ACADEMY OF
MEDICAL ROYAL
COLLEGES

Guidance for standard setting: A framework for high stakes postgraduate competency-based examinations

October 2015

01
Introduction

Postgraduate medical examinations have developed over many years across the specialties as a means of assuring that clinicians have an appropriate minimum level of knowledge and skill before progressing to the next stage of their career. The Academy and the General Medical Council (GMC) agree that the interests of the profession and of exam candidates coincide, and as such, efforts should be made to ensure that the processes underpinning pass/fail decisions in postgraduate medical examinations are conducted to an agreed standard across all specialties.

This guidance explores a set of guiding principles for standard setting in the context of summative postgraduate medical competency-based examinations. The aim of examinations within this context is to separate the competent candidates from the not competent candidates, according to how they compare to a minimum standard. This guidance therefore does not explore norm-referenced (or relative) standards.

Different examination formats and contexts lend themselves to different standard setting. This is particularly true in medical examinations where different methods will be needed for tests of knowledge and clinical assessments. In setting down these standards, it is understood that different colleges and faculties will have adopted different approaches to designing and delivering their examinations. It would therefore neither be possible, nor desirable to be prescriptive about which methods should be used and how they should be applied within each context. Instead, this guidance relates to standard setting, which can be applied regardless of the specific method or context.

The aim of providing these guidelines, based on key texts in the field of educational measurement, is to help ensure fairness, consistency and meaning in high stakes pass/fail decisions, in the interests of both trainee doctors and the public. There is no desire in doing so to inhibit excellence, or development and innovation which are the hallmark of postgraduate medical education and assessment in the UK. Irrespective of the standard-setting process adopted, the standard itself, a description of how it was derived and guidance as to how it should be interpreted must be clearly articulated and visible for both candidates and other stakeholders who might have a legitimate interest in the process of the examination.

The complexity of the standard setting process and the diversity of examination types and circumstances dictate that advice should be routinely sought from psychometricians who have experience and expertise in addressing these problems. The terminology for standard setting is particular to the subject, and therefore a glossary is included.

02
Have a clear rationale for the chosen method

Standard setting methods include the four following categories:

1. **Test-centred methods** where standard setters estimate how a group of imaginary *minimally competent* candidates would perform on each item in the test (Angoff, Ebel, Nedelsky, Jaeger)

2. **Examinee-centred methods** where the pass mark is determined by using the scores of candidates who were judged as belonging to different competence level categories either side of the minimal level of competence required for a pass (Borderline group, Contrasting groups, Borderline regression)

3. **Compromise methods** which combine aspects of norm-referenced and criterion-referenced methods (Hofstee)

4. **Statistical methods** which use the performance of candidates on *anchor* items to compute the pass mark for the examination (Item response theory, Linear equating, Raschmodeling).

Considering the multiplicity of methods available for standard setting, a clear rationale for why a particular method was chosen is an essential part of the defensibility of the pass mark,[1] and should therefore be clearly documented.[2] In high stakes examinations, the appropriateness of the method with regards to the examination's purpose, format, and practicalities (number of candidates, availability of standard setters, number of items in the examination, and so on) should be supported by research.[3,4,2]

Once an appropriate method has been chosen, consideration must be given to how candidates' marks on different items in the examination are used to reach a pass/fail decision. Examiners tend to intuitively favour partially compensatory standards, where candidates must both achieve a total score above the pass mark and meet additional criteria such as passing a minimum number of stations or never receiving a *clear fail*. However, due to the measurement error associated with individual items and stations, assessment specialists tend to favour fully compensatory standards, where candidates who achieve a total score above the pass mark are awarded a *pass* regardless of how they performed on individual items (see Standard setting in student assessment for a discussion on this topic).

