Royal College of Anaesthetists

FRCA Examination Review

Report of the Examination Review Group

January 2023

Contents

1. Terms of reference

The stated aims of the Examination Review Group (ERG) 2020 were -

- To oversee and conduct an in-depth review of the FRCA examinations' purpose, validity, reliability, format, delivery method, fairness and concordance with best practice in assessment theory, in order to ascertain fitness for purpose as the recognised test of knowledge, skills and attitudes within the training programme leading to the UK CCT in Anaesthesia and progression to high level training.
- To submit a report on the findings and make recommendations for change to the Examinations Committee, Education, Training and Examinations Board (ET&E), Council and Board of Trustees.

The ERG is ultimately accountable to the Education, Training and Examinations Board of the Royal College of Anaesthetists (RCoA), reporting to it through the established processes for the Examinations Committee. It was serviced by the RCoA Examinations Department, with the Head of Examinations and Director of Education, Training and Examinations in attendance.

The purpose of the ERG is summarised as follows:

1. To review all matters relating to the examination for the Fellowship of the RCoA listed below.
    a. Purpose and role of the FRCA examinations within the assessment strategy for the training programme leading to CCT Anaesthesia.
    b. Currently accepted best practice in assessment.
    c. Validity of the current components of the exams and the overall exam, including exploration of alternative options.
    d. Fairness of the exam, including but not limited to pass rates and differential attainment.
    e. Statistical reporting and data capture.
    f. Lay involvement.
    g. Technology and computer-based testing.
    h. Candidate communication and feedback.
2. To ensure any changes meet the GMC standards and other current standards for best practice.
3. Production of a comprehensive report, and to make any recommendations for change to the ET&E Board/Council/Board of Trustees. Publication of the rationale for recommendations for change to be made via the *Bulletin* and College website.
4. Ensure actions are followed up and changes made as required.

The first meeting of the ERG was held on 15 January 2020, with one further meeting held on 24 February 2020. Due to COVID-19, this work was paused and resumed on 4 March 2021, with a further six meetings held. This report was drafted and submitted to the Examinations Committee on 7 April 2022. The final report was sent to Council on 14 September 2022.

2. Membership of the Examinations Review Group

The composition of the ERG was agreed as follows:

- Dr Mark Forrest – Chair
- Dr Fiona Donald – Vice President
- Dr Roger Sharpe – Chair, Primary FRCA
- Dr Jo Budd – Vice Chair, Primary FRCA
- Dr Patrick Hopton – Chair, Primary MCQ
- Dr Emily Simpson – Chair, Primary OSCE
- Dr Jamie Macdonald – Lead, Primary SOE
- Dr Simon Vaughan – Lead, Primary SOE
- Dr Carl Stevenson – Lead, Primary SOE
- Dr Jason Walker – Lead Statistics, Primary FRCA
- Dr Kevin O'Hare – Chair, Final FRCA
- Dr Gary Lear – Vice Chair, Final FRCA
- Dr Arun Krishnamurthy – Lead, Final CRQ
- Dr Sameh Abdul-Latif – Lead, Final SOE
- Dr Satya Francis – Lead, Final SOE
- Dr Elaine Wilson-Smith – Lead, Final SOE
- Dr Alister Seaton – Anaesthetist in Training representative
- Adrian Mason – Lay Committee representative
- Dr Richard Fuller – External educationalist, Health Professional Assessment Consultancy
- Russell Ampofo – Director of Education, Training and Examinations
- Fiona Daniels – Head of Examinations
- David Rowand – Examinations Manager
- Val Perkins, PA to the Director of Education, Training and Examinations, Secretary to the Group

3. Executive Summary and Recommendations

This report summarises a review of the FRCA (UK) Examination, which was undertaken during the academic year 2020–2021. It is proposed that these reviews continue to be repeated every three to five years to ensure that the FRCA examination reflects best practice in postgraduate medical assessment.

Primary FRCA Review

a. Primary MCQ

The Primary Multiple-Choice Question (MCQ) examination was moved online in August 2020. Despite an issue in August 2020, the delivery method seems to be working well, and the College has decided that this component of the examination should remain online as the most natural mode of delivery for the written examination in the long term. The Review Group were of the opinion that the MCQ forms a critical part of the examination strategy for the testing of knowledge and applied knowledge and that the plan to increase Single Best Answer (SBA) questions and fully phase out the use of Multiple True/ False (MTF) should continue. The Review Group recommended that the Primary MCQ be updated to extend the range of SBA questions by, for example, the utilisation of multimedia content.

b. Primary OSCE

The Primary Objective Structured Clinical Examination (OSCE) has been online since November 2020 and returned to a face-to-face format in 2022. It was agreed that a clinical component of this exam should remain, but that there should be further development of the OSCE in the future to enhance the assessment of clinical skills. The Review Group recognised that the OSCE examination should continue to reform and evolve in order to remove pure knowledge aspects from the examination and be representative of an authentic OSCE examination.

c. Primary SOE

The recommendation of the Review Group was for this part of the examination – the Primary Structured Oral Examination (SOE) – to be re-evaluated with those elements of knowledge that could be tested in a written paper converted to a format using SBA questions and other elements, such as the testing of clinical performance and the basic sciences that underpin that clinical performance, incorporated into a reformed OSCE. The Review Group agreed that the current knowledge and understanding aspects of the SOE examination would be more appropriately assessed in the written and OSCE components of the examination.

Final FRCA Review

a. Final MCQ

This exam moved online in September 2020. In general, the online delivery was felt to work well and the Review Group agreed that this component of the examination should remain online for the longer term. The plan to increase SBA questions and fully phase out the use of MTF should continue in alignment with the Primary MCQ development. The Review Group also recommended that standard setting and item development processes in the Final MCQ should be fully aligned to the Primary MCQ.

b. Final CRQ

This exam moved online in September 2020 and, despite a specific issue with the paper build in September 2021, the Group agreed that, like the MCQ examinations, this component should continue to be delivered online via TestReach for the long term.

c. Final SOE

It was agreed that this part of the exam should remain, but a strong recommendation was made to review and potentially change the standard setting method, which is currently borderline regression (BLR). The new Examinations Development and Assurance Group (EDAG) will be tasked with investigating an alternative standard setting method for this examination.

4. Update on 2015 examination review

A full copy of the Action Log from the 2015 examination review is in Appendix A. The majority of actions have been completed or are actions that evolve over time and are therefore ongoing.

5. Preparatory work for the review

In preparation for the review, lead examiners undertook extensive and objective reviews of each component of the FRCA examination in an attempt to identify keys areas of practice and deliverability that should be considered for change. These reports included Primary OSCE (Appendix B) and Primary SOE (Appendix C). This work reflects the ongoing and evolving work that the examinations staff and examiners have been doing to consider updates to the examination in line with best practice. Many of the changes and proposals for innovating the examination contained in these reports are wide-ranging and have been produced following careful and deliberate consultation with the examiner body over time. These documents were used as an initial reference point and evidence base ahead of the review discussions. The Review Group would like to thank the lead examiners for their time, expertise and efforts in compiling these reports to inform the review.

6. Exam purpose and structure

The Review Group sought to begin the examinations review with the purpose and structure of the FRCA examination and to consider whether this was: fit-for-purpose in relation to the requirements of consultant practice, in line with the new 2021 curriculum, and covered a sufficient and appropriate range of anaesthetic knowledge and practice in the syllabus.

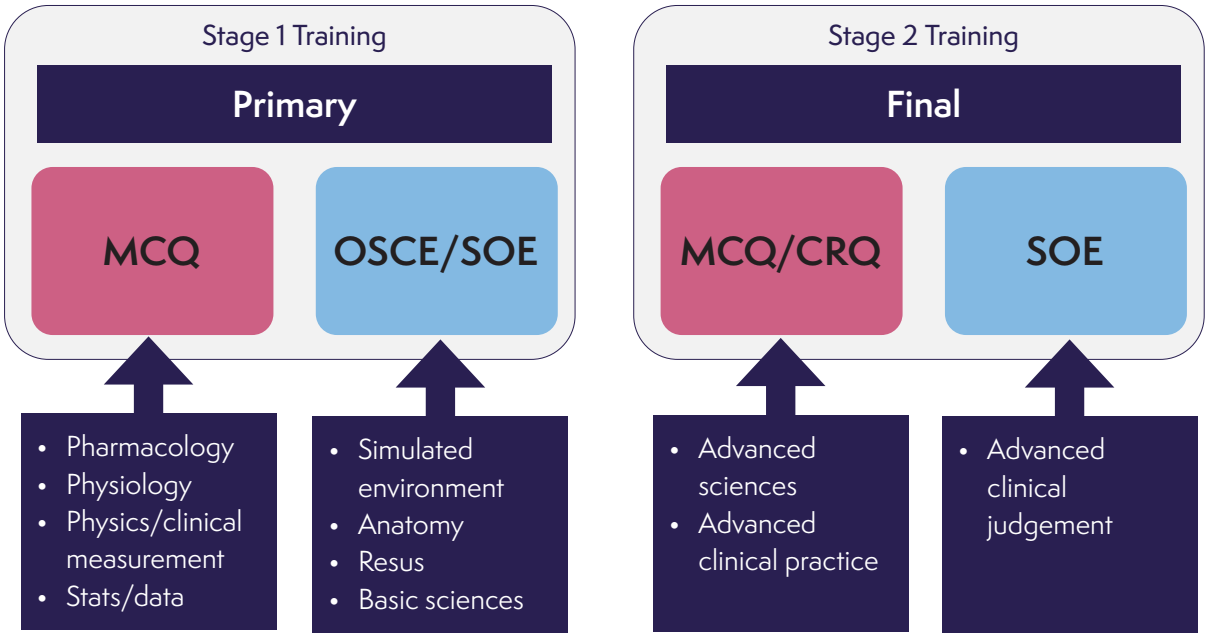
The Group quickly agreed that recommendations from the review should be based primarily on best practice in examinations and assessment theory that would deliver a valid and reliable examination. This concept was felt to be the foremost guiding principle. The Group were also guided by the *GMC standards for curricula and assessment Excellence by Design*,[1] the Academy of Medical Royal Colleges Guidance for Standard setting: A Framework for high stakes postgraduate competency-based Examinations,[2] as well as individual reviews of examination components by lead examiners (Appendices B–C).

Current format and structure of the examination

The FRCA Examinations are an integral part of the RCoA's programme of assessments. They provide Anaesthetists in Training with the opportunity to demonstrate at critical progression points the required outcomes of their training programme.

The examinations comprise a programme of summative assessments in two parts – Primary and Final. Each part uses validated assessment methods to test a broad spectrum of knowledge, understanding, skills, behaviours and attitudes, as defined by the anaesthetic training curriculum.

Current FRCA structure



The Primary FRCA examination consists of three components: a Multiple-Choice Question examination (MCQ), an Objective Structured Clinical Examination (OSCE), and a Structured Oral Examination (SOE). All Primary examination components are blueprinted to Stage 1 of the anaesthetic curriculum. Questions used in each component are tagged to the Primary examination's syllabus. The High-Level Outcomes (HLOs) and the capabilities relevant to each domain of learning are set out in the learning syllabus. Successful completion of all three Primary examination components is required to complete Stage 1 and proceed to Stage 2 of the anaesthetic training programme.

The Final FRCA examination consists of two components: a written examination in two parts (Part 1 is an MCQ examination and Part 2, a Constructed Response Question (CRQ) examination), and an SOE. The examinations used in the Final FRCA are blueprinted to Stage 1 and Stage 2 of the anaesthetic training curriculum. The questions used in the Final FRCA examinations are tagged to the FRCA Primary and Final examinations syllabuses. The HLOs and the capabilities relevant to each domain of learning are set out in the assessment blueprint. Successful completion of the Final FRCA examination allows progression to Stage 3 of anaesthetic training.

The Primary FRCA examinations

The Primary FRCA Multiple Choice Question (MCQ) examination

The Primary OSCE is a summative assessment of a candidate's clinical and communication skills, and applied technical knowledge of anaesthetic equipment, clinical monitoring and measurement. It comprises 16 assessed stations with no `killer' stations. Candidates have one minute to read the instructions for each station and five minutes per station for the assessment. To allow additional candidates to be examined, one or more `rest' stations may be added; candidates are not examined in these stations. The Primary OSCE stations currently assess resuscitation, technical skills, anatomy (general procedure), history taking, physical examination, communication skills, anaesthetic equipment, monitoring equipment, measuring equipment, anaesthetic hazards, and the interpretation of images. One or more of the stations may involve the use of a medium fidelity simulator. Pre-pandemic (March 2020), `follow on' stations were being tested in which a second station followed up on the information obtained in the station that came before.

The exam contains items from all areas of the syllabus, but with the majority centred around physiology, pharmacology, physics and clinical measurement. Anatomy and statistical knowledge are also included. At the time of starting this review, the Primary FRCA MCQ contained 90 MCQs: 60 MTF questions and 30 SBA questions. The 60 MTF questions assessed the following subject areas:

- 20 questions in pharmacology
- 20 questions in physiology, including related biochemistry and anatomy
- 20 questions in physics, clinical measurement, statistical methods and data interpretation

The 30 SBA questions are taken from any of the categories listed above.

In accordance with the Curriculum Assessment Group (CAG) document submitted to the General Medical Council (GMC) in 2019, this examination is undergoing a phased transition to increase the number of SBA questions while decreasing the number of MTF questions. The current format in academic year 2021-2022 is 45 MTF questions and 45 SBA questions. This ratio changed again in September 2022 to 30 MTF questions and 60 SBA questions. The final phase will take place in 2023, with the final, agreed format being informed by the analysis of each phase of this transition and best practice guidance from the external review, which started in January 2022.

Over its migration to an increased number of SBA items, there has been more opportunity to test candidates' understanding as well as basic knowledge (this is addressed in the section Discussion and Recommendations - Primary SOE). There have been many examples of feedback from candidates regarding the exams' ability to assess understanding. Candidates on the whole seem to appreciate this and see it as a positive attribute.

The passing mark for this exam is determined by a modified Angoff method. A definition of a '*minimally competent candidate*' has been agreed and questions are judged by a panel of experts who reflect on this standard which is determined for the stage of training.

