PLAB External Examiner Report 2025

Introduction

This External Examiner Report reflects observations of the PLAB assessment processes undertaken during a series of visits in 2025, with a particular focus on PLAB 2 (Clinical and Professional Skills Assessment – CPSA). The report has been compiled by External Examiners Dr Natalie Cope and Prof Kathy Cullen.

Across all observations, the General Medical Council demonstrated a strong commitment to fairness, validity, reliability, and transparency in high-stakes assessment. The assessment systems observed align well with the Medical Licensing Assessment (MLA) requirements and reflect ongoing investment in continuous improvement, informed by psychometric evidence, stakeholder feedback, and reflective practice.

Visits

We observed multiple stages of the assessment lifecycle across the 2025 assessment year. These included AKT board activity, CPSA Board meetings, station and standard-setting management group (SMG) meetings (both in-person and online), and live delivery of a Quad OSCE. Collectively, these observations allowed the examiners to review assessment design and development, examiner and simulated patient preparation, delivery processes, standard setting, data handling, and post-assessment quality assurance.

Of note, during the Quad OSCE on 4 November 2025, we were able to view all four circuits on the same day, and a high level of consistency between circuits was observed.

Design

- Assessment strategy

The overarching approach to both PLAB 1 and PLAB 2 is clearly explained and logically organised on the website, making it straightforward for candidates and examiners to access and understand. This includes samples of charts used, reasonable adjustments and how to obtain results. It also outlines compliance with the MLA requirements.

- CPSA design

To enhance the breadth and depth of the item bank, monthly item-writing lists are now being produced. These provide targeted guidance on areas requiring development while also identifying domains where there is already a significant volume of material. This approach helps ensure that new items are distributed appropriately across the blueprint, promoting balanced assessment coverage.

There has been a substantial focus on strengthening the development of simulator/METI stations. This work has increasingly emphasised collaborative item writing, with teams working together to improve the quality, consistency, and educational validity of station

content. Alongside this, further piloting of METI stations is being undertaken to ensure that they perform reliably and support robust assessment outcomes.

Existing stations have been systematically reviewed and updated to align with the MLA Content Map. This work ensures that assessments remain current, relevant, and fully mapped to curriculum requirements, supporting both transparency and quality assurance.

In addition, the team are actively developing new approaches for assessing prescribing within OSCEs, with a focus on both the act of prescription writing and the decision-making underpinning safe prescribing practice. This work is particularly important given that students will not undertake the Prescribing Safety Assessment. As part of this development, the team is also piloting the use of actors within relevant stations—an enhancement from previous iterations, which historically did not include an in-station participant. This change aims to create more authentic clinical encounters and strengthen the assessment of communication and prescribing competencies.

- Scoring

The marking guidance was discussed in detail, with emphasis on the intentionally broad structure used to support domain-based assessment. Checklist-style marking has been removed from station development, and anchor statements have been introduced to promote consistency and alignment among markers. These anchor statements are positioned at every alternate mark point and are designed to guide assessors in applying domain descriptors reliably.

Considerable attention was given to the clarity and precision of the language used within both the marking domains and the anchor statements. The group worked to ensure that all descriptors were concise, unambiguous, and reflective of the competencies being assessed. The anchor statements themselves were acknowledged as clear and effective, demonstrating a logical progression in the expected level of knowledge and skill across increasing grade descriptors.

During question development, there was thoughtful and reflective discussion about the borderline candidate profile. This included careful consideration of what performance at the borderline level should look like and how it would differ from that of both stronger and weaker candidates. This reflective approach supported the development of discriminating, fair, and appropriately calibrated station content.

The team linked these discussions to anticipated post-assessment psychometric analysis. They reflected on how well the stations were expected to discriminate between candidates, both within individual domains and across the station as a whole. This integration of design considerations with predicted psychometric outcomes ensured that the assessment framework would be robust, valid, and aligned to best practice.

- Standard setting

The OSCEs are standard set using borderline regression completed using robust process.

- Assessing professionalism

Cause for concern forms are available to ensure that any professionalism-related issues can be appropriately documented. Their use provides a clear mechanism for capturing concerns that fall outside the formal marking domains, ensuring that important behavioural or conduct matters are not overlooked during assessment.

There was constructive discussion regarding the importance of recording such concerns separately and providing meaningful feedback to candidates when appropriate. As these professionalism issues are not included within the formal marking guidance, they would not otherwise appear in routine assessment feedback. The group recognised the value of having a parallel system that supports transparency, supports learner development, and maintains the integrity of the assessment process.

Content

- Content sampling

Station development was discussed in detail, with all stations clearly mapped to the relevant assessment domain, the associated area of clinical practice, and the specific presentation or condition being assessed. This structured approach ensures that each station is grounded in defined competencies and reflects authentic clinical scenarios.