03
Select and train standard setters who are best placed to determine the minimum level of competence required for a pass

Where appropriate for the chosen method, the panel of standard setters are required to conceptualise the minimum level of performance required for a pass in the examination. In order to do so, they must be knowledgeable about the candidate population, the standard of practice in the field, the content of the examination and, importantly, the reference group (e.g. the day one specialist) which forms the basis of the performance standard. [6,3,4,5]

Standard setters should receive training, so that they can provide their judgements in an informed manner. This training should familiarise them with both the standard setting task and the conceptual level required for a pass. It should ideally provide an opportunity for standard setters to calibrate their expectations using past performance data.[2,7] The Academy recommends documenting the qualification and experience of the panel of standard setters,[2] and gathering their feedback on the training they received.[3]

04
Make the standard as reproducible as possible

Variability between standard setters is inevitable, but, in addition to training the standard setters, implementation details can help maximise reproducibility. Some examinations use close marking schemes to capture candidates' level of ability in each marking episode. Performance levels should have labels which reflect the purpose of the examination [3] (e.g. day one consultant). The minimum level required for a pass should be clearly described and discussed by the panel at the start of the standard setting process.[4,2] For examinee-centred methods, considerations should include the number of candidates whose data are used, and the strength of the relationship between global judgements and numerical scores. For test-centred methods, careful consideration should be given to the number of standard setters required. Estimates range from 5-12 [4] to 15-30 [7] for high stakes examinations, with a consensus around 10-15. [7,8]

Steps can be taken to minimise the statistical impact of outliers, and if needed, the process can be repeated with additional judges to achieve greater reproducibility. [3,4] These should be decided in advance of the standard setting process, to avoid any temptation to manipulate the outcome.

As previously stated, the standard will reside with the examiners, which reinforces the need for training in standard setting as well as the importance of documented decision making in deciding the standard.

05
Acknowledge the measurement error and have a clear rationale for what to do about it

Ideally, the Standard Error of Measurement (SEM) in the examination should be reported. [2]

The variability among standard setters should also be reported. [2] An indication of the reproducibility of the standard should be made available, either based on the available data (standard error of the mean in test-centred methods, strength of relationship between scores and global judgements in examinee-centred methods), or as inferred from the implementation details of the standard setting method.

Because the measurement error intrinsic to scores and standards is associated with a risk of false negative outcomes (failing candidates who deserved to pass) and false positive outcomes (passing candidates who deserved to fail), *'explicit consideration of the relative consequences of each type of [measurement] error, along with the purpose of the examination, should be translated into an explicit policy that helps to guide standard-setting decisions and any adjustments'*. [1]

06
Carry out a sense check on the outcomes

It is important to carry out a sense check on the outcomes, particularly in a new or unfamiliar context (e.g. a new examination, a change in standard setting method, a change in the candidate population) A sense check is necessary to ensure that the standard was set appropriately. [3,2,9] This can be done by triangulating the pass rate and pass marks to previous diets. How do the pass/fail decisions compare to other information available about the candidate? However, the nature and timing of the sense check are controversial. [3,4,8,2] If the minimally competent candidate was adequately described and discussed before standard setting, if training exercises used real data from past diets, and if due process was followed, there should not be a need to revisit the standard once it has been set. The ultimate responsibility for setting the pass mark will reside with the Chairman and board of examiners. In all cases it is essential that the process is transparent and clearly recorded.

07
Follow due process and be transparent

Since no single method and no single set of procedures can guarantee the defensibility of the standard, [9] there is a duty of transparency towards all stakeholders around the various decisions and their implementations. Documenting how due process was followed allows the stakeholders to see the systematicity of the approach, [1,2] and therefore forms part of the defensibility evidence for the standard. [10,3,4] Following due process may at times result in uncomfortable outcomes, such as a 0% pass rate, or a different pass mark on different days of an examination. Transparency and clear communication about the process should help maintain both good practice and the acceptability of its outcomes to all stakeholders.

08
Consider how decisions related to standard setting sit with the principles underpinning examination design

Within the Utility model, [11] assessment bodies are advised to consider the validity, reliability, educational impact, feasibility and acceptability of examinations. In addition, due consideration must be given to equality and diversity issues covered by the Equality Act 2010. [12] Applying these to standard setting requires us to ask the following questions:

Validity Does the chosen standard setting method allow us to categorise candidates meaningfully in line with the purpose of the examination?

Reliability Does the chosen standard setting method give us confidence that a candidate's pass/fail outcome would be the same, regardless of which diet they sat the examination?

Educational impact Does the documentation available to candidates about the level required for a pass allow candidates to prepare well for the examination?

Feasibility Is the chosen standard setting method appropriate within the constraints of our examination (number of candidates, availability of standard setters)

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?