A full audit trail for this exam exists covering: item shortlisting, new item development, item selection/inclusion, final item selection for examination, text proofing, candidate mark data, item level performance review, item removal, Angoff standard setting data, consideration of reliability, triangulation, cut score agreement, consideration of rater concordance and inter-rater reliability, and candidate result feedback (correct and incorrect items shown by item curriculum code).

The Primary MCQ has a test plan and is fully blueprinted; this is recorded in Annex B of the curriculum document; all areas of the curriculum detailed in the document are tested.

The Primary Objective Structured Clinical Examination (OSCE)

The Primary OSCE is a summative assessment of a candidate's clinical and communication skills, and applied technical knowledge of anaesthetic equipment, clinical monitoring and measurement. It comprises 16 assessed stations with no `killer' stations. Candidates have one minute to read the instructions for each station and five minutes per station for the assessment. To allow additional candidates to be examined, one or more `rest' stations may be added; candidates are not examined in these stations. The Primary OSCE stations currently assess resuscitation, technical skills, anatomy (general procedure), history taking, physical examination, communication skills, anaesthetic equipment, monitoring equipment, measuring equipment, anaesthetic hazards, and the interpretation of images. One or more of the stations may involve the use of a medium fidelity simulator. Pre-pandemic (March 2020), `follow on' stations were being trialled in which a second station followed up on the information obtained in the station that came before.

The Primary FRCA OSCE is taken together with the SOE and is blueprinted to Stage 1 of the anaesthetic curriculum. Candidates must pass both parts to be awarded the FRCA Primary. A successful candidate will have demonstrated the clinical skills and applied technical knowledge across multiple clinical scenarios required of a competent Stage 1 doctor in training. The OSCE is the only component using a simulated clinical environment, including communication with simulated patients.

The Primary Structured Oral Examination (SOE)

The Primary FRCA SOE is a summative assessment of a candidate's knowledge and understanding of the basic sciences; the foundations upon which further clinical knowledge is based. It also assesses clinical decision making and knowledge of equipment used in Stage 1 training. It comprises two parts, each lasting 30 minutes:

- SOE 1: an oral examination consisting of three questions in pharmacology, and three questions in physiology and biochemistry.
- SOE 2: an oral examination consisting of three questions on clinical topics (including a critical incident), and three questions in physics, clinical measurement, equipment and safety.

The SOE is taken together with the Primary FRCA OSCE, and is blueprinted to the Stage 1 training curriculum. The SOE complements formative assessments undertaken in the workplace, and provides assurance that candidates have reached the accepted national standard of knowledge and competence to progress to Stage 2 training.

The Final FRCA examinations

The Final FRCA Written examination

The Final FRCA written examination is a summative assessment, blueprinted to Stage 1 and 2 training curriculums. It assesses the knowledge required of an anaesthetist in training at the end of Stage 2 training in anaesthetics. It comprises two parts, the Final FRCA MCQ examination and the Final FRCA CRQ examination, both of which must be sat within the same diet.

The CRQ examination complements the MCQ paper, and whilst both parts test factual knowledge and understanding, the CRQ assesses judgement and the ability to prioritise information. The written examination currently forms a `gateway' to the Final FRCA SOE, which assesses application of this knowledge.

The Final FRCA CRQ examination contains 12 questions that are blueprinted to the Stage 2 curriculum. Each question is tagged to the Primary and Final FRCA syllabus. The CRQ paper assesses the mandatory areas of training. These areas may appear as stand-alone questions or as part of a question, for example pain management within a perioperative case. In addition to specific knowledge-based competences, examination material may be developed from guidance or recommendations published by healthcare organisations.

The public expects doctors to keep up to date with important developments and such material may be examined under the collective umbrella of 'professionalism'.

At the point at which this review was conducted, the Final FRCA MCQ contained 90 MCQs: 60 MTF questions and 30 SBA questions. The 60 MTF questions assess the following areas:

- 20 questions on advanced sciences to underpin anaesthetic practice
- 40 questions covering generalist and specialist topics within Stage 2.

The 30 SBA questions divide as follows:

- 15 questions on generalist topics
- 15 questions from specialist topics within Stage 2.

The Final MCQ is undergoing the same phased transition as the Primary MCQ to increase the number of SBA questions while decreasing the number of MTF questions. The current format in academic year 2021-2022 is 45 MTF questions and 45 SBA questions, changing to 30 MTF questions and 60 SBA questions in September 2022. The final format is set for release in 2023 and will be informed by the analysis of each phase of this transition and best practice guidance from the external review, which started in January 2022.

The Final FRCA Structured Oral Examination (SOE)

The Final FRCA SOE is a summative assessment of a candidate's knowledge, understanding and decision making abilities in clinical anaesthesia and the applied underpinning clinical science. It is the last component of the FRCA examination to be taken, and successful candidates are eligible for Fellowship by Examination of the RCoA. The SOE complements formative assessments undertaken in the workplace and provides assurance that candidates have reached the accepted national standard of knowledge and competence to progress to Stage 3 of anaesthetic training.

The Final SOE contains two components, SOE 1 and SOE 2:

- SOE 1 has two parts, Part A and Part B, which are taken consecutively. Each part is 26 minutes long and comprises two clinical short cases, each with a linked clinical science question. The clinical science question may come before or after the clinical short case. Candidates have 13 minutes to complete each short case and linked question.
- SOE 2 comprises a two-section clinical long case followed by two stand-alone clinical short cases taken in one sitting. This SOE is 36 minutes long, with 10 minutes to view clinical material, 13 minutes for the two-section clinical long case, and 13 minutes to answer the two clinical short cases on clinical anaesthesia, which are unrelated to the clinical long case. The SOE tests generalist and specialist topics from within Stage 2 of training.

7. Context of the Review

The review of the FRCA examination is undertaken every three to five years alongside ongoing improvements and innovations that are approved by the Examinations Committee, the Education, Training and Examinations Board, and Council. The last holistic review of the examination was conducted in 2015 and led to significant changes in the examination that took time to develop and obtain approval from the GMC.

This review of the FRCA examination began in January 2020 and the Review Group held two meetings up until March 2020, where the emergence of the global COVID-19 pandemic meant that the process needed to be paused. This was necessary as the examiners and exam staff needed to focus on transitioning the FRCA, FFICM and FPM examinations into an online format in a very short timeframe in order to continue delivering exams and supporting UK training programme and career development. The result was that between March 2020 and March 2021, all review work was postponed in order to divert resources to the operational delivery of the examinations during the pandemic.

In December 2020, the Examinations Committee agreed a proposal to restart the examinations review in 2021 during the pandemic due to the necessary changes that would be required in the examination, which would inevitably need to be planned and scheduled into the development programme.

During the early meetings of the Review Group in 2020, it was agreed that the current examination structure was well understood and worked well in terms of producing outcomes that training programmes supported and that were broadly in line with performances and professional competence in the workplace. However, the Group recognised that the examination had received criticism from candidates around the content, structure, assessment burden and style, and from external observers (RCoA, 2017, Clyburn *et al*,[4] 2022, Subramaniam *et al*, 2022).[8] The Group were keen to include this feedback in the ongoing discussions throughout the review process.

The Group agreed to keep an open mind during discussions and start from a basis of best practice and patient safety first, rather than opting for models of assessment that were either `traditional' or `well liked'.

8. Discussion and recommendations

I. External involvement

Externality of the review process was achieved by the involvement of Prof Richard Fuller, Deputy Dean of the School of Medicine, University of Liverpool, and consultant for the Health Professional Assessment Consultancy (HPAC), who attended meetings and held component specific meetings with separate examiner groups on particular issues that had been raised in feedback. Prof Fuller was briefed at the start of the review process, sent agendas and papers, and attended many of the meetings. Comments from Prof Fuller were fully incorporated into the discussions and decision making of the Group.

Recommendation 1

The Group agreed that external input and challenge of the FRCA examination and its processes was a positive thing and examiners and the examination as a whole have benefited hugely from exposure to HPAC training programmes and conferences. Moving forward, the FRCA examination should ensure that there is externality included within the development of the examination in order that it continues its journey towards best practice in medical education and assessment.

II. Assessment strategy

The Review Group agreed that the various examination components should continue to be developed, designed and aligned to produce a coherent package of assessment in Anaesthesia, Critical Care, Pain Medicine and Perioperative Care.

In terms of the FRCA, the Group agreed that the Primary and Final examinations should remain as complementary summative assessments that together form one examination that leads to the Fellowship of the Royal College of Anaesthetists. This was an important principle to agree early on as the examinations could be perceived as individual examinations. The Group therefore felt that the examination should continue to be referred to as the `FRCA examination', an examination comprising two parts: a Primary examination and a Final examination with each part containing written and clinical components.

To affirm this principle of a coherent package of assessments, the Group further agreed the following principles for the examination and the process for review.

- **Ensuring best practice in assessment methodology:** the examination should learn from contemporary and validated ways of delivering the examination. As part of the development and evolution of the examination, validated assessment methodologies will be sought, tested and applied to make improvements to the examination.

- **Implementing best practice standard setting methods in all components of the FRCA examination:** the standard setting methods used in the FRCA examination will follow recognised and validated methods for medical and professional assessments and not be based upon traditions within the exam.
- **Aligning standard setting across the examination:** where a standard setting method or approach is used in one component of the examination, it should also be used in the other. The examination as a whole should explore and strive for best practice in standard setting and seek to adopt and align that across the examination.
- **Breadth and depth of the examination syllabus requirements:** the examination should be mapped to the appropriate areas of the curriculum/syllabus. The examination will test appropriate depth of knowledge across syllabus domains. This will be set out by the establishment of a new examination blueprint and sampling grid that will be used across the examination.
- **Examiner recruitment, retention and requirements:** the requirements for becoming an examiner must be appropriate and fit for contemporary ways of working and aligned to the requirements of the examination.

Recommendation 2

The Review Group agreed to establish improved core examination documents (the examination syllabuses, blueprints, test specifications) to ensure that questions are appropriately set and derived from the curriculum. These documents will also help demonstrate and guide stakeholders to the appropriate level of depth and breadth that the examination tests.

Examination structure, format and constructs

The Group felt strongly that, when reviewing the format, structure and constructs of each component of the FRCA, they should only consider outcomes to the review that do not introduce or maintain any duplication in the assessment process, such as two components assessing the same thing.

Assessing knowledge and understanding

The Group agreed that it was essential that the examination assessed knowledge, understanding and performance throughout the summative FRCA. Furthermore, that a desirable aim was an examination that can assess a broad scope of the syllabus in sufficient depth at appropriate times in the training programme. There was significant discussion about whether written and clinical elements should be repeated in Primary and Final on the basis that few Royal Colleges have an OSCE component at Part 1 (Primary) or a written component at Part 2 (Final). Fundamentally, the Group questioned whether the presence of a written and clinical component in each part of the FRCA represented a duplication of assessment or a reasonable and appropriate sampling of the syllabus.

The Group felt that there were various areas of the current examination model that enabled an assessment of knowledge and understanding. The SOE offered a good opportunity for the assessment of knowledge and understanding, but the Group questioned whether this was best assessed in a written test (within an MCQ) or verbally (within an SOE). The Group accepted that the OSCE component could also assess knowledge when placed within a clinical scenario.

Since online diets have been introduced, the MCQ group had taken advantage of the new format in terms of its development of SBA items, with an increased number of SBA items now designed to test the candidates understanding. The style had been used since the original use of the SBA format, but the online platform afforded the ability to use items with:

i. a diagram to interpret
ii. datasets comprising clinical data on which the candidates are asked to make judgements

iii. tables of statistical data to interpret
iv. simple calculations to complete which test the application of principles in physics and pharmacology
v. theoretical previously unseen scenarios to apply general principles of understanding, for example hypothetical new drug data.

Additionally, there is the future possibility of using short video sections to look at understanding many areas, such as hazard recognition, understanding of moving image ultrasound, and three-dimensional anatomy. Some of the items used for OSCE, which are not true OSCE questions, could be well-suited to inclusion in such an exam.

The Group noted that since adaptive testing was becoming more and more interesting to assessment boards, the assessment landscape in five years' time may open up new ways of assessing and therefore it may be pertinent to `tread water' and withhold making major changes to the Primary written exam in order to see what technology brings. The point was made, however, that adaptive testing is a huge challenge for assessment providers and requires very large item banks.

The MCQ components

Candidate feedback on the Primary and Final MCQ showed that the exam was more than just a test of knowledge and needed an understanding of what was being asked. This reflects the item development work that has been taking place in the Primary and Final groups described above, in which SBA items have been created to test candidates' understanding.

Since the start of the phased transition in 2020 to increase the number of SBA questions in this examination component, SBA results are showing divergence from MTF questions in that those candidates who do well in the SBAs do not necessarily do well in the MTFs and vice versa. The Primary and Final core groups for MCQ have been reflecting on the direction these components should take in terms of the final structure. Aspects that have been discussed are, for example, whether the exams should be 100% SBA or include another item type, whether the total number of items should be increased beyond 90 items to include more data points, and if increased, to what new amount.

Advantages of written examinations

Advantages of written papers are centred around demonstrable validity. Many examiners are involved in the setting of the examination. The current core groups comprise members with considerable ethnic, geographic, gender and sub-specialty diversity, and members who are at different points of the examiner tenure. All contribute to the delivery of each paper. All candidates receive the same examination, there is consistency in the items, their delivery and in the marking scheme and method of application (machine scored). The current development strategy for the item banks ensures that approximately 70% of items have previous run data and these data can be used in pre-equating to achieve a consistent facility, ensure consistent horizontal equating, and inform availability and future use of unchanged test items or those with minor alteration after an appropriate period. This enables a healthy and well-designed bank of items to be assembled over time. Additionally, since all candidates receive the same exam, psychometric analysis is performed easily and accurate reliability data can be recorded.

To achieve a good item bank, considerable investment in examiner training is required. The writing of high-level performing SBA items may not be suited to all writers since these items do not necessarily sit in the normal skill set of a practising anaesthetist in the same way that oral discussion and exploration do.