It was also noted that including the explicit aim of each station is particularly valuable. Providing the examiner with a clear statement of intent supports focus, consistency, and alignment during marking, ensuring that assessors understand the core purpose of the station and the aspects of performance they are expected to prioritise.

- Quality of CPSA content

The question development group demonstrated strong diversity in its membership, bringing together individuals with a wide range of perspectives. Notably, the inclusion of a simulated patient provided valuable insight into role portrayal and delivery, enriching the authenticity and clarity of the scenarios being developed. The group worked within an excellent collaborative atmosphere, both online and in person, where participants felt comfortable contributing ideas, asking questions, and engaging fully in discussion.

The potential use of eBNF was highlighted as a means of enhancing authenticity within the stations, recognising that most doctors now rely on electronic prescribing systems rather than paper-based references. The BNF is no longer available in printed form in the NHS. Incorporating this approach would therefore support the realism and relevance of the assessment environment

The template structure for METI stations is to be commended. It provides a clear and comprehensive framework, including an excellent section specifically for technicians responsible for running the station on the day. This section, which goes beyond standard station information, includes details such as the availability of standard crash carts, with the

option to add station-specific equipment when required. This structured approach supports consistency and operational efficiency.

All questions were developed in accordance with a detailed, in-house style guide that ensures clarity, standardisation, and alignment across the full suite of station materials. This consistency contributes to a coherent and high-quality assessment experience.

The range of topics covered was wide and relevant to what an F2 doctor could expect to see at work.

- Security of CPSA content

All question and station development materials were managed securely, with access restricted through GMC Connect. This ensured that only authorised individuals could view or interact with the documents, maintaining a controlled and secure working environment. Members were regularly reminded—both by email and at the beginning and end of each session—of the requirement not to remove any material from the room or export content from shared computer folders.

During the development sessions, edits were made centrally to a single master version of each station. Contributors were able to view these changes live on screen but did not have direct access to download or modify the underlying files themselves. As a result, the final version of each station was considerably more developed than the limited version initially shared with members, adding an additional layer of information security. Comments submitted by members who were unable to attend the meeting were captured within GMC Connect, ensuring transparency and continuity while maintaining secure access controls.

Online meetings were also tightly monitored to preserve confidentiality. When an unexpected noise occurred during one session, the team carried out an immediate check to confirm that no unauthorised individuals were connected and that the meeting was not linked to any external calls. This vigilance reflects the high priority placed on maintaining the integrity of assessment materials.

Safeguards are also in place to detect and address potential station leakage. For example, if a station consistently shows unusually high pass rates over repeated uses—suggesting that candidates may have prior knowledge—the station is suppressed and removed from the question bank to protect the fairness of the assessment.

Strict operational procedures further protect station confidentiality during delivery. Candidates and examiners are kept separate, and morning and afternoon candidate groups are also segregated, preventing the sharing of information between sessions and maintaining the validity of the assessment process.

Preparation of and support for candidates

Candidates are provided with appropriate familiarisation with the assessment process to support their preparation and reduce anxiety. Although there are no visible timers within the

assessment rooms, candidates are given a clear two-minute warning, ensuring they can manage their time effectively during each station.

Processes for delivering results and feedback are well-structured. Examiners are required to leave feedback specifically for candidates who are judged to be borderline at a given station. This feedback is then visible to candidates when they receive their results. The use of the risr/assess platform supports this requirement effectively, as the system ensures that examiners must provide the necessary comments before completing the station record.

The student briefing was delivered in a supportive and positive manner, offering practical advice on staying calm, maintaining focus, and approaching each scenario as they would in real clinical practice. This contributed to a reassuring and professional environment for candidates at the outset of the assessment.

A recent change in the circuit layout means that no candidate now begins the assessment on a rest station. This adjustment has been beneficial, reducing pre-assessment anxiety and ensuring that all candidates rotate through stations in a more balanced and equitable manner.

Preparedness of examiners and patients

- Examiners

Examiners are provided with comprehensive reference material for each station, including relevant clinical guidelines and links to clinical knowledge summaries. This ensures that assessors have immediate access to authoritative resources and can review the underpinning evidence base for the station content. Examiners routinely review all station documentation in advance, alongside the candidate running order, for which they also receive a printed copy to support accurate checking and smooth operational flow.

Each station pack includes information on possible differential diagnoses, which is particularly valuable for non-specialist examiners. This supports consistency in expectations across the examiner group and helps ensure clarity regarding the key clinical considerations required within the scenario. Examiners are also able to request assistance, when necessary, with support typically provided during scheduled breaks. The introduction of enhanced risr/assess IT functionality has further streamlined this process, ensuring timely access to help when required.