References

1 S. M. Downing. (2006) Twelve steps for effective test development*, Handbook of test development*. (S. M. Downing and T. M. Haladyna, eds.), pp. 3–26, New York: Routledge.

2 R. K. Hambleton and M. J. Pitoniak. (2006) Setting performance standards*, Educational Measurement*. (R. L. Brennan, ed.), pp. 433–470, Westport, CT: Publishers, Praeger.

3 G. J. Cizek. (2006) Standard setting*. Handbook of test development* (S. M. Downing and T. M. Hala- dyna, eds.), pp. 225–260, New York: Routledge.

4 S. M. Downing, A. Tekian, and R. Yudkowsky. (2006) Procedures for establishing defensible absolute passing scores on performance examinations *Health professions education. Teaching and learning in medicine*, vol. 18, pp. 50–7, Jan. 2006.

5 M. Friedman Ben-David. (2000) AMEE Guide No. 18 : Standard setting in student assessment*, Medical teacher*, vol. 22, no. 2, pp. 120–130, 2000

6 B. E. Clauser, M. J. Margolis, S. M. Case (2006) Testing for licensure and certification in the professions *Educational Measurement* (R. L. Brennan, ed.), pp. 701–732, Westport, CT: Praeger Publishers.

7 M. R. Raymond and J. B. Reid. (2001) Who made thee a judge? Selecting and training participants for standards on complex performance assessments, *Setting performance standards: Concepts, methods, and perspectives* (G. J. Cizek, ed.), pp. 119–157, Mahwah, NJ: Lawrence Erlbaum Asso- ciates.

8 S. L. Fowell, R. Fewtrell, and P. J. McLaughlin. (2008) Estimating the minimum number of judges required for test-centred standard setting on written assessments. Do discussion and iteration have an influence? *Advances in health sciences education : theory and practice*, vol. 13, pp. 11–24, Mar. 2008.

9 AERA, APA, and NCME. (1999), *The standards for educational and psychological testing*. Washington, DC: AERA Publications.

10 D. F. Becker and M. R. Pomplun (2006) Technical reporting and documentation, Handbook of test development, pp. 711–724, New York: Routledge, 2006.

11 C. P. M. van der Vleuten. (1996) The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education*, vol. 1, no. 1, pp. 41–67, 1996.

12 GMC. (2015) *Approving changes to curricula, examinations and assessments: equality and diversity requirements*. GMC

Glossary

Anchor item
An item with known performance characteristics, which is included in more than one test in order to provide comparative information about the items in a new version of the test and also about the test takers attempting it.

Angoff Method
A method of standard setting where the passing score is computed from an estimate of the probability of a hypothetical borderline candidate answering each item correctly. After a discussion and agreement on the characteristics of a borderline candidate, each judge (subject matter expert) makes an independent assessment of the probability that a borderline candidate will answer the item correctly. The judges' assessments of each item are averaged to determine the probability of a correct response for that item. Each average probability assigned to each item in the exam is averaged to obtain the cut score.

Bookmark method
A type of standard setting method where questions are ordered by difficulty from easiest to hardest. Judges make *bookmark* determinations of where performance levels (e.g. cut scores) should be, i.e. As the test gets harder, where would a participant on the boundary of the performance level not be able to get any more questions correct? This becomes the cut score.

Borderline Candidates
Candidates in examinations whose results are close to the pass-mark and who may have passed when they should have failed (false positive), or failed when they should have passed (false negative). They are often identified statistically by using the Standard Error of Measurement. Examining boards need a policy regarding the identification of borderline candidates and, having identified them, what to do about them.

Borderline group method
The borderline group method is an approach to standard setting, used by many bodies including the GMC, for the PLAB examination. Judges are asked to identify borderline participants. The median test score for these candidates is calculated and used as the cut score.

Borderline Regression Method
In the Borderline Regression (BLR) method, judges observe candidates performing the clinical task at each station. They score the various components of the clinical task on the predefined checklist. Assessors then provide a separate overall rating or a global score of the candidate's performance (for example: Outstanding, very good, pass, borderline or fail). The pass mark for each Objective Structured Clinical Examination (OSCE) station is then calculated by statistically regressing candidates' checklist scores on global scores for each station. The overall pass mark of the OSCE is calculated by aggregating the pass marks for each of the separate OSCE stations.