Limitations of written examinations

The limitation of the written examination is that, in comparison to the SOE, the extent to which a candidate understands a particular area cannot be explored in as much depth. For example, simple computational error or misunderstanding of the details of an item could result in a candidate with partial knowledge arriving at an

incorrect answer. In comparison, in the oral situation, a skilled examiner may be able to circumnavigate these issues and be better able to explore and assess the depth of understanding. The SOE allows the opportunity to assess how candidates can organise and present concepts and show their appreciation of the relevance and importance of concepts, where some are key and others peripheral to the subject area. This is difficult to test using other formats, including the CRQ paper.

The Review Group agreed that although the writing of good SBA items was a complex task, the SBA model was one that should continue to be pursued since MTF had been all but phased out across other medical postgraduate examinations on recommendation from the GMC. However, the Group also recognised the likely consequence of this, for example, in terms of examiner time, the need for SBA-specific training and the candidate view on the difficulty of answering SBA questions.

In terms of what the final format will be in 2023 – the last phase of the transition plan – external advice will be sought (Recommendation 1 to include externality in exams), and a decision will be made after careful analysis of the data over the three phases by a new Examinations Development and Assurance Group, to be formed in 2022.

The Group noted that currently there is considerable variance in process between the Final and Primary MCQ exams. It agreed that there should be consistency between those components with regard to standard setting and item analysis. Additionally, the ability to store items with their associated data and access them easily was a prerequisite for a successful and professionally delivered written exam.

Final Constructed Response Question (CRQ) component

The Review Group reviewed the extent that the CRQ exam was able to assess understanding in the examination and remain a complementary component to the MCQ examination, comprising SBA items. The Group also considered whether the CRQ examination would be an appropriate additional written assessment for the Primary examination, again sitting alongside the current Primary MCQ component.

The format of the CRQ is twelve questions in total, with each question comprising a number of sub-questions that provide a total of 20 marks per question. The lead for the Final CRQ exam stated that the multi-part format of the CRQ makes it difficult to maintain the independence of questions because questions early on can give away an answer to subsequent questions. This makes question construction a difficult and time consuming process.

The Group looked at whether a CRQ exam could be a replacement component for the Primary SOE. The consensus of the Group was that a CRQ in the Primary exam may not be the best choice of exam format due to the difficulty of maintaining question independence, and the amount of work involved in writing questions, which is disproportionate in terms of the effort and time to produce an SBA MCQ examination.

The Group also considered the utility of the CRQ and agreed that due to the answering mechanism for this exam, it would almost certainly require: a) an examiner or, in the absence of examiners, b) expensive software or a platform to mark and produce results quickly. Since the Primary can be sat by practising anaesthetists from overseas, a written response in another language may create bias and widen differential attainment more than a selected response format exam. Further discussions on this topic are included under the section on the Primary SOE.

The Group considered the potential for the CRQ to assess understanding of basic science. Whilst the Group noted that it was possible to do this, such questions would be difficult to write. For example, to fully delve into assessing understanding by asking 'why', the answer to the previous question would essentially be revealed. The Group considered whether the current online platform used to deliver the FRCA written examinations had the ability to alleviate some of the concerns around item independence. Although the platform has the ability to deliver exams in a locked, linear format (`locked' means that once a question has been answered, it is not possible for a candidate to return to a previous question), it would be very different to the way exams are currently delivered in which candidates can move forward and backwards through the test. It was therefore felt that this would require further exploration in the next stages of development for the CRQ examination.

In addition to changing the way an exam is delivered, various types of visual format questions can be developed on the TestReach platform. This would provide examiners with the opportunity to use different styles of SBAs and elements of these visual resources could be reused in subsequent papers or for different questions, thereby enabling the item pool to be expanded more efficiently. Finally, the Group discussed a return to single-use CRQs. However, although a greater number of resources for candidates are needed to support preparation for this exam, single use is an issue once the time and resource to produce an item is considered. The point was made that Very Short Answer (VSA) questions could be an alternative to CRQ since it is easier to maintain question independence and they can potentially be machine marked. Although it is easier for this question type to be leaked over time due to the brevity of the format, the Group felt that if the question bank is of a reasonable size and quality, this should not be an issue. Furthermore, the RCoA is in essence just assessing the syllabus, which is in the public domain.

Recommendation 3

The Group were in agreement that the current Primary and Final SBA and CRQ formats should continue. The CRQ is only used in the Final exam and this should not change. More consideration of VSA questions was required and this will form part of the discussion around the final format of the MCQ examinations in 2023.

Assessment of clinical knowledge and skills

The Primary SOE assesses understanding of scientific principles and their application to clinical practice. Candidates receive three questions in pharmacology, three in physiology, three in clinical practice and three in physics. The format allows examiners to probe deeper into a candidate's understanding of these areas.

The Final SOE assesses clinical anaesthesia and clinical science. The assessment of clinical anaesthesia complements workplace-based assessments to examine the understanding and theoretical application of this knowledge in clinical practice through a standardised format. Candidates need to appreciate the anaesthetic significance of clinical situations and demonstrate correct interpretation of clinical investigations in the planning of perioperative anaesthetic care. Candidates support their chosen management options with well reasoned discussion based on sound scientific principles. The aspect of clinical science assesses a candidate's understanding of basic medical science as applied to the practice of clinical anaesthesia, intensive care, and acute and chronic pain management.

The Group felt strongly that there should be a clinical summative assessment in both parts of the examination. The Group reviewed the structure of the new curriculum and agreed that anaesthetists would need to be clinically assessed both independently and in the workplace after Stage 1, as this would be a stage in practice where more distant supervision is started to be introduced for clinical practice.

The Group considered the role of each component in the clinical examinations; the SOE and the OSCE. They spent several meetings discussing the merits of the SOE examination component in both the Primary and Final parts of the examination. The Group heard proposals as to the extent to which the content and material in the SOE examination could be more efficiently and objectively assessed via a written examination, particularly when considered against proposals to robustly assess knowledge and understanding via the use of SBA-type questions within the Primary MCQ. This would help address the criticism of too much basic science in the SOE component of Primary.

Objective Structured Clinical Examination (OSCE)

An OSCE can be defined in the following way "A multi-station clinical examination (typically having 15 to 25 stations). Candidates spend a designated time (usually five to ten 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" (AoMRC 2015).[2]

The Group felt strongly that it was appropriate and reasonable to continue to have an OSCE within the FRCA and within the Primary examination more specifically. It did not feel that an OSCE would be a useful addition to the Final part of the FRCA.

The Group agreed that as a clinical examination assessing skills, the Primary OSCE is correctly placed at the end of Stage 1 training as it ensures an appropriate level of knowledge and practice before moving to stage 2.

The Group received detailed reports by the OSCE working party (WP) on developments to that exam which will ensure that pure knowledge, which could be tested in a written format, will not be assessed in OSCE. The WP has been involved in ongoing work to review and evolve the examinations incrementally over time. This work has been incredibly useful background to understanding the current processes, benefits and deficiencies of the current OSCE model.

The WP had been attempting to improve the quality of the scenarios within the constraints of the examination structure by piloting the concept of `follow-on' stations. It is believed that these stations can allow for a deeper and more authentic probe into subject matter or a clinical case over the course of two stations, while compensating for the relatively short amount of time (five minutes) allocated in each station; other OSCE examinations tend to provide seven to eight minutes of assessment per station.

The Group felt that the current delivery of OSCE did not fully fit the description of an OSCE assessment due to the considerable number of knowledge-only stations. This was particularly noted in the unmanned kiosk stations where there is no candidate/ examiner interaction. It was felt that these types of stations lend themselves better to be tested in an MCQ written context rather than an OSCE.

In considering best practice for OSCEs, the Group considered whether 14 stations would impact on the validity of the exam and therefore whether 16 stations is the optimum. Balanced with this is a need to ensure sufficient capacity for the candidates that book and an increase to 17 or 18 assessed or rest stations would accommodate greater numbers of candidates.

Overall, it was felt that the FRCA OSCE should probably consist of approximately 14 stations, two stations fewer than the current total of 16 assessed stations. This was based on the removal of fact-based stations in the examination that are better suited to a written format.

The Group agreed that there is potential to increase the station duration beyond the current time of five minutes, but that the total time should be dependent on the scenario. Short stations were felt to support discrimination as there is insufficient time to fully assess the candidate. It was therefore felt that best practice supports longer stations but fewer of them. It was noted, that if such changes were implemented, there may be a knock-on effect to capacity.

In terms of standard setting methodology, the Group noted that previous trials of borderline regression method (BLR) with a checklist and global score had not been successful due to the construct variance caused by the kiosk stations present in the OSCE at the time. The Group agreed that, moving forward, the exam should utilise BLR with domain marking to enable skills, knowledge and behaviours to all be assessed in an objective way. It was noted that the Primary is currently working towards `follow-on' stations with domain marking, and that this work needs further development.

The Group had some concerns about whether the curriculum would be covered in its entirety with an approach of reduced assessed stations and follow-on stations. The Group agreed that the simulation station, interactive resuscitation, history and communications stations should all remain as core stations and that there may be a possibility of incorporating anatomy, technical skills, physical examination and the applied sciences (pharmacology, physiology) into these. Other more knowledge-based questions could be moved to a written format. Overall, a key aspect is that stations should move away from being a series of questions.

Recommendation 4

The OSCE should remain a fundamental clinical component in the Primary examination to assess performance in a simulated environment appropriate to the new Stage 1 of training.

Recommendation 5

The kiosk stations should be removed as these stations do not represent a true OSCE assessment. The contents of these stations could be converted into SBA questions in the Primary MCQ examination and clinical data interpretation stations in the Primary OSCE.

Recommendation 6

The OSCE should be updated and reformed to reflect an authentic OSCE assessment as defined by the Academy of Medical Royal Colleges and the GMC, for example, there should be a reduction in assessed stations and an increase in station duration that is in line with the constructs being assessed.

Recommendation 7

The OSCE should have a clearly defined set of areas of practice that it assesses and a separate test specification that supports the sampling of examination content against the curriculum.

Structured Oral Examination (SOE)

The SOE components

The Group noted that the Primary SOE played an important part in preparing candidates for the Final SOE and that the format of these exams enables examiners to test candidate understanding of subject matter in some depth. However, prior to this review, the Group was made aware of various potential deficiencies with the SOE format of examination from both internal and external sources. It was also aware that a viva-style exam format is no longer a favoured format for postgraduate medical exams and is only used by a few Royal Colleges.

The Group felt that the weakest component in terms of a justifiable assessment argument for its use was the Primary SOE. This SOE contains a lot of basic science and that examining this verbally might not be the best and most efficient way to assess these aspects of the syllabus. The SOE also has a tendency to resemble a question-and-answer form rather than an authentic clinical discussion. The Group therefore considered whether this component could be removed from the examination. However, despite these arguments, removal would only be justifiable if the MCQ component of the examination was updated to include a full suite of SBA-type questions that could reasonably and effectively assess both knowledge and understanding in written form rather than verbally. If this is possible, the Group agreed that the removal of the Primary SOE component from the FRCA would be a reasonable pathway to consider.

In addition to moving elements of the SOE into SBA questions, the Group also discussed at length potential changes that could be made to the OSCE to further support the removal of the SOE. It concurred that if the OSCE was updated to properly assess clinical knowledge, skills and performance, then there would be no clear need for the SOE component. The Group noted that it would be reasonable to remove the SOE if the new OSCE format and change to the SBA was made clear to stakeholders. A discussion and review of the OSCE component is considered further on in this report.

The Chair concluded the discussion with a request for more detailed information on the ability to assess understanding and applied knowledge in written form.

Primary SOE examination

Consideration of the future of the SOE was discussed at several points during the review process and there was an agreed appreciation among Group members of some of the limitations of SOE in terms of consistency for candidates, fairness and hence validity.

During the course of the review, the Group evaluated the extent to which the material and content of the Primary SOE could be assessed in other ways or in other components of the exam, such as the MCQ, OSCE or CRQ. The Group concluded that the Primary SOE contained approximately 75% basic science content.

The Group noted the view that the SOE enables candidates to develop skills in retaining knowledge and `thinking on their feet' while responding to questions, and the Group accepted that these were valid skills to have in anaesthetic practice. However, the Group did not feel that within the scope of basic sciences this was an appropriate requirement to be retained.

The Group acknowledged that moving the SOE basic science questions to a written format would remove the aspect of examiners being able to probe for further information. However, the Group felt that although the ability to elicit information through probing is a useful and valid assessment, it is more suited to the contextualising and explaining of clinical decision making present in the Final SOE (see Final SOE section) and therefore is not an implicit requirement of the Primary SOE. In terms of moving SOE questions to a written mode, previous discussions on this subject that were supported by external, expert advice and example questions made it clear that higher levels of cognition can be examined in a written format. In reviewing the remainder of the content of the SOE, the Group acknowledged that the clinical element could be modified easily for use within OSCE stations.

CRQs for basic science might be possible but would be time consuming to write for each exam and items could not be used in the same format again for a very long period. The CRQ format did not offer the same advantage as a bank of test items and needed much more examiner input at the marking stage. Currently, few colleges use this type of format; RCEM continue to run an SAQ, RCPYSCH are in the processing of evaluating the utility of VSAQ format.

Two shorter SBA papers, each with a different focus, was the format which received the most support. Very short answer questions were thought to be a path away from the testing of higher-level understanding as they are designed to test knowledge rapidly.

Extended Matching Questions (EMQs) were not supported by anyone in the panel and the external panel member commented that they were not considered particularly favourably by many researchers.

Principles and needs that were agreed:

i. experienced and trained psychometric support
ii. need to retain the testing of higher-level understanding
iii. consistency for candidates in experience
iv. improvement in professionalism of delivery
v. consistent brand across all parts of examination
vi. high functioning, user friendly, efficient platforms
vii. need for consistency of standard setting between diets and between parts of the exam
viii. further examiner training to underpin the changes to delivery.

In conclusion to these discussions, the Group agreed that the Primary SOE content would be more efficiently and objectively assessed in other parts of the examination with no loss to validity. The Group therefore recommends a re-evaluation of the SOE component in the Primary examination which considers the extent to which elements of knowledge could be tested in a written paper, and other elements, such as the testing of clinical performance and the basic sciences that underpin that clinical performance, incorporated into a reformed OSCE. This recommendation is based on the following reasons -

- The SOE examination is deemed a subjective examination by internal stakeholders, external experts and the GMC due to the potential lack of parity in the candidate experience caused by the format of this particular type of exam.
- Throughout the course of the review, the Group felt confident that a well-constructed SBA question had the ability to test understanding and higher levels of cognition.
- The Group modelled how to write some of the SOE materials (basic science questions) into an SBA-type format.
- The examiner time required to set and deliver an SOE examination far exceeds the challenge and time required to create good SBA questions.
- Both clinical elements and basic science questions in the SOE examination can be converted into material suitable for delivery in the OSCE examination by, for example, placing basic science questions in a clinical context.

