External Examiner Report

GMC PLAB 2024

Introduction

The Professional and Linguistic Assessments Board (PLAB) is an assessment designed to ensure that doctors who graduated with a primary medical qualification outside of the UK have the right knowledge and skills to practise medicine in the UK. This year the PLAB assessment has been successfully integrated into the UK Medical Licensing Assessment (UKMLA) framework and will form the UKMLA requirement for international medical graduates (IMGs) wishing to practice medicine in the UK. The implementation of the UKMLA aims to unify assessment processes for UK medical graduates and IMGs, creating a standardised pathway for graduates to demonstrate the required clinical knowledge, skills and professional behaviours to practice medicine safely in the UK. Eligibility to sit PLAB has not changed with the implementation of the UKMLA and IMGs will need to pass 2 key elements; the Applied Knowledge Test (AKT) still currently known as PLAB part 1, and the Clinical and Professional Skills Assessment (CPSA), currently known as PLAB part 2.

This report will focus on the implementation of the UKMLA AKT at PLAB, referred to in the report as PLAB 1, as it is currently appearing in public facing documentation at PLAB. In 2024, there were approximately 21,000 candidates taking PLAB 1, with over 5,000 candidates per sitting. Assessment diets were available in February, May, August and November. Candidates could sit PLAB 1 in 9 locations across the UK, 2 European Union locations, and over 15 countries worldwide, with variations on location availability at each sitting.

Design of PLAB 1

Standard Setting

The pass mark for PLAB 1 is set using a Modified Angoff methodology, an evidence-based method which ensures differences in test difficulty between assessment diets is reflected in the pass mark for candidates. The pass mark is set to reflect the level of knowledge required for safe practice at foundation training year 2 (FY2) level, to reflect the standard to gain full GMC registration with a licence to practice.

There is a large pool of standard setters, a selection of whom form a standard setting panel for a given number of items (single best answer (SBA) questions). There has been a very successful recruitment drive to increase the pool of standard setters, which has resulted in a wider representation of medical specialities and doctors in different stages of training. Each standard setting panel comprises a good number of standard setters to provide robust undertaking of the Modified Angoff methodology. At the start of the standard setting meeting, panel members introduced themselves, including their specialty, level of training and experience in standard setting. There was a good mix of specialties represented including primary and secondary care and medical and surgical specialties. The panel was demographically diverse and represented clinicians at different levels in their own training. All members had been fully trained and had experience of recent working with FY2 doctors to ensure they had the knowledge to visualise the level of the assessment. The chair of the session ensured that all voices were heard and emphasised the need to respect and value all opinions.

The panel did not set pass marks for any individual assessment diet itself, but standard set new questions which had been written to be put into the question database. They also standard set questions which hadn't been reviewed in terms of standard for a few years.

The session started with a detailed introduction to standard setting and the process of Angoff/Modified Angoff and a clear anchor statement on the expectation of candidates was read. The standard represents the application of knowledge required to work safely on appointment to an FY2 role without additional support than that provided in such a role. Emphasis was placed on what would be expected in terms of core knowledge and safe practice, and it was emphasised that the assessment is not to select for excellence. Expectations were outlined, and emphasis on these being in line with UK practice and what would also be expected of UK medical graduates. There was discussion on the changes to processes and the assessment which have ensured PLAB 1 achieved compliance for the UKMLA AKT. For example, changes to the content covered (via the content map) and there was further discussion to include ED&I processes for example on avoiding stereotypes and ensuring UK practice was the focus. Throughout the standard setting process and discussion on questions, ED&I

permeated, and great discussion was had to ensure stereotyping was not present in questions or scoring. Where there were concerns, for example within question wording, changes were made to the question or, following process, where more detailed changes were required, the questions were referred back to question management group (QMG) for review.

A new set up has been implemented for the running of the standard setting panel to streamline the process for members. Each standard setter was allocated to a station with a large screen, laptop and mouse. Standard setters could set their standards for each question in private and by saving their results, these were automatically available to assessments team members to align the results for the second part of the session where members discussed their scoring and were able to make adjustments to their scores based on others' opinions. This streamlined the process and also ensured smooth running of the session without IT hitches such as loss of Wi-Fi. Whilst the changes aided ease of sharing question data and more robust security on standards and file sharing, the process of standard setting has not changed significantly with UKMLA compliance. Members independently enter their initial marks for each question and where a large range of scores was found across the panel, questions and standards were discussed. Discussions were relevant, showed detailed understanding of the questions, level and expectations of practice and showed understanding of the candidate population. Panel members were not afraid to ask for clarification and made clear comments to show recognition of the difference between standard setting for FY1 level vs FY2 level.