Calibration
Calibration is the exercise of checking the accuracy of a standard setting method, by comparison with another standard, other evidence or data. It may also include adjustment of the standard in question thereafter to bring it in line with comparative data.

Classical test theory
Classical test theory (CTT) assumes that a test score consists of a true score (the latent trait) plus an error score (that is due to measurement error). The concept of reliability (the extent to which a diagnostic instrument shows consistent results in repeated application) is central to CTT: It increases with decreasing error variance.

Clinical Competence
A student's ability to do what is expected at a satisfactory level of facility, at a certain point in time. The acquisition of a body of relevant knowledge and of a range of relevant skills including personal, interpersonal, clinical and technical components. In the case of clinical education, which is primarily based on an apprenticeship model, teachers define what the student is expected to do and then test their ability to do it.

Compensatory (also referred to as fully compensatory)
A compensatory standard is one in which any candidate who achieves a defined total score will pass, regardless of performance in different aspect/areas of the examination. In an OSCE where the standard is set on the total test score, which is the average performance across stations, the skill standard will constitute a compensatory standard. This method of scoring permits candidates to compensate for relatively poor skill performance on some stations. For example, so long as the sum of scores obtained on each station is greater than the pass mark, the candidate is awarded a pass; the performance on individual stations is not considered in the pass/fail decision.

Competency
The knowledge, skill, attitude or combination of these, that enables a candidate to effectively perform the activities of a particular occupation or role to the standards expected.

Conjunctive
Conjunctive standards are set by stressing the most important parts of the assessment or making performance on a given item decisive rather than or as well as relying on overall score. In an OSCE, it may be decided that candidates should competently *pass* a number of stations before a passing score is warranted. Similarly in a written test, developers may decide that candidates must exceed each assessment area separately (e.g. pharmacology, genetics, neuroscience), and thus candidates are not able to compensate for relatively poor performance in one area with relatively better performance in another.

Correlation
Defined as the relationship between two or more events and expressed as a number. Two identical events have a correlation of +1.0, events not related at all have a correlation of 0.0, and two completely opposite events have a correlation of -1.0. In psychometrics tests, correlation is used to convey the validity between a test and the criterion being measured.

Confidence intervals
The confidence intervals for a specific statistic (e.g. the mean) give us a range of values around the statistic where the *true* (population) statistic can be expected to be located (with a given level of certainty.

Contrasting groups method
A method of standing setting where judges are asked to categorise the participants in their classes according to the performance category descriptions (e.g. *qualified* and *non qualified*). The test is administered to all of the categorised participants and the test score distributions for each of the categorised groups are compared. Where the distributions of the contrasting groups intersect is where the cut score would be located.

Criterion referencing
Criterion referenced assessment measures performance against an absolute standard. Each candidate's performance is reviewed against a benchmark (usually the pass mark), rather than other candidates in the examination cohort (Norm referencing).

Cronbach's alpha
The most commonly measured aspect of reliability of a test – internal consistency. It is an average of all possible split half reliability measurements, expressed as a value between 0 and 1. The generally accepted value of Cronbach's alpha for a test is 0.7-0.8.

Cut score
A specified point on a score scale, such that scores at or above that point are interpreted differently from scores below that point (see also Standard Setting).

Dichotomous Angoff method
In the dichotomous modified Angoff approach to standard setting, instead of using difficulty level type statistics (proportion of borderline candidates who would get it correct), judges are asked to simply provide a 0/1 for each question (0 if a borderline acceptable participant would get the question wrong and 1 if a borderline acceptable participant would get the item right). A pass mark is calculated by averaging the scores.

Discriminator
A test item that discriminates well between weaker and stronger test takers, with stronger candidates performing statistically better than weaker ones.

Ebel Method
The Ebel standard setting method involves rating an examination question on two dimensions. The first dimension relates to the importance of the assessed item – is it essential, important or acceptable (nice-to-know) material for the students being assessed? The second dimension relates to whether the assessed item is easy, medium or hard. Each member of a panel of standard-setters completes a 3x3 grid, allocating every question to one of the nine boxes in the grid. After the panel of subject matter experts agrees on the definition of a minimally competent examinee, a determination is made as to the percentage of questions in each of the nine categories such an examinee would answer correctly. This percentage is multiplied in each of the nine categories by the number of questions it contains. The passing score is set by averaging the category scores.