The examiner briefing was clear, structured, and concise. It included sufficient time for calibration, with detailed discussion of marking domains, the scoring approach, and the importance of fairness and consistency across all candidates. Calibration was well organised, allowing for multiple run-throughs with the actors and thorough discussion of borderline performance and domain-specific expectations. Station setup was reviewed carefully, with facilitators on hand to provide support, and all checklists were examined in detail to ensure accuracy and alignment prior to use.

Any incidents or problems arising on the day can be escalated to the chief invigilator, who is responsible for investigating concerns and determining appropriate actions. This includes managing unexpected issues such as equipment failure or decisions regarding the need for candidates to resit a station. This structured escalation process ensures that problems are handled promptly, transparently, and in a manner that protects the fairness and integrity of the assessment.

- Simulated/real patients

Simulated patients (SPs) were actively involved in the development of questions and stations, contributing meaningfully to the refinement of scenarios and helping to remove any unintended stereotypes within roles, such as assumptions regarding a character's occupation or background. Their input ensured that role descriptions were realistic, balanced, and free from unnecessary bias.

They also played an important part in designing stations and drafting the behavioural guidance for actors. Their familiarity with the structure and layout of station materials meant that the written instructions were clear, concise, and accessible for all role players. Care was taken to avoid oversimplification, recognising that overly brief descriptions may lead to inconsistent acting or assumptions about how an actor should interpret a single statement. Although actor training was not observed, the group acknowledged that detailed guidance is vital to avoid variability in delivery.

While the aim is to keep station instructions succinct, attention was given to ensuring that brevity did not come at the expense of essential detail. Comprehensive role information is important to support consistent performances across lanes and maintain the integrity of the assessment. The development team also identified stations that may place higher demands on an actor—such as those requiring prolonged demonstration of pain—and considered where it may be appropriate to deploy more than one actor for sustainability and performance quality.

There was some discussion regarding the limited use of real patient examination within scenarios, which may present an opportunity for future development. Role players continue to contribute to examiner refresher training, supporting calibration and ensuring examiners better understand candidate-actor interactions.

The use of actors has now returned to pre-COVID levels, following a period of necessary reduction during pandemic restrictions. Ongoing reflection on the demographics of the actor pool is conducted to ensure representation aligns appropriately with station requirements. All actors hold Equity cards and consistently deliver high-quality performances. They receive relevant training, and support is available on the day to help maintain performance standards and respond to any issues that arise.

- Collaboration between examiners and patients

Examiners and simulated patients worked well together. No concerns were noted.

- Feedback to examiners and simulated patients

Actor performance is closely monitored as part of a robust quality-assurance process. Facilitators review performances via video, and comprehensive reports are produced—for example, 782 reports were published in 2025—providing systematic feedback on actor consistency and adherence to the scripted behaviour for each station. Feedback is actively acted upon. Instances such as consistent deviation from the script across a full day (where all candidates were exposed to the same variation), failure to provide laminates or required information, inconsistent body language, or variation in the pace at which information was delivered have all been identified through this process. Where concerns persist despite feedback, some actors have been removed from the pool to maintain the integrity and reliability of the assessment.

Examiners similarly receive detailed feedback on their marking performance at the end of each year. This includes comparative data that allows them to benchmark their marking against that of other examiners. When the metrics indicate areas of concern—such as consistent divergence from expected marking patterns—examiners may be invited to undertake retraining to support ongoing development and ensure continued fairness and accuracy in assessment

Policies and resources

- Policies and procedures

Clear anti-bribery and misconduct procedure policies available on the website.

- Resources and space

The assessment took place in an excellent environment that was consistently quiet, calm, and spacious. There was strong concordance between the different circuits, with each station located in its own dedicated room, supporting privacy and reducing noise transfer. Staff were clear in their roles and provided effective signposting for candidates to toilets, refreshments, and rest areas as required. Overall, the atmosphere was supportive and well-organised.

The risr/assess platform functioned extremely well, providing examiners with fully accessible digital packs that included detailed station information, copies of station laminates, marking guidance, and anchor statements. The marking interface was clear and intuitive, enabling smooth and consistent scoring during the assessment.

Invigilators had access to live video feeds of the stations and could view all associated documentation, allowing them to oversee proceedings in real time. They were also able to monitor total scores and standard-setting scores as they were entered into the system. This live visibility supported immediate identification of any missing marks or anomalies, enabling prompt correction.

In addition, risr/assess facilitated efficient communication between examiners and the facilitation team. Examiners were able to request assistance directly through the system, and

facilitators or invigilators could acknowledge receipt of these requests, ensuring timely support and maintaining operational flow throughout the assessment.