Panel members were well trained to understand psychometric data on questions, with the sharing of some indices although these were not read out for every question discussed and metrics were referred to in quite vague terms such as 'it had a high facility last time it was used' or 'it discriminated reasonably well'.

Content

PLAB 1, is a 3-hour exam, comprising 180 SBA (60 seconds per questions) where candidates select the best option from 5. Unlike UK Medical Schools, PLAB 1 comprises only 1 paper. This helps lessen the financial burden on candidates who, in addition to paying to sit the assessment, will often have expensive travel arrangements and accommodation to pay for. It balances the time and cost well with ensuring there are enough questions to ensure the high reliability (for PLAB 1 typically over 0.9) required of this high-stake assessment.

Sampling

PLAB 1 question writers use the MLA Content Map (General Medical Council, 2021) based on Outcomes for Graduates (General Medical Council , 2020), Generic Professional capabilities framework (General Medical Council, 2017) and relevant parts of the Foundation Programme curriculum (UK Foundation Programme, 2021) to write questions. This has replaced the old PLAB 1 blueprint, but does not represent a significant change for candidates, with much overlap between the two.

Questions for each paper are selected against a sampling grid to ensure the breadth of the content map is covered in each assessment. Questions are selected to cover areas of clinical practice, including both medical and surgical presentations, and areas of applied knowledge from basic medical science, through investigations, diagnosis and management and covers aspects such as health prevention and wellbeing and professional practice. Balance of questions between both areas of clinical practice and areas of applied knowledge within the sampling grid have been determined using an expert panel with knowledge of the AKT requirements and the candidate cohort. The balance shows understanding of the aims of the assessment, to ensure that IMGs have the required knowledge to safely practice medicine in the UK at Foundation Training 2 level with much focus on investigations, diagnosis and management in UK contexts. Professionalism and ethical and legal responsibilities have good representation and are relevant to the level of the assessment.

An automated system has been developed which selects questions (those which have been written and passed through quality assurance processes including compliance to the UKMLA) within the PLAB 1 database to be used in an assessment diet. The 240 selected questions are sent to the question selection group, who select 180 of the questions for use in the final assessment. The group selects questions to ensure fair representation of patient demographics and elements of care, and to ensure that there is balanced medical content and demographic representation in images.

The chair of the question selection group outlined the role of the group and the expected outcome and defined what they were able to do with questions within the meeting including question removal or minor edits. The group also acts to ensure fair coverage of the content map via the use of the sampling grid, but also ensures inclusion of different elements of care, patient demographics and that questions are free from bias and stereotypes. There were some very innovative questions designed around patient demographics which will clearly test candidates' ability to understand the intricacies of the patients they will treat, moving away from candidates being able to answer through being test-wise or looking for key words to trigger certain answers.

The atmosphere and chairing of the group ensured everyone was included and panel members were happy to ask questions, ask for clarification or justification of issues.

The question selection group included 10 individuals, with an excellent mix of ages, ethnicity and gender. There was representation from all 4 nations of the UK. This was particularly important given the different guidelines and processes in each of the 4 nations ensuring fairness to candidates. Many questions were discussed to ensure all questions were relevant and had the same correct answer irrespective of the nation a candidate may have experience/knowledge of. Clear processes exist and are followed where if significant changes to questions are needed e.g. to ensure relevance to all UK nations, questions are sent to the QMG who can edit and standard-set questions if necessary. Members of the question selection group also attend QMG, which helps with the discussions between the two groups and understanding the changes that could be made to questions. Lay membership of the question writing group has been advanced through the use of OSCE role players to advise on for example, the use of language in questions and the description of patients.

It was good to see involvement of all panel members in discussions, who represented a wide variety of clinical background including medicine, surgery, women's health, psychiatry and general practice. There was also a pharmacist present who gave invaluable advice on prescribing and treatment plans etc. There were no dominant voices, which ensured both experienced and newly appointed members of the panel were able to express their opinion, all bringing valuable ideas to the discussion which resulted in more robust decisions being made on the inclusion of questions and the balance of questions within the paper.

Some members of the panel had not seen the completed sampling grid, and how many questions fell into each part of the grid. It would be good to present this to panel members and how it was changed as the 240 questions were reduced to 180, some being concerned that it felt like some aspects were being removed more than others - whilst sampling was accurate and covered the sampling grid at the end of the session, reassurance could be given to the panel by seeing the grid as the session progresses.

