>
<th>Number of candidates who did not answer</th>
<th>Number of candidates answering</th>
<th>Percentage who did not answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>191</td>
<td>115</td>
<td>1770</td>
<td>6.1%</td>
</tr>
<tr>
<td>192</td>
<td>112</td>
<td>1773</td>
<td>5.9%</td>
</tr>
<tr>
<td>193</td>
<td>109</td>
<td>1776</td>
<td>5.8%</td>
</tr>
<tr>
<td>194</td>
<td>110</td>
<td>1775</td>
<td>5.8%</td>
</tr>
<tr>
<td>195</td>
<td>106</td>
<td>1779</td>
<td>5.6%</td>
</tr>
<tr>
<td>196</td>
<td>120</td>
<td>1765</td>
<td>6.4%</td>
</tr>
<tr>
<td>197</td>
<td>141</td>
<td>1744</td>
<td>7.5%</td>
</tr>
<tr>
<td>198</td>
<td>145</td>
<td>1740</td>
<td>7.7%</td>
</tr>
<tr>
<td>199</td>
<td>141</td>
<td>1744</td>
<td>7.5%</td>
</tr>
<tr>
<td>200</td>
<td>148</td>
<td>1737</td>
<td>7.9%</td>
</tr>
</tbody>
</table>

28. In addition, the discrimination index provided by the psychometrician showed that, apart from two flawed items with negative discrimination which were suppressed, all of these final ten items had good discrimination. Discrimination is simply a calculation of how well candidates performed on a single item, compared with how well they performed on the whole exam. It is a figure unique to each test, and for this test the index shows us the candidates were still thinking carefully about their answers even at this stage of the exam, rather than rushing to finish and guessing the last few.
29. Taking account of the extensive statistical analysis when responding to complaints about the average mark being so much lower than the pass mark, the response has been that this simply means that the candidates in this exam were less able than in exams with a higher pass rate (when the gap was not so great). Overall, there is no doubt that the pass mark has been set at the appropriate level for this examination.

30. One view is that the average mark has been influenced by the high number (722) of candidates who have, now, inadequate evidence of their English language proficiency.

The delay in the results

31. There was an unforeseeable and unprecedented technical problem in loading the results files and this delayed the release of results by 24 hours.

32. Results are received from the marking bureau, loaded into FAIM (standalone software) and then transferred to ‘staging tables’ in Siebel. Then the data is released into individual candidate records, which can be viewed through GMC Online.

33. The technical problem meant the file had to be recreated so that it could be processed through FAIM and onto Siebel. The problem related to the technical specification of the file and no data was lost or corrupted in the process.

34. Some candidates believed this delay was orchestrated to prevent them from complaining under the PLAB regulations which state that “If you want to complain about the examination you should send a detailed written report to our complaints section or to the GMC, Complaints and Correspondence, 3 Hardman Street, Manchester M3 3AW within 28 days of the date of the examination.”

35. The intention of the time limit is that candidates should complain immediately after an examination about issues which prevented them from performing to the best of their ability. It is not intended to be a route by which those who have failed can seek some form of reassessment. In the current circumstances, it is worth highlighting that any complaint received about the exam after the 28 day time limit in the regulations, is treated in exactly the same manner as those received within the time limit.

36. Whilst the complaint of deliberate delay is completely without foundation, it is another factor which has created dissatisfaction for some unsuccessful candidates.

Organised complaints

37. A final observation is that many of the complainants copied a template, with perhaps four or five versions being used by 60 (to date) complainants. It is likely that social and professional internet networks played a significant role in galvanising responses. This meant we received a high volume of similar complaints in a short period.
Conclusions

38. The balance of SBAs/EMQs made no difference to the time required to take the examination. SBAs are easier to read, assimilate and answer and this is supported by extensive expert opinion and research.

39. The website guidance should have been updated prior to exam. In future it will be checked and updated more regularly.

40. We employ robust exam methodology supported by expert question preparation, marking, assessment and administration. The average score and pass mark for this examination were determined in accordance with this.

41. IS failure was unforeseen and rectified quickly.

Recommendation: To note the analysis of the PLAB part 1 examination of 15 March 2011.

Resource implications

42. No resource implications arise from this paper.

Equality

43. No equality implications arise from this paper.

Communications

44. This matter has been reported to Council via the Chief Executive’s Report and this paper will be published on the GMC website.