Recommendation 8

The removal of the Primary SOE component from the FRCA and for the materials from the SOE to be converted into SBA questions and OSCE questions. The Primary MCQ may need to extend the range of SBA questions to ensure that the appropriate content continues to be assessed across the FRCA.

Final SOE examination

The Review Group scrutinised the validity arguments for an SOE examination in the Final part of the FRCA examination. The Group noted the challenges and criticisms of an SOE format and reviewed the changes that had been made to the Final SOE following the 2015 examination review to lessen some of these criticisms around, for example, bias and subjectivity.

The Group compared the Final SOE format and delivery to the Primary and noted in Final the benefits of rotating candidates through three pairs of examiners to reduce possible bias and the candidate-centred method of deriving the pass mark.

The Group highlighted some weaknesses with the current format, namely the limitation of the 0-1-2 marking structure, which does not allow for clear discrimination between the borderline and the excellent candidate. BLR was not felt to be of substantially greater utility than the fixed pass mark currently used for the Primary SOE as regardless of the cohort makeup, the cut score generally fell between 39, 40 or 41.

In this evaluation, the Group recognised a key difference between Primary SOE and Final SOE. In the former, although a candidate is be expected to be able to operate safely, they are also expected to recognise when support is needed and call for assistance as required. In comparison, the expectation of a candidate at Final level is that they are essentially ready to move to much more independent practice. Basic science questions are currently used in both Primary and Final SOE, but the Final leads had been considering the extent to which these were needed in order to formulate a management plan for a clinical scenario.

The external expert, Prof Fuller, noted that the position of the Final SOE examination in the assessment strategy was appropriate because these tend to involve more senior candidates who need to be assessed on their ability to make valid decisions on what course of action the care of a patient should take.

In conclusion, the Group agreed that the Final SOE should continue to be a clinical component in the Final part of the FRCA examination, but that some parts of this component may be modified in the future. It was agreed that future discussions should include potential changes to the marking system, such as domain marking, and for these domains to differ according to the needs of the question.

Recommendation 9

The retention of the Final SOE in the FRCA exams strategy as a valid mode to assess knowledge, understanding and clinical decision making.

Recommendation 10

A new and appropriate method of standard setting for the Final SOE examination should be investigated with a view to being developed, piloted and implemented in the exam.

III. Standard setting

The Review Group felt that the written papers appear to be setting a 'low bar' when arriving at a cut score. The current method for deriving the Primary and Final MCQ pass mark is via a modified Angoff less one standard error of measurement (SEM). The pass mark is then rounded down at each point of calculation. The Group recognised that the removal of an SEM is not in keeping with assessment best practice or practice in other medical Royal Colleges and should therefore be discontinued.

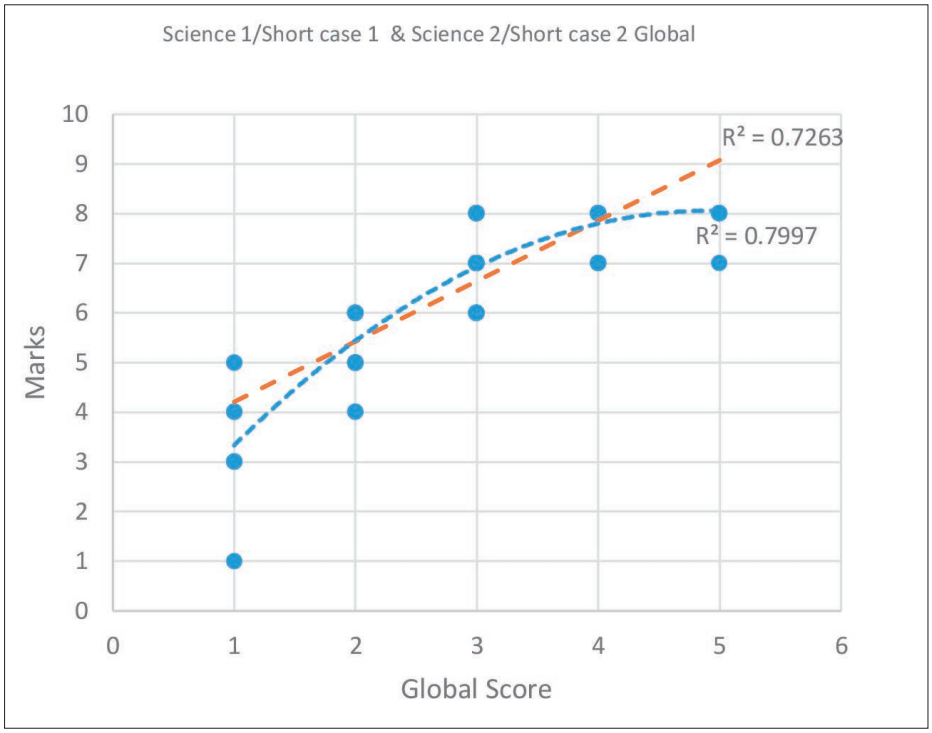
Linked to the removal of the SEM, the Group discussed the aspect of a 'gatekeeper' function for the MCQ exams as points of assessment which ensure candidates have sufficient knowledge to be able to take the next part; Primary OSCE SOE, Final SOE. The Group did not feel that this was justification enough for reducing the passing score via removal of an SEM. More detailed information on this subject appears later in this report.

Borderline Regression (BLR) was discussed in detail during this review since it is considered best practice in terms of standard setting for OSCEs and is used in the FRCA but only in the Final SOE. BLR is a linear regression method which correlates the marks awarded and the global score for each part of the exam by examiners. The borderline score becomes the pass mark.

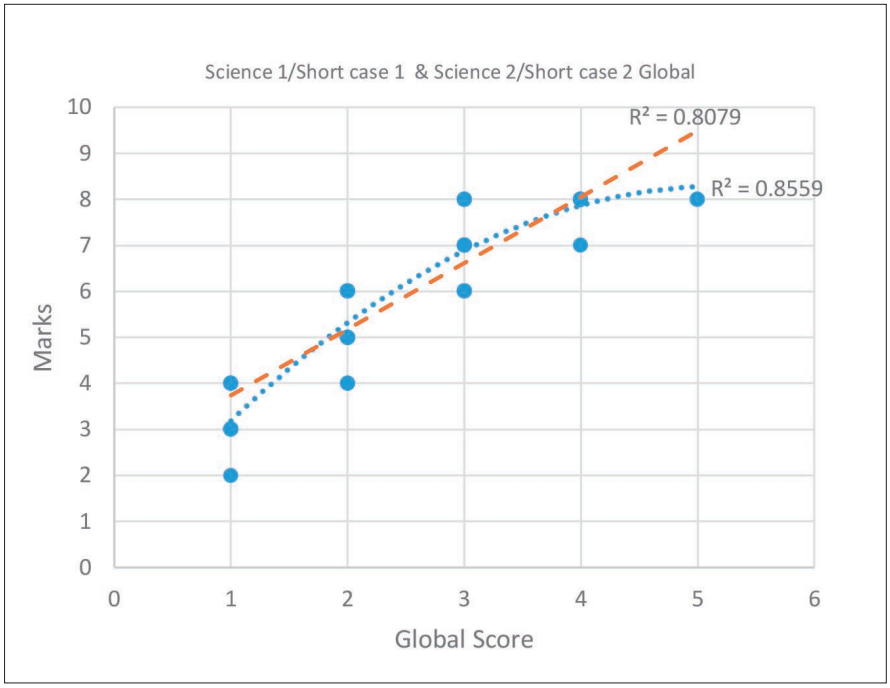
The Group reviewed an analysis of data from the Final SOE diets during the pandemic and compared this to pre-pandemic, face-to-face diets. Strong relationships between the marks given by the examiners and the global scores were evident in pre- and post-pandemic diets. The Group felt this relationship is to be expected since the limited range of marks for performance (0-1-2) only allows for a global view of a candidate's performance in each question and this is therefore likely to be reflected in the additional global score on overall performance. On a number of exam days, the apparent best fit was not a linear model and the reason for this was felt to be this narrow range of marks which does not allow for a distinction between, for example, those candidates who did very well to those who were just competent. Examples of this are shown in graphs 1 and 2 for face-to-face and online delivered examinations.

In the graphs we can see that candidates awarded a global score of a clear pass (4) are achieving full marks and those with an excellent global score (5) are also achieving full marks, and it is this combination of scores that means a linear model no longer becomes the best fit. The ceiling effect created by such results, shown at the top of each graph, could also be the result of discrepancies in how examiners are applying the global score to the total scores. A point was made that the examination should not be aiming for full correlation between the global score and the marks awarded because a perfect correlation would suggest a close similarity between scoring systems.

Graph 1: distribution of global scores for a face-to-face examination



Graph 2: distribution of global scores for an online examination



The Group felt that for BLR to work effectively, the marking system would need to be symmetrical about the central core of the data, ie good candidates. However, BLR will only work well in this setting if a way can be found to make the marking system more granular and the Group felt that this may be difficult to achieve in an SOE.

The Group reflected on the fact that BLR with domain marking had already been trialled in the Final SOE, but that attempts to implement this within the structure of the SOE had been unsuccessful. This may be due to the fact that BLR does not fully suit the style of the SOE and that there was not a suitable level of sampling across a range of domains. A more suitable application may be to have different domains according to the needs of the question, but again this would not be easy to implement. The Group reflected on whether to move away from using BLR, but there was no clear alternative beyond the test-centred standard setting method, Angoff.

The Group considered the utility of BLR and domain marking in the OSCE assessment but recognised that there were additional factors that first needed to be implemented for this to be successful. Such factors are ensuring that the examination represents a `true and authentic' OSCE that tests clinical skills, and increasing the duration of the stations from five to seven minutes in order to have an appropriate amount of time to assess these skills.

The Group explored the concept of using a Rasch model of standard setting in which each examiner/ candidate interaction could be assessed for the contribution from the examiner, from the candidate and from the question, and thus create a cut score. The Group understood that this method of modelling was complex but was something that they felt should be explored and appropriately implemented in the examinations. The Group agreed to take this action away and explore the viability of Rasch.

Recommendation 11

The Group recommends a review of the standard setting methods in the OSCE and Final SOE components of the examination, with the view of making recommendations for the most appropriate standard setting methods for the respective examinations.

IV. `Gateway' examinations and the SEM

In many of the documents and descriptions of the examination there is reference to the written component as a `gateway' examination to the next clinical component of the FRCA examination. This has been a subject of debate for some time in terms of understanding what this term means and whether in a high-stakes exam with a patient safety aspect, this is appropriate.

Clearly, since the exams cannot be taken in any order (a candidate must pass the Primary MCQ to be eligible to sit the Primary OSCE/ SOE), the MCQ will always be classed as a 'gateway' exam, but does `gateway' imply something else, such as an easier exam to pass? The Group felt that in order to move away from such terminology, the regulations would need to be changed to allow the candidates to sit the exam parts in any order. There was no strong consensus regarding changing the order in which the components could be taken, nevertheless, it was agreed that a clear purpose of the examination as a whole and of the individual component parts was central to the ongoing strategy.

A result of further exploration highlighted that the term `gateway' in the context of the FRCA may be related to the standard setting method. A key element of this is the use of the Standard Error of Measurement (SEM). The SEM is a statistic that tells us by how much a true score might differ from an observed score. We assume that a candidate has a `true score' on the exam that reflects their ability on the construct being measured. The observed score of the candidate may be different as a result of error. The SEM is used to calculate a confidence interval around an observed test score. A 95% confidence interval falls at 1.96 standard deviations. We can be confident that the observed score will fall within a range of + or – 3. In a very reliable exam, error may mean that a score falls within a range of 7 raw marks. This is important when a score is close to a decision point/cut-off as we do not know with certainty whether the candidate is truly above or below the cut score.

The FRCA MCQ exams are standard set with application of the SEM, which is derived from Pearson product moment correlation statistics for reliability. The aim is to progress candidates who may have a realistic chance of success in SOE/OSCE. An analysis of retrospective data and modelling of the potential success of borderline candidates in MCQ on the next components of the exam (OSCE/SOE) has been undertaken. It was clear from this analysis that performance in the MCQ for these candidates did not determine good performance in the OSCE SOE. This raises the question why an SEM is removed since it does not necessarily strongly reflect success in the next component.

In the FRCA written exams, one SEM is removed in order to reduce the pass mark. This action increases the pass rate and provides the `benefit of doubt' to candidates. However, it is also possible that the action of removing one SEM undermines the Angoff process.

The reduction of the pass mark by one SEM is a historic strategy and there is no recent or known historic paperwork to explain the origin of this approach. The Group emphasised that the College is an outlier in this respect, since most Royal Colleges add one SEM to the cut score as a patient safety measure to control for `false positives' in the examination. If the College changed their approach, the pass rates would be predicted to reduce and perhaps to a level which would be unacceptable to some or all of the stakeholders. This is because going from the removal of one SEM (-1) to adding an SEM (+1) is making a change of two SEMs from the current method. The Group suggested that since the current method of Angoff is linked to the SEM, it could be recalibrated if we were to change our use of the SEM.

The Group were in general agreement that regardless whether the written exams are `gateway' exams or not, the removal of an SEM was less than optimal and should be reviewed. A suitable time to do this may be on moving to a full SBA paper in 2023. This should be accompanied by further modelling around the borderline candidates in terms of tendency to fail or pass the clinical, and the impact changing the application of the SEM would have on prior pass rates.

Finally, the Group discussed the differences in standard setting and item analysis processes between the Primary and Final MCQ exams and agreed that both core groups should meet to come up with one method aligned to best practice.

Recommendation 12

Further modelling should be undertaken on the performance of borderline candidates in the written exams and on subsequent clinical exams.

Recommendation 13

The FRCA should move to a position where no or one SEM is added to the pass mark at the point of transitioning to an SBA paper in 2023.