Much emphasis was placed on ensuring questions assessed application of knowledge to relevant clinical scenarios rather than direct recall of facts. This has been a big positive change over the last year and aligns with best practice in assessments. For all questions, panel members asked themselves if this was application of knowledge and in instances where there was recall they discussed whether this was appropriate for the level and aims of the assessment.

Clear question writing guidance was used to check formatting of questions and use of abbreviations for example, ensuring consistency throughout the paper. Additionally, panel members checked previous questions did not give answers or cues to other questions. Language, ensuring questions are easy to read and understand, and updating of any terminology to more recent and inclusive terms was also undertaken.

Many questions written represented different genders and pronouns, skin colour, ethnicity and sexual orientation.

PLAB 1, standard set to Foundation Training 2 (FY2) level, ensured much of the discussion on questions was directed to focus on level of training and expectations within the FY2 role. Where there was misalignment, for example selected questions were not at the appropriate level, these were removed from the paper and send back to QMG for review. Where ideas for other questions were identified, these were noted and passed to the QMG too.

Quality Management

There are clear processes in place for question writing and development, reviewing and standard setting. New questions are developed and older questions reviewed through the QMG. Membership of the QMG group has increased significantly over the last year to include representation from more specialties including pharmacists. Another recruitment round is taking place for more panel members too. Panels were very well attended and represented a diverse group of practitioners and demographics. Members of the panel had been trained thoroughly and understood the complexities of assessment and best practice. Throughout the observed session, new question writers were supported and were allowed time to ask about the questions they had written.

In the past, questions have been written in situ within QMG meetings, panel members being given a list of topics on which to focus. This process has been updated, and a new process is being evaluated so that panel members write questions before they come to the meeting and within the meeting questions are reviewed and changed following discussion. This worked very well, with panel members being able to spend time thinking about questions outside the meeting and then review occurring in the session. I think this has helped to drive the development of questions to test application of knowledge to clinical scenarios. This was clearly evident within the OMG meeting. New questions were clearly meeting this requirement and when older questions were being emended/reviewed, where they had tested recall of knowledge they were changed to assess application of knowledge.

All members have completed comprehensive mandatory training on data protection and GDPR along with ED&I. There are secure methods of ensuring questions are not leaked and a new management system is currently being piloted to strengthen this further.

Support for Candidates

Familiarisation

Detailed information on PLAB 1 is presented to candidates clearly through the PLAB website. This has been updated significantly since 2023 to represent the changes to PLAB 1 to comply with the UKMLA. Where there have been significant changes this is clearly stated on the website and there is a new section to explain PLAB 1 compliance and UKMLA requirements.

The number of questions available for candidates to use for practice has increased significantly with a sample of 85 questions available through the PLAB 1 website. Questions represent the style of questions candidates will receive in PLAB 1 and including questions with images, ECGs and test/investigation results to help orient the candidate to what types of information will be given within an assessment.

Reasonable adjustments for candidates are constantly being reviewed and are implemented by the administration team. A consultant is available to advise on strategies to support candidates fairly, with the understanding that the level of competence shown by the candidates is the same (e.g. the level of knowledge has to be the same to be safe to practice) but the support to help them show this level can be put in place.

Safety

The assessment team go to great lengths to ensure that all candidates receive have a positive experience of the assessment regardless of the location of the assessment in the world. Safety of candidates is paramount to the team and the team go to great lengths to ensure that all location venues are safely accessible, whilst maintaining positive candidate experiences. Over the year, whilst a cancellation has been unavoidable due to safety concerns on the advice of the British Council, many issues have been successfully resolved to ensure fair experiences by candidates, such as moving start times to avoid city road closures and protests in one country, to employing extra invigilation in sites near to locations of the UK riots. With mitigations in place, all assessments ran without incident. Candidate feedback is positive, appreciating the judgements made by the assessment teams.

Results and Feedback

Candidates receive their results and feedback in a timely manner approximately 6 weeks after the assessment. This timeframe includes the compiling of over 5,000

candidates results from around the world, the detailed quality assurance processing of results and the ratification of results at Exam Board.

Feedback to candidates after an assessment, has been updated to reflect the assessment blueprint. Students receive their score, the average score for all candidates and the pass mark for the assessment diet. Further, candidates receive feedback on 10 main areas of the clinical practice based the MLA Content Map – being given the percentage of questions they answered correctly in each area. This has been reflected in changes to the PLAB 1 guide published on the PLAB website for candidates to view (General Medical Council, 2025). The site clearly outlines result dates for each diet of the assessment, how the total score is calculated, how the pass mark is set, and the feedback and metrics candidates will receive as part of their results. The feedback has been significantly improved with the introduction of the AKT requirements as candidates can now see how they performed in different areas of clinical practice rather than 3 skill domains as it was historically.