Examination
A formal, controlled method or procedure to access an individual's knowledge, skills and abilities. Examinations involve written or oral responses, or observation of the candidate performing practical tasks (such as in an Objective Structured Clinical Examination).

Generalisability theory
This is an extension of classical reliability theory and methodology that is now becoming the preferred option. Analysis indicates the magnitude of errors from various specified sources, such as number of items in the assessment, the number of assessors etc. The analysis is used both to indicate the reliability of the test and to evaluate the generalisability beyond the specific sample of items, persons and observational conditions that were studied.

Hofstee Method
The Hofstee method of standard setting combines aspects of both relative and absolute methods. Panelists are asked to give their impressions of what the minimum and maximum failure rates should be for the exam, as well as what the minimum and maximum percent correct scores should be. These minimum and maximum failure rates and percent correct scores are averaged across panelists and projected onto the actual score distribution to derive a passing score.

Item response theory
A set of mathematical models for relating an individual's performance in a test to that individual's level of ability. These models are based on the fundamental theory that an individual's expected performance on a particular test item is a function of both the level of difficulty of the item and the individual's level of ability. Item response theory also examines individual items in relation to each other, and to the test as a whole, quantifying such characteristics as item difficulty and their ability to discriminate between good and poor candidates.

Marking Descriptor
Descriptions used to judge the standard and extent to which each learning outcome has been achieved in order to arrive at an appropriate mark.

Measurement error
The difference between the *true* score (the appropriate score for the candidate) and the score actually obtained in an assessment. Measurement error is present in all assessments, but can be minimised by good item design and, up to a point, by increasing the number of test items. It is usually calculated as the Standard Error of Measurement.

Multiple choice questions
A type of question in a written exam comprising of a lead-in statement (typically a short clinical description) followed by a homologous list of options (five is generally considered the optimum) from which the candidate selects the best answer.

Nedelsky method
A standard setting method where judges make decisions on a question-by-question basis regarding which of the question distracters they feel borderline participants would be able to eliminate as incorrect. The number of plausible options remaining determines the probability that the candidate will answer correctly; i.e. one plausible response = 100%, two = 50%, three = 33%, etc. The average of all probabilities determines the cut score. This method is generally used with multiple-choice questions only.

Norm referencing
A method of establishing passing and failing candidates based on their performance in relation to each other, rather than to an established standard (criterion referencing). So for example, only the top n number or x% of candidates pass, irrespective of how strong or weak the cohort is as a whole. Norm referencing should be used only in certain special circumstances, for example, where there is a limited number of posts available for successful candidates to move on to.

Objective Structured Clinical Examination
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.

Outlier
An observation or data-point that lies an abnormal distance from other values in a random sample from a population. What will be considered abnormal is context-dependent. However, an outlier is commonly defined as being outside a specified number of standard deviations above or below the mean, or a point which falls more than 1.5 times the interquartile range above the third quartile or below the first quartile.

Partially compensatory approach
The candidate's total score must exceed the total pass mark, and the candidate must also fulfil one or more additional conditions. For example, the candidate's total score must exceed the pass mark AND the candidate must pass at least eight stations AND the candidate must receive no *catastrophic fail* marks in any given station.

Pass mark
The score that allows a candidate to pass an examination or assessment. Also commonly termed *cut-point*, a passmark serves as a boundary between those who meet a required standard and those that do not.

Raw score
A test mark that has not been modified (for example, in the light of reliability calculations).

Reliability
The chosen standard setting method should give confidence that the candidate's pass/fail outcome would be the same regardless of which diet of the examination was taken. Various forms of reliability include:

- **Inter-rater reliability** – the extent to which different assessors give similar ratings for similar performances
- **Intra-rater reliability** – the extent to which a single assessor would give similar marks for almost identical performance, or would be consistent if re-marking a test item
- **Parallel forms reliability** – the consistency of results between two or more forms of the same assessment, testing in the same domain. This, along with test-retest reliability to which it is closely associated, is highly significant in examinations such as Royal College membership and Fellowship examinations, and involves such concepts as producing matching items, standard setting and examiner training
- **Test-retest reliability (or stability)** – the degree to which the same test produces the same results when repeated under the same conditions.