Recommendation 14

A group to be formed with the express aim of aligning standard setting methods and item analysis process between the Primary and Final MCQ examinations.

V. Psychometrics

The importance of psychometrics to assess item performance was emphasised at many points during the review. The external panel member confirmed the need for psychometric support to collate and analyse the data, provide data to examiners in a simple and usable format, and report on test performance to internal and external stakeholders and the regulator. This was considered in the context of the two separate platforms, TestReach and Fry IT's Practique, which are currently being used to deliver online examinations. The Group agreed that reliability data and individual item performance captured over time was vital to being able to deliver and progress the written examinations. The discussion included the extent to which item fatigue appears in long assessments and the ability to have statistical evidence of candidate misconduct that can be used to exclude reuse of affected items.

Differential attainment in association with differential item functioning should be scrutinised to look specifically at cohort differences, such as the performance of candidates with Temporary Examinations Eligibility (TEE) on the Primary components of the FRCA. A longer-term strategy could include collaboration with schools of anaesthesia, college tutors and TPDs to conduct predictive validity studies through the correlation of the summative assessments to continual workplace-based assessments.

At present, College exams use classical test theory in standard setting processes. Item response theory (IRT) is gaining in popularity in high stakes exams and future development should take a path which does not preclude its use.

VI. Delivery methods: online or face-to-face

Platforms and the user experience

The Group noted the use of two platforms for the different parts of the examination – TestReach for the written exams and Practique for the clinical exams. The Group recognised that the use of two platforms was informed first by the pre-pandemic landscape of traditionally delivered exams and a tender for an exams management system released in 2018, and second, a short-term strategy to maintain exam delivery through the pandemic; Practique was the only platform able to offer an online OSCE in the summer of 2020. The Group agreed that it may be less economic in terms of training and overall cost of delivery to continue with two platforms, however, it also recognised that after a series of presentations from competing suppliers, there was no one platform that emerged as a best fit.

In evaluating the platforms, the Group agreed that it was important for the user interface to be professional in appearance, consistent in terms of font, layout, units, and to have the ability to support reasonable adjustments.

The Group were in agreement that the online delivery was a positive step forward for the written exams in terms of efficiency in delivery and marking and benefits to candidates who do not need to spend time and money travelling to test centres. The online delivery of clinical exams was felt overall to have a greater number of disadvantages and was also a threat to examiner engagement if continued over the long term. Prof Fuller raised the concept of the `5Ps': purpose, programme, people, processes and product for the Group to consider as a way of evaluating the mode of delivery and creating an assessment use argument in general for exams.

Written exam delivery

The Group agreed that the candidate view of remote online written exams was generally positive, citing the ability to sit an exam in a more comfortable, less stressful environment, with no/less travel and reduced costs. Online delivery also provides a greater opportunity for overseas candidates to sit the Primary MCQ than offered by a UK test centre delivery.

A key concern was the need to run multiple cohorts for written exams, however, it was acknowledged that this will soon cease once the supplier's proctoring capacity increases to 500 candidates per sitting.

The Group agreed that the TestReach system offered a more reliable way of managing the CRQ exam since it negated the need to mark handwritten papers. There was a request to explore whether TestReach's system could deliver the CRQ in a way that would support the independence of questions, such as displaying each sub-question one by one.

The Group discussed the security aspect of online delivery and the fact that the College had previously had issues with the harvesting of question banks during the pen and paper exams, which was learned via information from exam candidates. It was felt that most candidates recognise the seriousness of academic dishonesty and seem to be more concerned about ensuring there was a level playing field. Overall, the Group felt that the 'live proctoring' in TestReach whereby invigilators are present for the duration of the exam, a small ratio of proctor to candidate (1:6), and locked browser delivery preventing screen recording meant that the risk to exam security was not significant. It was further noted that with the technology available, the online format was much more likely to reveal any academic dishonesty.

The Group noted that feedback is collected after each exam via a post-exam survey sent to all candidates. This survey information was also shared with the GMC during 2020-2021 as part of the reporting on the online delivery. However, the Group felt that feedback is often provided only when a candidate has had a negative experience and therefore may not be an accurate representation of the candidate experience overall of sitting an online exam. It is clear that whilst candidates do not wish to return to pen and paper exams, they also want a good online experience, and there is a balance to be struck between security and accessibility.

Clinical exam delivery

The Group yet again acknowledged the generally positive view of candidates to the online clinical exam delivery and that this was based largely on the same reasons given for the online written exams: a more conducive environment for performing well in an exam, reduced travel time and cost.

In terms of examiners, the Group spent time discussing the advantages and disadvantages of this form of delivery for clinical exams. Some disadvantages of this delivery mode were:

- the long days sitting in front of screens
- lack of group discussion and networking which takes place during a face-to-face exam delivery
- appropriate space to examine in the home environment/conflict with family life
- security of the exam
- connectivity issues leading to increased reviews
- reduced involvement in paper development.

The Group also discussed alternatives to delivery, such as hybrid (examiners at the College, candidates remote) or hubs of people in regional centres.

Exam security was discussed for the clinical delivery and the Group noted that since (at the time of writing) there is no ability to lock the browser on the platform, Practique, clinical exams could be recorded without examiners' knowledge. While this may not benefit the candidate sitting the exam and making the recording, and therefore may not be classed as academic dishonesty per se, it would allow the candidate to share exam material post-test.

The Group spent some time discussing the issue of scheduling in online clinical exams. This was linked in particular to the May 2021 sitting of the Primary OSCE/SOE in which examiner breaks were reduced as far as

possible in order to increase candidate capacity for what was an oversubscribed sitting. The scheduling in May resulted in very long days for everybody involved and this may have emphasised the negative aspects of this form of exam delivery. The Group discussed the potential to retain office hours for online exam delivery since early starts and late finishes conflicted with family activities, such as getting children ready for school.

Examiner leads in the Group reported that questions used in the online SOE are the same as those used in a face-to-face format, albeit with some tweaks. These minor changes were necessary since SOE questions have been written in such a way to test understanding of concepts and this was felt to be a more difficult construct to assess online when, for example, candidates use a drawing as part of their explanation. This change to the way the exams are being delivered may lead to examiners taking differing approaches and thus an inconsistent experience for the candidate. The point was also made that anything designed to go online should be purpose built for online delivery, but that because of the pandemic and the speed at which the exams had needed to be transformed, this had not been possible.

The Group reviewed the reports on the online delivery of OSCE and noted that some questions had been performing better than expected despite re-Angoffing after a rewrite for the online delivery. The Group wondered whether this was due to Angoffing questions as if the exam were still face-to-face rather than online. There is still some examiner judgement involved in this objective checklist delivery of an OSCE and therefore the Group considered whether examiners were either consciously or unconsciously applying a benefit of the doubt to compensate for the change in format and delivery.

The point was raised that with all the data collected, an area that needed further investigation was examiner performance in terms of inter- and intra-rater reliability. A member of the Group stated that some work had been undertaken in this area for the Primary MCQ in terms of Angoff. However, it was felt that this should only be pursued once changes to the exams had been made.

The Group reviewed the data analysis undertaken for the GMC regarding the delivery of online exams, which compared pass rates for online exams to previous in-person/pen and paper deliveries. The data suggested that pass rates and the reliability of the questions set were not changed significantly by the change in format, despite the change in the demographic, eg the increase in TEE candidates.

The Group did not feel that this data set helped in the assessment of validity and that the best way to review validity was through the link from the curriculum to the blueprint. The importance of looking at item-level metrics contributes towards the validity of the exam and should also be encompassed in any review. The Group felt that the integration of the written question banks into the system should allow for a more efficient review of metrics and paper performance.

The issue of data on differential attainment (DA) was raised. The Chair confirmed that work on DA will initially concentrate on the attainment gap of those with a primary medical qualification from the UK with protected characteristics and will be done when there is the psychometric capacity and the necessary data available (a data collection exercise is required). Whilst it was accepted that this data collection should initially be focused on UK trainees, the Chair stated that the College would like to understand if the attainment gap had changed as a result of the online delivery and the associated increase in overseas candidates sitting the exams.

Recommendation 15

To retain a remote, online delivery with live proctoring for written exams.

Recommendation 16

A return to face-to-face delivery for clinical exams.

Recommendation 17

To undertake a careful review of the link between the curriculum and blueprint as part of a validity study of the FRCA examination.

Recommendation 18

Once changes to the exams have been made, the College should investigate best practice for analysing inter- and intra-rater reliability for its exams.

Recommendation 19

Research to be undertaken into the attainment gap of those with protected characteristics. A data collection project will be first required to ensure sufficient data are available on which to base a meaningful conclusion.

9. Summary of Recommendations

Recommendation 1

The Review Group agreed that external input and challenge of the FRCA examination and its processes were positive undertakings and that examiners and the examination as a whole have benefited hugely from exposure to training programmes delivered by the Health Professional Assessment Consultancy (HPAC) Moving forward, the FRCA examination should ensure that there is externality included within the development of the examination in order that it continues its journey towards best practice in medical education and assessment.

Recommendation 2

The Review Group agreed to establish improved core examination documents, the examination syllabuses, blueprints, test specifications, to ensure that questions are appropriately set and derived from the curriculum. These documents will also help demonstrate and guide stakeholders to the appropriate level of depth and breadth that the examination tests.

Recommendation 3

The Review Group were in agreement that the current Primary and Final SBA and CRQ formats should continue. The CRQ is only used in the Final exam and this should not change. More consideration of Very Short Answer (VSA) questions was required and this will form part of the discussion around the final format of the MCQ examinations in 2023.

Recommendation 4

The OSCE should remain a fundamental clinical component in the Primary examination to assess performance in a simulated environment appropriate to the new stage 1 of training.

Recommendation 5

The kiosk stations should be removed as these do not represent a true OSCE assessment. The contents of these stations could be converted into SBA questions in the Primary MCQ examination and data interpretation questions in the OSCE.

Recommendation 6

The OSCE should be updated and reformed to reflect an authentic OSCE assessment as defined by the Academy of Medical Royal Colleges and the GMC, for example, a reduction in assessed stations and an increase in station duration that is in line with the constructs being assessed.

Recommendation 7

The OSCE should have a clearly defined set of areas of practice that it assesses and a separate test specification that supports the sampling of examination content against the curriculum.

Recommendation 8

The removal of the Primary SOE component from the FRCA and for the materials from the SOE to be converted into SBA questions and OSCE questions. The Primary MCQ may need to extend the range of SBA questions to ensure that the appropriate content continues to be assessed across the FRCA.

Recommendation 9

The Final SOE in the FRCA exams strategy to be retained as a valid mode to assess knowledge, understanding and clinical decision making.

Recommendation 10

A new and appropriate method of standard setting for the Final SOE examination should be investigated with a view to being developed, piloted and implemented in the exam.

Recommendation 11

A review of the standard setting methods in the OSCE and Final SOE components of the examination, with the view to making recommendations for the most appropriate standard setting methods for the respective examinations.

Recommendation 12

Further modelling should be undertaken on the performance of borderline candidates in the written exams and subsequent clinical exams.

Recommendation 13

The FRCA should move to a position where no or one SEM is added to the pass mark at the point of transitioning to an SBA paper in 2023

Recommendation 14

A group to be formed with the express aim of aligning standard setting methods and item analysis processes between the Primary and Final MCQ examinations.

Recommendation 15

To retain a remote, online delivery with live proctoring for written exams.

Recommendation 16

To recommend a return to face-to-face delivery for clinical exams.

Recommendation 17

To undertake a careful review of the link between the curriculum and blueprint as part of a validity study of the FRCA examination.

Recommendation 18

Once changes to the exams have been made, the College should investigate best practice for analysing inter- and intra-rater reliability for its particular exams.

Recommendation 19

Research to be undertaken into the attainment gap of those with a primary medical qualification in the UK with protected characteristics. A data collection project will be first required to ensure sufficient data is available on which to base a meaningful conclusion.

References

1. Excellence by design: standards for postgraduate curricula. General Medical Council, 2017.
2. Guidance for Standard Setting: A framework for high stakes postgraduate competency-based examinations. Academy of Medical Royal Colleges, 2015. [Accessed April 2022]
3. Boulet, J R. (2003). Setting defensible performance standards on OSCEs and standardised patient examinations. *Medical Teacher* 25 (3), 245–249.
4. Clyburn, P *et a*l. (2022). Association of Anaesthetists. Are we creating another 'lost tribe' of anaesthetic trainees? . [accessed April 2022].
5. Kane, M. (1994). Validating the Performance Standards Associated With Passing Scores. Review of Educational Research, 64, 425-461.
6. Norcini, J. J. (2003). Setting Standards on Educational Tests. *Medical Education*, 37, 464-469.
7. Pell, G. (2010). How to measure the quality of the OSCE: A review of metrics – AMEE guide number 49. *MedicalTeacher*,32, 802-811.
8. Subramaniam, J *et al*. (2022). Recruitment to higher specialty training in anaesthesia in the UK during the COVID-19 pandemic: a national survey. *Anaesthesia*. 2022 Jan 21. doi: 10.1111/anae.15660. Epub ahead of print. PMID: 35064578.

Bibilography

- Angoff, W. H. (1971). Scales, norms and equivalent scores. In: Thorndyke, R. L. (ed.) Educational Measurement. Washington DC: American Council on Education.
- Designing and maintaining postgraduate assessment programmes. General Medical Council, 2017. [Accessed January 2019]
- Miller, G. E. 1990. The assessment of clinical skills/competence/performance. *Academic Medicine*, 65, S63-S67.
- Polanyi, M. (1967). The Tacit Dimension, London, Routledge.
- Ricker, K. L. (2006). Setting Cut-Scores: A critical Review of the Angoff and Modified Angoff Methods. *The Alberta Journal of Educational Research*, 52, 53-64.
- A report on the welfare, morale and experiences of anaesthetists in training: the need to listen. Royal College of Anaesthetists. [Accessed April 2022].
- Van Der Vleuten, C.P. M. (1996). The Assessment of Professional Competence: Developments, Research and Practical Implications. Advances in Health Sciences Education, 1, 41-67.

Appendix A

FRCA Examination Review 2015 – Action Log

A review was conducted of the FRCA examination in 2014/15 under the Chairman of the Examinations Review Group, Dr Platon Razis.