Policies and Resources

Security

In each diet, PLAB 1, is sat across the world, locations being in many different time zones. It is not possible for the assessment to start at the same time across the world as very early or late start times would be unfair to candidates sitting the assessment in some parts of the world. Fairness in ensuring candidates have the same experience of sitting the assessment is a focus of the team. Very early/late starts would be unfair to candidates who may be fatigued, resulting in poorer performance. To help align times, the assessment team, after discussions with many stakeholders, have moved the start time of some assessment locations. This was balanced with ensuring fair and equitable student experience of the assessment regardless of assessment location. Whilst it is impossible to align them perfectly, other strategies have been discussed to try and identify instances where a possible question leak could occur between locations which finish the assessment before the start of the assessment in another location. Psychometric evidence could be used to show questions are found to be similarly difficult at different locations i.e. those starting early/late where there is more opportunity for question leakage, although this needs to be contextualise with question content; some parts of the world will have the same practice as the UK and others may have different practice resulting in differential performance of the same item in candidate.

Psychometric analysis to determine the effect of exposure on question performance has been done and has yielded evidence to show that there is a very slight increase in

facility (i.e. questions are found to be very slightly easier) after multiple uses of the question. This is unlikely to significantly affect candidate performance within PLAB 1 in the future given that there has been a significant drive to increase the questions in the question database, resulting in questions being used less often, but more importantly, evidence suggests that questions needed to have been used in up to 20 assessment diets for any increase to be detected. Any increases were also negligible. To see a measurable increase in facility within the hour or so between assessments would take considerable effort to remember questions and to disseminate to candidates in later sittings. Therefore, the risk of this is very low. To assess the risk in more detail, the assessment team are now working on extracting raw facility data on individual questions from each assessment to determine the exact impact on facility each time a question is reused.

In addition to psychometric analysis, the assessment team have been trailing collusion tracking software within PLAB 1. In work so far, this has been used to look at scoring in assessments and where issues have been identified, these have been investigated appropriately.

Evaluation and Quality Assurance

Psychometric Analysis

Exam Boards for PLAB 1 take place after each diet and have a structured agenda covering the quality assurance and psychometric analysis processes after each assessment.

Trained teams each examine aspects of the assessment process including analysis at the whole assessment level and individual item level. Any metrics for a question which fall outside of the normal range outlined in process documents or showing anomalous results are discussed at Exam Board and decisions on action determined by the Exam Board. For example, for the November diet PLAB 1, 23 questions from 180 were discussed. The discussion was thorough, showed board members understood the metrics and their strengths and limitations and were able to apply educational research evidence to make quality decisions over the removal of any questions. Removal was appropriate where questions were removed. Discussion highlighted some very good aspects in new questions which had moved them from rote recall to application of knowledge, this meant candidates were not able to ‘question spot’ using key words as aids to the answer and instead needed to assimilate the given information to answer the question.

Where questions are removed from an assessment, the pass mark is re-calculated based on the reduced number of questions (a standard error of measurement (SEM) is

also appropriately added to the pass mark). This is confirmed with Exam Board members after the Exam Board meeting along with the number and percentage of candidates who passed the assessment. Throughout 2024, the pass marks and number of candidates passing each diet has been consistent.

A highly detailed report on differential item functioning (DIF) of PLAB 1 questions based on candidate demographics was written by the psychometrician. The aim was to look at the prevalence of DIF in PLAB 1 items, focusing on sex and ethnicity. The analysis determines whether subgroups of candidates at the same overall level of ability perform differently of specific items. The number of candidates sitting the assessment makes this an appropriate analysis and subgroups were defined appropriately to ensure numbers were appropriate for analysis and interpretation. There was some DIF for both ethnicity and sex on individual items (although as a consequence of the methodology employed, the effect on the assessment overall, that is candidate total score, can not reliably be made). Further steps are now being taken to look at individual questions to see if there are aspects to the question design which may reduce the DIF. I think it is important in understanding the results of the analysis to consider the context of PLAB 1 within this analysis. Candidates have very different backgrounds and experiences, some aligned to UK practice (which is tested in PLAB 1), some less so. It is important to remember that these differences may in part explain the DIF in questions rather than the presence of bias per se and should be considered in more qualitative aspects of reviewing questions.