Data management

- Data acquisition

Examiners found the digital marking system easy to use and the real-time monitoring of data produced from stations was useful.

- Production of results

Results were generated using data exported from the risr/assess system

Evaluation and quality assurance

- Psychometric analysis

The psychometric profile of the assessment remains strong, with an average reliability coefficient of 0.71, which is considered acceptable for a high-stakes examination. Current data show no statistically significant differences when compared with previous years, despite the transition from an 18-station format to the current 16-station model, indicating stable assessment performance across cycles.

Actor-related variability is monitored closely, and stations are suppressed when significant inconsistency in actor performance is identified, although such occurrences are very rare. CPSA Board meetings demonstrate rigorous review of OSCE data, with post-exam suppression rates remaining extremely low—around 1%—reflecting high overall station quality.

Facility data indicate an appropriate spread of station difficulty, supporting a balanced and discriminative assessment. Standard setting is monitored through detailed analysis of the regression line slope for each station, with graphs demonstrating good discrimination: on average, candidates increase by approximately two marks with each successive grade. Where extreme score differences are identified, the team undertakes a thorough review of station performance beyond the regression analysis to determine whether any issues require intervention. These graphs are presented to the board with clear explanation and well-reasoned justification for retaining each station.

Station cut scores appropriately reflect differing levels of difficulty across the circuit. Psychometric analysis has also shown that variation in scoring patterns across lanes has reduced in recent cycles, suggesting improving consistency between examiners and station delivery teams.

Stations where regression lines for borderline regression (BLR) differ significantly between examiners in different lanes are suppressed, although this is uncommon. Any such stations are returned to the review group for revision of content and marking guidance to ensure future alignment.

On the day of the OSCE, invigilators conduct active quality assurance across all stations, observing both the scenarios and the examiners to ensure consistency, alignment, and fairness across the different lanes.

- External examiners

Having two external examiners from different professional backgrounds has proven highly effective, as it allows us to view the programme through two distinct and complementary lenses. We have been warmly welcomed throughout the process, and all procedures and practices have been shared with us in a transparent and open manner, supporting a clear and thorough understanding of the system.

Recommendations

The assessment system observed was of a high standard, and most findings related to examples of good practice. The following recommendations are intended to support further enhancement rather than to address areas of concern:

1. **Further development of authentic prescribing assessment**
   The ongoing work to strengthen prescribing assessment within OSCEs is commendable. Continued exploration of integrating electronic prescribing tools (such as eBNF) within stations would further enhance authenticity and alignment with real-world clinical practice.
2. **Increasing authenticity**
   Simulated patients are used effectively, particularly in communication- and consultation-based stations. However, there would be benefit in more seamless integration of physical examination using models with the human interaction component of these stations. For example, in some instances the use of a plastic model felt disconnected from the consultation, particularly when the model was located in a different part of the room from the simulated patient. In addition, the inclusion of real patients within the CPSA would be a valuable enhancement. Involvement at both the stage of station development and within selected stations on the day of assessment would help to broaden the range of clinical skills assessed and further enhance the authenticity of the examination.
3. **Continued monitoring of actor workload in demanding stations**
   Stations requiring sustained physical or emotional performance should continue to be reviewed to ensure actor welfare and consistency, including consideration of multiple actors where necessary.
4. **Ongoing review of examiner feedback utilisation**
   The rich feedback provided to examiners through end-of-year reports is a clear strength. Continued encouragement and monitoring of how this feedback informs individual examiner development would further support consistency and calibration.

Conclusion

Overall, the PLAB assessment framework observed during the 2025 cycle was robust, well governed, and continuously improving. Assessment design is clearly aligned with contemporary educational and psychometric best practice, supported by strong quality-assurance processes and a culture of reflection and openness.

As external examiners, we were consistently impressed by the professionalism of GMC staff, examiners, simulated patients, and facilitators, as well as by the transparency with which all aspects of assessment development, delivery, and evaluation were shared. The integration of detailed psychometric analysis into decision-making, alongside attention to candidate experience and examiner support, provides strong assurance of fairness and validity.

In summary, the PLAB assessments observed meet their stated aims and provide a defensible, high-quality evaluation of candidates' readiness for practice. The minor recommendations outlined above are intended to support the GMC's ongoing commitment to excellence in assessment.

Dr Natalie Cope Senior Lecturer in Clinical Education – Psychometrics, Keele University

Keele UNIVERSITY

Prof Kathy Cullen Consultant Respiratory Physician and Director, Centre for Medical Education, Queen's University Belfast

QUEEN'S UNIVERSITY BELFAST