In the subsequent report, a number of actions were identified. This log provides an update of the actions taken and work ongoing.

The log provides a resource for recording the work undertaken, which was identified in the 2014 review, and will enable additional tasks to be addressed as required. It will also provide a source of evidence for the planned subsequent FRCA Examination review 2020.

Serial	Exam	Action point	Responsible	Status	Target/Review Date	Comments/Action required
1.	Final and Primary	**2011 Review follow-up**; improve sound proofing dividing screens in examinations galleries.	Director Manager Facilities	Completed	To be considered along with refurbishment plans.	■ To be discussed further with Facilities/SMT.
2.	Final and Primary	**2011 Review follow-up**; consider moving the waiting area for candidates away from the College reception area.	Director Manager Facilities	Ongoing	To be considered along with refurbishment plans.	■ To be discussed further with Facilities/SMT.
3.	Primary	**2011 Review follow-up**; complete remaining Primary FRCA related e-LA modules.	Chairman Exams Committee and Primary Director Manager	Ongoing	Exams Committee February 2016.	■ Discussions with e-LA Leads as required. ■ Complete remaining FRCA related e-LA modules. ■ Improve liaison to ensure future update.

Serial	Exam	Action point	Responsible	Status	Target/Review Date	Comments/Action required
4.	Final	**Proposed change to the Final FRCA examinations**; timing of the exam relative to training – proposal that Final exam to be taken from ST3 to end of ST5	Chairman Exams Committee and Final Director Manager	Completed	2 to 3 years, Exams Committee to review annually in May each year.	This proposal will require GMC approval as follows- ■ Complete a curriculum change form. ■ Confirm support of lead dean. ■ Confirm support from NHS Employers if required. ■ Conduct an Equality Impact Analysis. ■ Consider the impact on other assessments within the training programme (WBA and Primary FRCA). ■ Consider issues related to moving closer to the CCT date. ■ Consider the impact to current trainees ,
5.	Final	**Proposed change to the Final FRCA examinations;** Final Written exam – The structure of the SAQ paper to change to Constructed Response (CR) questions	Chairman Exams Committee and Final Director Manager	Completed	2 to 3 years, Exams Committee to review annually in May each year.	■ Inter-examiner reliability will be assessed by double marking a SAQ exam paper. ■ College employ educational experts to assist, support and train examiners in CR question writing. ■ Move format to CBT. ■ Consider impact on trainees ■ Conduct an Equality Impact Analysis ■ Notice for change – 1 year ■ Approval for change to GMC

Serial	Exam	Action point	Responsible	Status	Target/Review Date	Comments/Action required
6.	Final	**Proposed changes to the Final FRCA examinations;** Final SOE structure to be changed to ensure increased number s of different examiner-candidate interactions, formal matching to cover all aspects of the curriculum and the merging of basic science with clinical practice.	Chairman Exam Committee and Final Director Manager	Completed	Within 2 years, Exams Committee to review May 2016.	■ To reduce the long case to two rather than three elements. ■ Introduce an additional table for all candidates increasing total examiner-candidate time. ■ Introduce linked basic science/ clinical questions ■ Consider impact on trainees ■ Conduct an Equality Impact Analysis ■ Notice for change – one year ■ Approval for change to GMC
7.	Final and Primary	**Standards for examiners;** introduce PS/JD for examiners, highlight shortfalls in numbers of examiner applicants	Chairman Exam Committee Director Manager	Completed	To be reviewed as required.	■ Create JD and PS for examiners ■ Ensure examiner standards meet the AoMRC/ GMC requirements ■ Highlight examiner applicants must; hold training post or actively involved in exam preparation/teaching. Visited exam within five years. ■ Continue E&D training on an annual basis
8.	Final and Primary	**Equality and Diversity;** GMC Standard 17 compliance.	Chairman Exam Committee Director Manager	Completed	To be reviewed as required.	■ Equal opportunity policy in place and in the public domain. ■ Learn from findings and statistical evidence ■ Ensure candidate rights are respected ■ E&D training a core competence of examiner training.

Serial	Exam	Action point	Responsible	Status	Target/Review Date	Comments/Action required
9.	Final and Primary	**Equality and Diversity;** differential attainment	Chairman Exam Committee Final Primary Director Manager Exams statistic department	Ongoing	Review annually.	■ Ensure database is of a sufficiently high standard to allow longitudinal studies of large enough numbers to allow future analysis. ■ Longitudinal study of differential pass rates must form part of future exam reviews. ■ College to consider if audit department needs to be expanded. ■ College should encourage completion of voluntary Equal opportunity monitoring forms.
10.	Final and Primary	**Equality and Diversity;** Examining Board diversity	Chairman Exam Committee Final Primary Director Manager	Ongoing	Review annually.	■ Actively encourage women and BME anaesthetists to apply to become examiners, ensure they understand what factors are considered during the selection process. ■ Encourage all examiners to complete their gender and ethnicity data. ■ Collect and publish data and use it to encourage greater participation from under-represented groups. ■ Ensure that helpful initiatives such as LTFT examining are adequately publicised. ■ Work actively with the PSED Compliance Group. ■ Lobby the Secretary of State for Health and the CEOs of NHS Foundation and other Trusts to look more favourably on releasing anaesthetists of all ethnic groups to serve as examiners.

Serial	Exam	Action point	Responsible	Status	Target/Review Date	Comments/Action required
11.	Final and Primary	**Statistical reporting and data capture;** supplying reports and data	Chairman Exam Committee Final Primary Director Manager	Completed	Chairman's reports to be submitted annually at September Exams Committee meeting.	■ Chairmen of Primary and Final to submit an annual report to the Examinations Committee at their September meeting and once agreed these reports should be placed on the exam pages of College website no later than 1 November each year. ■ Examinations dept to continue to produce stats reports on completion of each exam and place on examiner secure area. ■ Pass rates to be published on exam pages of College website. ■ Exams dept to continue to supply individual candidate data to the GMC.
12.	Final and Primary	**Lay involvement;** in the examination process and content	Chairman Exam Committee Final Primary Lay Committee Director Manager	Completed and ongoing activity	Review as required.	■ To liaise with the Chairmen of the relevant exams and the question-setting examiners to arrange for specific lay input into aspects which involve assessing the full range of communication skills of candidates. ■ To support this action by researching the results of current projects such as the simulation exercise currently in hand at St Bartholomew's Hospital. ■ To continue to observe exams, make comments on the observers' sheets, and abide by the guidance for Lay Visitors to Exams.

Serial	Exam	Action point	Responsible	Status	Target/Review Date	Comments/Action required
13.	Primary	**Technology and CBT;** increasing the use of e-marking in the Primary OSCE, prove e-marking in the SOE	Chairman Primary Examiner sub-group Director Manager	Completed	SOE e-marking – May 2016. Review kiosk in May 2016. Introduce E-marking for all OSCE stations within 2 years.	■ Assess technology. ■ Form examiner sub-group to work with developer and College TSR. ■ Prove E-marking in SOE; monitor for 12 months, if reliable cease manual marking. ■ Consider which OSCE stations can be converted to kiosk format. ■ Complete kiosk project then introduce E-marking to remaining stations.
14.	Final	**Technology and CBT;** prove e-marking in the Final SOE	Chairman Final Director Manager	Completed	June 2016.	■ Prove e-marking in SOE; monitor for 12 months, if reliable cease manual marking.
15.	Final and Primary written exams	**Technology and CBT**; to move Primary and Final written exams from paper and pencil to computer-based solutions	Chairman Exam Committee Final Primary Director Manager	Completed	Within three years.	■ Follow up scoping document with proposal setting out risks, costs (initial and ongoing with comparison to current costs) pros and cons, timescales, governance. ■ Seek approval of SMT, Council and finance committee. ■ Form examiner sub-group, with trainee and lay representative. ■ Appoint dedicated project manager/supervisor. ■ Consider impact on trainees. ■ Conduct an Equality Impact Analysis. ■ Notice for change – one year. ■ Approval for change to GMC. ■ Carry out local testing.

Serial	Exam	Action point	Responsible	Status	Target/Review Date	Comments/Action required
16.	Final and Primary	**Overseas examinations**; Continued recognition, QA and examiner assistance to overseas examining and questions.	Chairman Exam Committee Final Primary Director Manager	Ongoing	Review as required.	■ The MCAI will no longer be recognised as an exempting exam for the Primary FRCA as of April 2015. ■ Awaiting GMC's decision on timing of cessation of recognition of the FCAI for training in the UK. ■ Ensure clear communications plan and that no candidate is disadvantaged. ■ Continue to recognise agreed exemption qualifications as part eligibility towards the Final FRCA. ■ Once GMC position confirmed add statement to regulations. ■ QA of exempting examinations by sending FRCA examiners to visit. ■ Examining assistance will continue as long as it is considered to be mutually beneficial.
17.	Final and Primary	**Candidate feedback;** Standards and policy	Chairman Exam Committee Final Primary Director Manager	Completed	Policy to be agreed and added to August 2015 regulations.	■ Exams Committee to review AoMRC standards document on candidate feedback and ensure College meets requirements. ■ To agree a candidate feedback policy, add to exam regulations as appendix.

RAJB
DoTE
29 Nov 13

Appendix B

Royal College of Anaesthetists

Primary FRCA – OSCE Review

1. Introduction

This review has been conducted as part of the continuous examinations review process. The purpose of the review is to consider whether the primary FRCA OSCE fulfils the recent standards and requirements of the GMC (2017) and is aligned with contemporary educational theory and examinations practice. The review has focused on 1) the current OSCE content, format and delivery; and 2) standard setting and marking.

The OSCE exam is currently part of the oral component of the Primary FRCA. This exam assesses whether candidates have the knowledge, skills and behaviours required to progress from basic anaesthetic training to intermediate training. OSCEs have been developed as a method of assessing skills and behaviours and they are intended to assess the `shows how' level of Miller's pyramid. Furthermore, whilst, skills and behaviours can be assessed using WPBAs, there is no other format that can assess these attributes in an exam setting.

A number of options are proposed for consideration by the RCoA Examinations Committee.

OSCE Review Group

Damian Doyle (DD), Roger Sharpe (RS) and Mark Forrest (MF) were asked to undertake this review. All are experienced Primary FRCA examiners. DD was previously OSCE lead and is currently vice chair of the Primary FRCA. RS is current OSCE lead. MF is current chair of the RCoA Examinations Committee. Visits have been made to OSCEs held by other colleges. DD has visited the PA OSCE exam held by the RCP, the RCPCH OSCE exam and the Professional & Linguistics Assessments Board (PLAB) exam held by the GMC. RS has visited the RCS OSCE exam and MF has also visited the PLAB exam. All three members of the Review Group have in 2018 attended the Assessment Masterclass for Health Professional Educators, a two-day course focusing of current practices of standard setting and delivery of OSCE examinations. DD and MF have both previously attended the International Advanced Assessment Course, a two-day course on postgraduate examination assessment. DD and RS conducted a focus group discussion during the January 2019 exam to get feedback from OSCE working party members.

2. Background

The current Primary FRCA OSCE format

The OSCE consists of 17 - 18 consecutive stations (16 assessed plus 1-2 rest stations) of five minutes duration each with a one-minute break between stations.

The following areas (each led by a section lead) are examined:

1. Interactive resuscitation and simulated critical incident management.
2. History & communication (the new assessment system continues to be rolled out into the history and communication stations. This awards marks for qualitative aspects such as organisation and professional manner. This approach to assessment has strong support of our lay committee members).
3. X-rays & technical skills, (X-ray questions and some anatomy questions are computer-based interactive stations. Computer-based stations do not require a dedicated examiner).

4. Anatomy and physical examination.
5. Equipment, measurement and monitoring.

Currently, the exam is paper based. The long-term aim is to move across to an electronic platform, but the need for absolute reliability and to avoid significant delays has precluded this at present. This will continue to be reviewed by the exam board.

A candidate may score a maximum of 20 marks at each station, and the sum of the mark at every station produces their final score. This is compared against a target score created by use of Limen referencing based on the Angoff score (pass mark) of each of the individual stations.

Each day the results are analysed to ensure consistency of the process. In particular, candidates who score one mark under the pass mark have their performance reviewed.

Questions and answers in each station are developed on an ongoing basis by the OSCE Working Party (WP), composed of up to 18 consultant examiners. Standard setting is performed by the OSCE WP using a modified Angoff process. Each station will therefore have a fixed individual pass mark. The sum of sixteen Angoff marks represents the pass mark for each OSCE round. The matrix for the exam week is set up such that each round should have a similar overall pass mark, ensuing that each round is of a similar overall difficulty.

Post-exam data analysis is provided to the OSCE WP by Dr Andrew Bowhay (a previous examiner). Cronbach's Alpha, Discriminatory Index, Item Difficulty and Point Biserial are calculated to assess question performance. Poorly performing questions are either archived, amended or changed as required.

Best practice and concerns regarding the current OSCE format

The list below summarises the current best practice in OSCEs (HPAC 2018):

1. use OSCEs to test clinical and communication skills
2. blueprint OSCE to curriculum outcomes
3. sample sufficiently.
4. construct authentic stations (content, reward synthesis and decision making, appropriate length of station, appropriate SP scripts)
5. use appropriate standard setting - borderline group/regression method
6. appropriate scoring – checklists vs rating scales
7. ensure training of examiners and simulated patients
8. carry out QA - range of metrics to evaluate OSCEs.

The following factors have been identified as problematic areas for OSCEs (HPAC):

- too few stations (less than 12) – poor reliability
- poorly blueprinted OSCE – insufficient spread of skills being tested - poor reliability and validity
- MCQs/ SAQs/Oral exams disguised as OSCEs – poor validity.

These factors are considered in more detail below.

1. **Use OSCEs to test clinical and communication skills** – In the current format, many of the stations, eg anatomy, monitoring and measurement, the computer-based stations, etc. are effectively a knowledge assessment using a structured series of questions in the form of a checklist. Arguably, many of these questions are mini-structured SOEs which affects the validity of the exam. Knowledge is better assessed in other ways, i.e. MCQs or other written assessments.
2. **Blueprint OSCE to curriculum outcomes** – The current blueprint may require review if the structure of the OSCE is revised and knowledge-based questions are moved out of the current OSCE format (see below). In addition, the current curriculum review may require further adjustment.