Sequential testing
In this approach those who fail a test will undertake another one.

Simulated patients
Individuals who are not ill but adopt a patient's history and role for learning or assessment in medical education. Actors are often used to accomplish this goal, these are also known as Role Players.

Standard deviation
The square root of the variance, used to indicate the spread of group scores and a component of the equation to calculate Standard Error of Measurement. A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values.

Standard error of measurement
Standard Errors of Measurement (SEM) gives an indication of the accuracy with which an individual candidate's performance has been assessed.This is important in identifying borderline trainees. In high stakes examinations, borderline trainees would be those within two or even three SEMs of the pass mark.

Standard Setting
The process of establishing a cut score(s) for a test. By defining a *performance standard* i.e. the minimally adequate level of performance for a given purpose, a cut score or *pass mark* is determined.

Standards
In medical education standards may be defined as a model design or formulations related to different aspects of medical education and presented in such a way to make possible assessment of graduates' performance in compliance with generally accepted professional requirements.

Summative assessment
Assessment carried out for the purpose of (usually pass/fail) decision making e.g. an examination. It is important to note the distinction between formative assessment (to aid improvement) and summative assessment (for decision making).

Systematic Bias
The tendency to consistently underestimate or overestimate a true score or value. Leads to systematic error, a source of measurement error which is predictable and consistent.

Triangulation
Using multiple data sources (e.g. passmarks, passrates) in an investigation to corroborate findings or facilitate understanding.

True score
A candidate's score on a test without measurement error. The true score is the observed score minus the error.

Validity
In the case of assessment, validity refers to the degree to which a measurement instrument truly measures what it is supposed to measure. It is concerned with whether the right things are being assessed, in the right way, and with a positive influence on learning. It is increasingly regarded as a unitary concept, but there have been several different dimensions of validity distinguished in the past including:

- **Content validity**. An assessment has content validity if the components reflect the abilities (knowledge, skills or behaviours) it is designed to measure
- **Face validity**. This is related to content validity. Face validity can be described from the perspective of interested lay observers. If they feel that the right things are being assessed in the right way, then the assessment has good face validity
- **Construct validity**. The extent to which the assessment, and the individual components of the assessment, tests relevant professional constructs. For instance, an assessment has construct validity if more advanced doctors in training achieve higher scores than less advanced doctors in training
- **Concurrent validity**. The extent to which the results of a test are consistent with the results of another test that is intended to assess the same thing.
- **Predictive validity**. This refers to the degree to which an assessment predicts expected outcomes. For example, a measure of attitudes (behaviour) toward preventive care should correlate significantly with preventive care behaviours
- Consequential validity (educational impact). This refers to the effect that an assessment has on learning, and in particular on what doctors in training learn and how they learn it.

Weighting
Assigning different values to different items, reflecting, for example, their importance or difficulty in order to increase the effectiveness of a test.

Z-Score
A standardised score that indicates how many standard deviations a data point (or test score) is from the mean. Calculated by subtracting the population mean from the individual score, and dividing this difference by the population standard deviation ($z = (X-\mu)/\sigma$).

Working group

The Academy Assessment Committee set up a small working group to develop these standards.

Mr Ian Richie	Chair / Royal College of Surgeons of Edinburgh
Professor Chris McManus	University College London
Ms Daniela Warr Schori	Membership of the Royal Colleges of Physicians of the United Kingdom
Prof Philip Turner	Joint Committee on Intercollegiate Examinations
Mr Grant Fisher	Faculty of Public Health
Dr Mark Blunt	Royal College of Anaesthetists
Ms Kiran Sanghara	Royal College of Psychiatrists
Ms Lucy Foard	Royal College of Paediatrics and Child Health
Dr Nicki Williams	Royal College of General Practitioners
Ms Paula Robblee	General Medical Council
Dr Cara Featherstone	Royal College of Surgeons of Edinburgh
Dr Yvonne Hurst	Royal College of Surgeons of Edinburgh

Academy of Medical Royal Colleges
10 Dallington Street
London
EC1V 0DB
United Kingdom

Telephone
+44 (0)20 7490 6810

Facsimile
+44 (0)20 7470 6811

Email
academy@aomrc.org.uk

Website
www.aomrc.org.uk

Registered Charity Number
1056565