Quality assurance processes are strong and continue to be developed further. Where incidents occur, processes are scrutinised and developed to stop the same happening again. For example, at the question level, a print error in one paper, has led to changes in the standard operating procedures (SOPs) to prevent the same error happening again. Whilst the print error led to differences in candidate experience in the UK compared to overseas, the responses were appropriate and will not have significant impact on candidate performance. Indeed, responses were in favour of the candidates, who were notified of the error, so they did not waste time answering a question printed incorrectly. New SOPs based on the current paper development processes will stop this happening again, but trials of new question management software will also potentially stop this occurring too if it is adopted after evaluation.

The Assessment Delivery Team and Assessment Development Team have been visiting sites delivering PLAB 1 to complete quality checks. This year 4 UK sites and 1 overseas site have been evaluated, with a mix of teams running the assessments including VICTVS and the British Council. All aspects of the assessment were scrutinised, from the run up to the assessment, through to the implementation of the assessment and post-assessment processing. Compliance on following of GMC guidance was assessed. Very few issues were raised and the few issues that did occur have been

actioned appropriately and solutions identified. Following the visits, a plan for quality assurance visits has been made and plans to visit more locations are being made. These will include 2 overseas and more UK visits each year. Locations are given little notice of the planned visit to ensure that the team view a true reflection of the delivery of the assessment.

Conclusion

There have been a number of changes to PLAB 1 over the past year to ensure that it is now compliant with the UKMLA requirements. The changes have brought about a more robust assessment of application of knowledge at the FY2 level and for assessing knowledge based on UK practice. Changes have been clearly outlined to candidates through the PLAB website and are fair, robust and evidenced. The challenges remain on ensuring a large number of candidates, across the world, have the best experience in taking the assessment and the team go to great lengths to support candidates and ensure their safety.

It is good to see some groundbreaking evidence being produced on DIF in medical education assessments with sound justifications and clear understanding of the data, sample size and assumptions of the analyses being undertaken. The results are being contextualised in a thoughtful and appropriate manner in moving the research forward into the next year.

Recommendations

Aspect	Recommendation
Standard Setting	More specific psychometric data could be shared with the standard setting panel on the past performance of questions. Panel members had a good understanding of psychometric indices and the strengths/weakness of measures. This would aid thinking in the future and given that many panel members sit on the question writing group, gives members a good idea of what the candidates find easy and more difficult.
QMG – Sampling Grid	It would be useful to show panel members the completed sampling grid numbers at question selection group meetings and show the change through the session as questions are removed. This would reassure the members that questions were not all removed from the same part of the sampling grid.
Quality Assurance – Paper Development	Check that new SOPs work to prevent print errors in final question papers.
Psychometric Analysis – Location Performance	To determine if there is any potential sharing of questions between locations where candidates have finished before others have started, psychometric analysis could compare questions performance at these locations. However, it is important to also consider that there could be some differences in performance based on the education system and experiences in general at different locations. For example, if in one location the standard management is the same as the UK, you may expect them to better than if the other location has a different management guideline – just based on familiarity and experience.
Psychometric Analysis – Location Performance	Candidates who sit PLAB 1 in the East could potentially finish the assessment and share questions with those sitting the diet in the very Western locations. Whilst this risk is very low, it could be investigated whether it is possible to alter the order of the questions in the locations where there is no attendance overlap, to try and reduce the strategic learning of questions 1-10, 11-20 etc which could, in an organised manner help pass on questions to later sit candidates.
Psychometric Analysis - DIF	Some items in PLAB 1 revealed evidence of DIF by sex and/or ethnicity. Question writers could review these questions to look for possible ways to improve the question(s). It is important to remember the context of PLAB 1 too, and consider aspects of fairness, validity and effects of education too.

Dr Natalie Cope

Senior Lecturer in Clinical Education (Psychometrics): Keele University School of Medicine

Prof. Adrian Freeman

Emeritus Professor of Medical Education: University of Exeter School of Medicine

05/02/2025

References

General Medical Council . (2020). *Outcomes for Graduates.* General Medical Council .

General Medical Council. (2017). *Generic Professional Capabilities Framework .* General Medical Council.

General Medical Council. (2021). *UKMLA Content Map.* General Medical Council.

General Medical Council. (2025, 01 28). *PLAB 1 Guide*. Retrieved from General Medical Council: https://www.gmc-uk.org?registration-and-licensing/join-our-registers/plab/plab-1-guide

UK Foundation Programme. (2021). *UK Foundation Programme Curriculum 2021*. UK Foundation Programme.