3. **Sample sufficiently** – The current number of stations (16) is sufficient. The number of stations determines the total number of candidates examined per week.
4. **Construct authentic stations** – The computer stations generate recurring and frequent incidents, e.g. computer failure, server disconnection, candidate allocation error. These issues, together with the concern that the computer-based stations are simply knowledge assessment stations, have caused some examiners to question whether they should remain as stations within the OSCE. Currently it is not easy to examine some of the material covered within kiosk stations within other formats, e.g. MCQ. However, if the College were to move to computer-based testing this should not be the case.
5. **Use appropriate standard setting** – *Borderline group/regression method* – The Angoff method is currently used. HPAC supports the use of BLR as the gold standard in OSCEs. In 2017 Richard Wakeford, a standard setting expert, reviewed the RCoA OSCE examination. He was concerned about our use of a 'modified Angoff' process, the use of computer-based stations and supported the introduction of BLR. A trial of BLR over three sequential exams was undertaken (see below). The trials suggested that BLR was `useable' as a standard setting format.
6. **Appropriate scoring** – *checklists vs rating scales* – In the BLR pilot, the results / cut scores generated showed significant inconsistencies particularly in the less interactive stations. This highlighted concerns regarding the use of the OSCE as an assessment of knowledge rather than competency.
7. **Ensure training of examiners and simulated patients** – All examiners are required to attend a day's introductory training plus additional training in the application of simulations for assessment. The actors used in the OSCE do not receive specific training, but the regular use of the same agency ensures experienced actors. Further training will be required if the method of standard setting is changed
8. **Carry out QA - range of metrics to evaluate OSCEs** – The OSCE questions are reviewed using examiners' and visitors' feedback and post-exam analysis using a number of metrics including item difficulty, difficulty index, point biserial, Cronbach's alpha.

Other challenges of the current RCoA OSCE

Fixed and restricted location to run the exam
The footprint of the OSCE floor (and SOE) is relatively small. The floors are expected to be used flexibly throughout the year. Room dividers are used to divide stations. They do not provide adequate sound protection, causing high noise levels and distraction for candidates, eg noise protectors are now offered for candidates in the computer-based stations (although complaints from candidates are not frequent).

High number of candidates
The January 2019 exam week for example is expected to accommodate 341 candidates through the OSCE. This equates to 68 candidates per day, 17 candidates per OSCE round based on a four-round day. These numbers are fairly typical of candidate numbers. The numbers taking part in the November 2018 exam required one additional evening OSCE round.

Limited number of primary examiners
Each exam sitting requires approximately 56 examiners. At present, there are 82 primary examiners. Despite nine examiners expected to take part in all three sittings of the exam next year, potential examiner shortages have been identified. It is uncertain what the impact of flexible or less than full-time examinership will be. Examiners represent a significant resource implication for the College, both human and financial.

Fixed number of exam sittings
The Primary FRCA exam in its current format is delivered as three sittings per year. Any changes are restricted by examiner availability, facilities and the demand that the exam schedule be linked with key points of trainee progression.

Uncertain future IT planning

A review of the IT facilities and resources is ongoing within the College. It is not clear to this Review Group what potential impact this will have on future IT resources available to the exam. Ongoing reliability of the unmanned computer stations are a significant concern. Recent marking errors have highlighted the problems of manual collation and transfer of marks and results between the different data systems used by the exam. The SOE presently uses both a paper based and an iPAD platform. The OSCE is for the most part based on manual checklist marking by examiners and the use of OMR sheets

3. Relevant literature and external standards

The most recent standards on postgraduate assessment programmes were published by the GMC in May 2017. The document defines five key principles.

- Safety – assessments assure the profession, patients and the public that doctors are safe.
- The maintenance of professional standards.
- Excellence – enables learners to develop the skills, knowledge and performance for excellent patient care.
- Fairness – affords all learners opportunities to demonstrate outcomes and considers their performance consistently in line with clear and transparent criteria.
- Meeting patient and population needs – current and future.

The document highlights a greater emphasis on validity (defined as: "interpretations and uses of tests that make sense and are supported by appropriate evidence" [Kane 2013]).[5]

Van der Vluten (1996) defined five aspects that are required to ensure an effective assessment methodology. AoRMC (2015) stated that to apply these to standard setting requires us to ask the following questions:

1. Validity: does the chosen standard setting method allow us to categorise candidates meaningfully in line with the purpose of the examination?
2. Reliability: does the chosen standard setting method give us confidence that a candidate's pass/ fail outcome would be the same, regardless of which day they sat the examination?
3. Educational impact: does the documentation available to candidates about the level required for a pass allow candidates to prepare well for the examination?
4. Feasibility: is the chosen standard setting method appropriate within the constraints of our examination (number of candidates, availability of standard setters)?
5. Acceptability: is there sufficient transparency around how the pass/fail decisions are made to satisfy the candidates, the profession, the governing bodies and the public that the process is fair?

In addition, AoMRC (2015) added that due consideration must be given to equality and diversity issues covered by the Equality Act 2010. These factors need to be considered in any review of the OSCE format.

Setting the pass mark

The pass/fail decision based around the pass mark sets the standard of the examination. The borderline candidate should sit at this point on the continuum of examination scores. It is noteworthy that the standard set is based on the judgement of selected examiners and as such is a social construct. Norcini (2003)[6] describes several steps required to arrive at the pass mark: a) the type of standard; b) the method for setting it; c) who sets the standard; d) how the standard will be converted into a pass mark; and e) the review process that follows afterwards.

a. The type of standard

In the FRCA, criterion referenced standards are used in which candidates are assessed against performance to set criteria. This is an appropriate method for professional exams and does not need review.

b. Methods of standard setting

Over 38 methods for setting the standard/pass mark have been described (Berk, 1986). In the context of this OSCE review the current method (the Angoff process) and the perceived gold standard (borderline regression) will be considered in more detail. Both methods use the construct of the borderline candidate to determine the pass mark.

The Postgraduate Medical Education Training Board (PMETB) suggested that the particular approach chosen for standard setting may not be as critical as the fidelity and care with which it is conducted, and defined three main requirements in the choice of method (PMETB, 2007, p.13):

Defensible to the extent it can assure the stakeholders about its validity

Explicable through the rationale behind decisions made

Stable as it is not defensible that standards vary over time.

c. Who sets the standard?

The OSCE WP examiners set the standard. A clear definition of the target examination group is required (Boulet 2003)[3], which also defines the purpose of the exam, ie the level of knowledge and skills required for a core trainee to progress to specialty training.

d. How are the standard converted into a pass mark?

The pass mark is a numerical point on the score scale that defines the boundary point between acceptable and unacceptable levels of knowledge and competency (discussed below).

e. The review process

After the examination it is necessary to review the standard to ensure reasonable results are produced; these requirements are defined in the GMC document. Kane (1994) also recommends that examiners should be exposed to the consequences of setting pass marks. Arguably this happens, as all examiners are required to be clinically active and therefore work in departments with core trainees passing through the exam.

Angoff versus Borderline Regression for standard setting

1. Angoff

In the Angoff method (1971), examiners visualise a borderline candidate and estimate the probability of this candidate answering individual test items of the examination correctly. The overall pass mark is then derived from the sum of the Angoff scores for each component of the examination (In the context of the OSCE – the sum of the stations). This is a 'test-centred' method where the pass mark is determined prior to the candidates sitting the examination and the resulting pass mark is considered just adequate. This method is currently used in setting the OSCE pass mark.

Advantages

- A modified Angoff method has been used to set the OSCE pass mark for the past 15+ years and the examiners are familiar with this technique.
- The Angoff method allows the pass mark to be set ahead of the exam allowing earlier publication of exam results.
- The reliability of the current question bank is based on historical Angoff scores.
- The Angoff for each question is set using the mean score derived from all members of the OSCE WP, giving a wide input into each question.
- The Angoff method can be used for all the question formats in the current OSCE format.

Disadvantages

- The Angoff process is laborious and time consuming for the OSCE WP members.
- Boulet (2003) argues that checklist items are interrelated and therefore raters' judgements are not totally independent, reducing the validity of using the Angoff process for checklist items.

2. Borderline regression

This is an examinee-centred test and evaluates the candidates performance/scores during the exam. A regression line is constructed by plotting the global performance rating against the checklist items to derive the pass mark after the examination. The BLR method is described in more detail in supporting documents [A].

Advantages

- Utilises the expertise of the examiners during direct observation of the performance of the candidates.
- Examiners are in a position to make a (global) judgement about the performance based on:
  - their clinical expertise
  - expected standards for the level of the test
  - knowledge of the curriculum/teaching.
- Credible and defensible: based on expert judgment in direct observation.
- Reliable: cut-off score based on a large sample of judgements.

Disadvantages

- Passing score not known in advance.
- Global judgements may not be fully independent of checklist scoring (although there is some evidence that this may be mitigated by using domain-based checklists and weighting).
- Requires expert processing of marks immediately after the exam:
  - checking of results
  - delay in producing results.
- The cut score for each station will be based on the judgement of a single examiner for each round.

In summary

The key themes of this brief review highlight that there are many ways of determining an examination pass mark. Despite much research, there is no firm opinion to suggest which method should be used and when. However, there is a growing consensus (HPAC, 2018; Boulet, 2003) that the borderline regression model is optimal for OSCE standard setting. However, the Angoff process is also acceptable (personal communication, HPAC). Many methods employ the concept of the borderline candidate to judge where to set the pass/fail point. It is acknowledged that setting the pass mark involves a judgement process.

4. Observation of other examinations

Full descriptions of the visits are included in the supporting documents [B-D]. Common themes were observed:

- Ten-minute OSCE stations seemed to work well
- Check-list combined with BLR marking was commonly used
- Knowledge-based questions are used in the OSCE of other colleges
- Dividing each station into an initial `doing / task part' and a later `question' part would allow an examiner to assess a candidate's performance and knowledge, using a varied level of domains.

5. OSCE working party focus group

A focus group was held during the January 2019 exam. This was facilitated by DD & RS and attended by nine members of the OSCE WP group and David Rowand from the exams team. Common themes that emerged included:

- Consideration of disconnecting the OSCE from the SOE
- Longer duration stations (10 - 15 mins) was widely supported
- A longer exam week (e.g. Mon – Sat) was acceptable (if mitigated by flexible working).

6. Borderline regression pilot

A pilot of Borderline Regression (BLR) marking was conducted for the three exam weeks during the 2017-18 academic year. The BLR method was conducted in parallel to the standard Angoff method and checklist marking. The standard method was used to set the pass mark and for marking, the BLR marks did not contribute to the results. The full reports are available in the supporting documents [E-F].

The pilot included a trial of a four-point scale (November 2017 and January 2018): 1(clear fail), 2 (fail), 3 (pass), 4 (pass). A five-point global rating scale was trialed in May 18: 1(clear fail), 2 (fail), 3 (Borderline pass), 4 (pass), 5(clear pass). For all three trials, global scoring was not applied to stations beginning with X, A3 and R as these are unmanned/kiosk station and therefore in the absence of an examiner to apply a global score, cut scores were set using the historical average pass mark for each of these stations.

The pilot of the BLR analysis indicated that a five-point global score with a cut point of borderline pass (3) plus one SEM produces a pass mark (and therefore pass rate) that is most similar to the Angoff process.

Other issues

- The global score was omitted in 1.6%, 1.1% and 2.1% of the marking sheets for the three pilots
- In some candidates, there were large discrepancies between the checklist score and the BLR global score – for example some candidates scored 18-20 on the checklist but were scored a global fail in BLR (and vice versa)
- A different set of quality metrics are used to determine question/examiner reliability.

The pilot indicated that examiner training is important to address the above issues.

7. Options appraisal

General

Ensure that all stations are appropriate to OSCE format. The OSCE is a valid test to assess technical, non-technical skills and behaviours. It is currently the only component within Primary and Final to do that. This should include the continued (and greater) use of history and communication stations, interactive resuscitation and simulation stations. Technical skills, eg siting of epidural, central line, arterial line, correct prescribing, etc. could be included as could clinical examination and airway assessment. There is considerable scope for the development of patient-based stations focusing on pre op assessment skills and practice.

Stations which currently focus on knowledge alone should be removed from the OSCE. This is likely to include a significant proportion of areas covered by anatomy, monitoring, measurement, equipment and hazards. These `knowledge'-based areas should be assessed by means of tests designed to assess knowledge. The different OSCE WP subgroups should review their own assessment areas and identify which topics are suitable for an OSCE, which should be assessed by means of SBA / SAQ.

Use of BLR for standard setting. The changes above would facilitate the effective introduction and use of BLR as a means of standard setting. A simplified (but station-specific) domain marking should be used in combination with a global rating mark

Electronic marking should be developed for the OSCE. This would mitigate against marking errors. Feedback could be integrated within electronic marking platform, e.g. a candidate scoring below a threshold on the rating scale should trigger a requirement for feedback by examiner (as per PLAB).

Logistical options. Note: Apart from Option 1, all options will require revision of OSCE station format (as described above) to focus on assessment of communication and clinical skills. This will allow the introduction of BLR with domain marking and global rating. It will require revision of questions banks. Knowledge-based questions would need to be removed and assessed in other parts of the examination.

Option 1: no change

Effectively continue with current format of OSCE – ie maintain current OSCE structure with 16 assessed stations plus up to two rest stations. Each station five minutes plus one minute gap. Total time 108 minutes per round.

Advantages:

- allows current OSCE / SOE relationship to continue, i.e. maintain the current candidate numbers and to run OSCEs and SOEs in parallel
- the major advantage of this option is that it is the least onerous in terms of resources (human and financial)
- no question re-writing required
- historical performance data available.

Disadvantages:

- BLR cannot be used for standard setting in some questions. The result would be a mixture of standard setting within the same OSCE round
- does not address recent concerns, i.e. insufficient emphasis on assessment of clinical skills and overemphasis on knowledge-based assessment
- may not be compliant with 2017 GMC standards
- does not 'future proof' the OSCE.

Option 2: minimal change

Maintain 16 stations (five mins plus one minute gap), but change the format of each station to allow the use of BLR and remove the knowledge-based areas of questioning.

Advantages:

- maintains current OSCE / SOE relationship, ie maintain the current candidate numbers and run OSCEs and SOEs in parallel
- partially addresses concerns around GMC standards
- simplest option for change, less need for resources, ie question re-writing.

Disadvantages:

- longer stations seem more appropriate for assessment of clinical skills in an OSCE setting
- unlikely to 'future proof' the OSCE.

Option 3: Moderately longer stations

This option is to make all stations moderately longer, but to maintain (as near as possible) the current timeframe of 108 minutes per round. It is suggested that a minimum of 14 stations is required to ensure validity of an OSCE round. For example 14 x 7 min plus 1 min gaps = 112 minutes total.

Advantages:

- this allows longer stations which are thought to be more valid
- longer stations allow improved assessment of candidate.

Disadvantages:

- seven minutes could be considered too long for some questions (e.g. equipment/technical based OSCEs) and not long enough for others, such as clinical performance
- reduction in numbers of stations will result in reduced candidate capacity. Current capacity is 320-360 candidates per week depending on the number of rest stations (16-18 per round x 4 rounds/day x 5 days). Reducing the round to 14 stations equates to a maximum capacity of 280 candidates per week. There are a number of options for mitigating this moderate reduction in capacity:
  - an extra (5th) round per day (= total 350). This is unlikely to be popular (examiners, candidates, exam staff)
  - an extra (5th) round per day excluding the evening of the dinner. (= total 336)
  - four rounds per day Mon – Sat (n=336). Maybe appealing if people allowed to work flexibly through week, e.g. Mon - Fri or Tues – Sat
  - long days/Saturdays would require increased examiner numbers, examiner engagement and cost. This could be facilitated by flexible examiner working week, e.g. Monday / Friday vs Tuesday / Saturday.

Any further reduction in station numbers to achieve longer stations results in greater pro rata impact on capacity due to increased time required for each round.

Option 4: hybrid model (in series)

Mixing longer and shorter stations within same round, eg eight short stations (five minute station and one minute gap) with six longer stations (nine minute station and one minute gap).

Total stations 14 (which meet minimum agreed valid number for OSCE validity) and total OSCE round duration kept at 108 minutes.

Advantages:

- maintains current OSCE / SOE relationship and candidate capacity
- would allow a mixture of longer stations for assessing clinical competencies and shorter stations for technical skills.

Disadvantages:

- extremely complex logistics required
- rest stations and /or gaps are required if running long and short stations in series, which would impact on capacity
- would potentially require two sets of long stations (run in parallel) and one set of shorter stations to maintain candidate numbers. This would require: 1) an increase in the number of examiners; and 2) a bigger footprint of floor space.

Option 5: hybrid model (in parallel)

This would involve separating the OSCE round to two separate 'floors' running simultaneously in parallel.

Floor A: 8-10 sequential stations, five minute stations with one minute gap. Total round duration 48-60 minutes. (Potentially could be run in the first-floor briefing area).

Floor B: Composed of two parallel rounds of five to six identical stations, e.g. B1 and B2. Each round has five to six stations, ten minute OSCEs plus one minute gap. Group of five to six candidates split into two equal groups. Each candidate completes the five to six stations. Total round duration 55-66 minutes. This could be reduced to 55-60 minutes if gaps were removed.

Floor A and Floor B would rotate – total time in stations 120 minutes (plus time needed for rotation).

This would provide the capacity to examine 20 candidates per round (Floor A and B rotating).

Four rounds cold be run per day, providing a total capacity of 400 candidates per week.

Advantages:

- could potentially increase OSCE capacity depending on the number of rounds/days per week
- facilitates the introduction of longer OSCE stations
- questions could be allocated to either five mins (such as technical/equipment) or ten mins (clinical, H&C)
- allows a short break in the middle of the OSCE round (some examiners have commented that two hours OSCE rounds are too long with no breaks).

Disadvantages:

- needs additional floor space
- would requires long working days if four rounds were held per day
- moderately complex logistics
- challenge for floor managers
- requires increased numbers of examiners to maintain capacity

- calibration between identical parallel OSCEs critical
- each OSCE round extended to 120 minutes – potential difficulty of maintaining OSCE / SOE relationship. (This could be mitigated by 'disconnecting' the requirement to sit the SOE/OSCE on the same day).

Option 6: a complete rethink

Separate SOE from OSCE (days/weeks)

Take clinical out of SOE2 and use materials from SOE to develop performance based clinical OSCE scenarios

Largely examine clinical/technical skills in OSCE format

12 - 14 x 10-15 minute stations.

Advantages:

- would allow a radical review to allow in-depth clinical and technical skills assessment
- candidates are increasingly aiming revision for one part of the SOE/OSCE day – therefore could be acceptable to candidates.

Disadvantages:

- would need some redesign of the SOE
- would probably double the number of days required for OSCEs
- costly – examiner/exam team college time
- extensive resource required to restructure the OSCE and prepare new material.

Transition phase

Once the Examination Committee has decided on an option (with the exception of option 1) a phase of new question development will be required. Other areas to be considered are:

- question development time resources
- pilot a single 10-minute station in OSCEs next academic year
- examiner training
- communication to candidates.

8. Summary/conclusion

The purpose of this review was to consider whether the FRCA OSCE in its current format is aligned with best contemporary practice. We have conducted wide discussion and consultation within the examiner body, attended an assessment masterclass run by HPAC, and observed OSCEs run by other colleges. We conclude that although the current OSCE examination is largely fit for purpose there is an opportunity, and more importantly, a need, to change the format of the OSCE. This would address concerns that there is insufficient assessment of clinical skills and too much knowledge-based question material. It would also allow the introduction of BLR marking which is currently regarded as the optimum method of standard setting for the OSCE examination format. A number of options are presented for discussion. Some of the options would involve changes to other parts of the exam (SOE/MCQ).

Dr Damian Doyle
Dr Roger Sharpe
Dr Mark Forrest

January 2019

9. References

1. Excellence by design: standards for postgraduate curricula. General Medical Council, 2017.
2. Guidance for Standard Setting: A framework for high stakes postgraduate competency-based examinations. Academy of Medical Royal Colleges, 2015. [Accessed April 2022]
3. Boulet, J R. (2003). Setting defensible performance standards on OSCEs and standardised patient examinations. *Medical Teacher* 25 (3), 245–249.
4. Clyburn, P *et a*l. (2022). Association of Anaesthetists. Are we creating another 'lost tribe' of anaesthetic trainees? . [accessed April 2022].
5. Kane, M. (1994). Validating the Performance Standards Associated With Passing Scores. Review of Educational Research, 64, 425-461.
6. Norcini, J. J. (2003). Setting Standards on Educational Tests. *Medical Education*, 37, 464-469.
7. Pell, G. (2010). How to measure the quality of the OSCE: A review of metrics – AMEE guide number 49. *MedicalTeacher*,32, 802-811.
8. Subramaniam, J *et al*. (2022). Recruitment to higher specialty training in anaesthesia in the UK during the COVID-19 pandemic: a national survey. *Anaesthesia*. 2022 Jan 21. doi: 10.1111/anae.15660. Epub ahead of print. PMID: 35064578.

Bibilography

- Angoff, W. H. (1971). Scales, norms and equivalent scores. In: Thorndyke, R. L. (ed.) Educational Measurement. Washington DC: American Council on Education.
- Designing and maintaining postgraduate assessment programmes. General Medical Council, 2017. [Accessed January 2019]
- Miller, G. E. 1990. The assessment of clinical skills/competence/performance. *Academic Medicine*, 65, S63-S67.
- Polanyi, M. (1967). The Tacit Dimension, London, Routledge.
- Ricker, K. L. (2006). Setting Cut-Scores: A critical Review of the Angoff and Modified Angoff Methods. *The Alberta Journal of Educational Research*, 52, 53-64.
- A report on the welfare, morale and experiences of anaesthetists in training: the need to listen. Royal College of Anaesthetists. [Accessed April 2022].
- Van Der Vleuten, C.P. M. (1996). The Assessment of Professional Competence: Developments, Research and Practical Implications. Advances in Health Sciences Education, 1, 41-67.

10. Supporting documents

A	Borderline Regression Summary
B	Observation of RCS OSCE
C	Observation of RCPCH OSCE
D	Observation of GMC Professional and Linguistics Assessments Board (PLAB) OSCE

Appendix C

Royal College of Anaesthetists

Primary FRCA Clinical SOE review

1. Background

The clinical SOE currently consists of three questions of five-minute duration, based upon a clinical scenario that the candidates are given prior to the examination. The questions are produced by the clinical core group and are designed to assess the knowledge and non-technical skills of clinical anaesthesia, conduct of anaesthesia and critical incidents.

We have been asked to review:

- the purpose of the assessment (clinical SOE) and its role as a whole
- whether it tests what it is supposed to do and/or are there other assessments that might be used
- if there are any changes in the current process that could improve the validity of the assessment (validity in this context is: does the clinical SOE test measure what it intended to measure?).

2. The purpose of the clinical SOE

This can be defined as the assessment of a candidate's knowledge and non-technical skills of perioperative assessment, conduct of anaesthesia, recognition and management of critical incidents. It is also intended to identify areas of dangerous clinical practice and aspects of patient safety.

3. Whether it tests what it is supposed to do and /or are there other assessments that might be used

The following are considered valid areas of assessment **within the Primary exam** and are assessed in the SOE. However, the SOE format is perhaps not best suited to continue to assess the areas listed below:

Non-technical skills:

The test of understanding and communication of understanding to a fellow clinician – the initial section of the SOE requires comprehension of the significance of, and communication of aspects of, a patient vignette to the examiners. This requires prioritisation of the issues, decision-making processing and synthesis of a plan, with challenge by a more knowledgeable other.

Understanding of the impact of pathophysiology on anaesthetic practice:

There is a test of deeper understanding of clinical anaesthesia. Pathophysiological considerations must be married with best practice information.

However, the clinical SOE format fails to assess these areas reliably for the following reasons.

- The examiner guidance and candidate prompts are too general. There is insufficient guidance about the expected 'correct answers', areas of controversy for exploration or indeed the range of subject matter to be covered – which is therefore variable. This very likely effects the reliability of the questions.

- Given the marking guidance is very limited, and weighting is non-existent, the 2/1/0 scoring is too narrow to allow for reliable and adequate candidate discrimination (see section on validity).
- There are often not binary 'correct' or 'wrong' answers in the clinical SOE as there is at times an inadequate evidence base. Some areas examined in the SOE could be considered a matter of opinion; most answers would be based on textbook response or the accepted practice at the time, which is variable. This is probably a fault in the way the questions are set or phrased to fit the format of this free-flowing style of assessment. We feel this reliability issue could be improved (see below), but believe the underpinning clinical subject matter should be and can be an important part of the examination to test the candidate's knowledge, skills and behaviours compatible to moving from basic to intermediate training. Deeper understanding is demonstrated by the ability to weigh the merits of several clinical options. It seems feasible that with more adequate examiner guidance and content structure, clinical decision-making skills could be assessed in the FRCA.
- The SOE doesn't really pick up dangerous practice but more encourages the recital of memorised techniques that are perceived to be uncontentious. Candidates respond with 'stock' answers' (eg 'I would do a full history and examination'). A negative educational impact results from assessments that reward learning without exploring the depth and breadth of knowledge – this is a problem with the clinical SOE.
- There is candidate and examiner bias. A candidate's response will be determined by their clinical experience and training to date and their origin of training (eg the candidate may not have done their obstetric module, they may be from Asia and deem their anaesthetic technique more acceptable than in the UK), whereas examiner bias will be their perception of the correct answer based on their own experiences and local / regional practice (e.g. an intensive care consultant asking an obstetric question may be less astute than an obstetric one).
- Standard setting is an issue. There is no 'benchmarking' akin to Angoff scoring of the questions – indeed the database is produced by a small number of primary examiners in the clinical core group who determine the acceptable responses, albeit linked to the curriculum, latest guideline publications, and feedback from fellow examiners after the question has been asked in the exam.
- In the current format *some aspects of* the clinical SOE could be examined in the OSCE part of the primary examination. This would have cost and time implications in more actors would be needed and more simulation stations required.

4. Validity

The validity of the current format of the clinical SOE must be questioned.

- Some of the questions lag behind current accepted practice and guidelines.
- There is a degree of repetition across the database. Although this occurs in other sections of the SOE, there are perhaps limited clinical scenarios to move away from the 'stock answers' problem (above).
- The questions are not focused enough and allow too much candidate leeway in their responses (i.e. not specific), compared to other aspects of the SOE. Candidates can give vague answers, which while not containing incorrect information, amount to a verbal avoidance strategy and lack of commitment to any definitive answer. It is hard then to be clear about what they would do in practice. Adjudication of answers is vulnerable to variable interpretation by each examiner, based on impressions (and personal experiential bias discussed above) rather than facts (this is a mixed reliability and validity problem).
- The clinical SOE is not discriminatory enough and may allow a borderline candidate to pass the examination overall. A candidate could score an 11 or 12 and scrape through to the 37-pass mark when their performance in the other aspects of the SOE has been borderline. However, this does open the debate as to whether different aspect of the SOE should be weighted more than others. An example – should the understanding of Regnault's hygrometer be as important as safe competent anaesthesia in a six-year old?

The clinical SOE could be improved by asking focused questions on perioperative medicine and clinical anaesthesia (for example: what are the current DAS guidelines with difficult / failed intubation; what is the current consensus statement for the management of anaemia in elective surgery; principles of consent; how do you manage massive haemorrhage in obstetrics, etc). Focusing on more tangible sources of expert consensus and evidence from the literature may eliminate some of the current problems highlighted in the above sections.

5. Summary

With a new curriculum due, a focus on perioperative medicine and the FRCA examination review, there is an ideal opportunity to restructure the clinical SOE with a more focused evidence-based questions and moving some current aspects into the OSCE examination. A thought should be given to how the SOE is scored.

Dr Simon Vaughan – Lead for Clinical SOE

Royal College of Anaesthetists
Churchill House, 35 Red Lion Square, London WC1R 4SG
020 7092 1500 | exams@rcoa.ac.uk | rcoa.ac.uk
@RCoANews
RoyalCollegeofAnaesthetists

Information correct as at January 2023