



Designing and maintaining postgraduate assessment programmes

Working with doctors Working for patients

General
Medical
Council

Contents

Section	Paragraphs
About this guidance	1-24
Part 1: Designing programmes of assessment	
Setting out the purpose of the programme of assessment (Theme 1, CR4.1a, CR4.1b, CR4.2, CR4.3, CR4.4)	25-37
Linking curriculum content and assessment (CR4.1d-e)	38-44
Content decisions: formative and summative (CR2.4, CR4.1h, CR4.2, CR4.8)	45-52
Evidencing format decisions: acceptability, feasibility, cost effectiveness (CR2.5a-f, CR4.1b-c, CR4.4)	53-54
Part 2: Managing programmes of assessment	
Evidencing content quality: setting up structures to ensure the quality of assessment (CR2.5a-f, CR4.1g, CR4.5, CR4.6)	55-63
Setting out expectations: learners (CR2.4, CR4.1g, CR4.2, CR4.4)	64-65
Setting out expectations: examiners and assessors (CR4.1g, CR4.4, CR4.9, CR4.10, CR4.11, CR4.12, CR4.13)	66-70
Evidencing decisions about assessment structure: standard setting (CR4.1f, CR4.11, CR4.12)	71-76
Evidence about assessment structure: statistical analysis (CR4.1a, CR4.5, CR4.6, CR4.7)	77-85
Part 3: The impact of a decision	
Evidencing the impact of assessments (CR4.2, CR4.3, CR4.4, CR4.5, CR4.6)	86-90
Appendix 1: Annual publication of exam data	
Key terms used in this guidance	
Bibliography	

About this guidance

What's the purpose of this guidance?

- 1 This guidance is supplementary to our curriculum standards, [Excellence by design](#). It gives advice for those making decisions in organisations (typically colleges and faculties) about how programmes of assessment are designed and maintained.
- 2 *Excellence by design* emphasises five 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.
- 3 The standards require curricula to describe fewer, high-level generic, shared and specialty-specific outcomes, which will support all doctors better in understanding what is expected of them in their training programme. They require curricula to identify shared areas of training and to have a greater focus on the generic professional capabilities common to all doctors. These requirements for approval will enable improvement in the flexibility, relevance and consistency of postgraduate medical training.
 - Generic outcomes are those that are common across all specialties.
 - Shared outcome are those common components of training across groups or families of specialties. The [flexibility review](#) identified this as an area of work to be developed.
 - Specialty-specific outcomes are defined those that relate to specific areas of specialist practice.
- 4 Our standards also require assessment to be proportionate and impose a reasonable and necessary amount of assessment activity on learners and their trainers.
- 5 The guidance sets out the steps and principles that you should follow when:
 - planning and designing a programme of assessment

- determining expected levels of performance for critical progression points (points or transitions in (or between) all training programmes that may represent significant risk to patients, the service and those in training) and satisfactory completion of training
 - maintaining its quality and validity in practice.
- 6 It also suggests approaches that you could use, but these are not prescriptive. By focusing further on the fundamental principles and practical considerations outlined in *Excellence by design* it is hoped this guidance will support organisations as they develop their assessment philosophy, strategy and programme of assessment.
- 7 While this guidance does not describe or advise on the approvals process, the principles contained in this document are the same as those underpinning our curriculum standards *Excellence by design* and associated approvals process.

Who is this guidance for?

- 8 This guidance gives advice and information for colleges and faculties to support those making decisions about the programme of assessment in their curricula. It is written with the intention of assisting in the design, development and management of the programme of assessment. The guidance relates to the applicable section of the standards and explores what approaches and evidence might be helpful to consider and provide when submitting a curriculum to the formal approvals process.
- 9 This guidance may also be relevant to other organisations involved in medical education in the UK. Local training bodies conducting assessments in the workplace are subject to the standards in [Promoting excellence](#) (see especially R1.18 and S1.2).

What we mean by assessment

- 10 We define assessment as all activity aimed at judging a learner's attainment of curriculum outcomes, whether for summative (determining satisfactory progression in or completion of training), or formative (developmental) purposes. An outcome can be defined as a level of performance or behaviour that a trainee is expected to achieve as part of their development according to their stage of training within their curriculum. This can include an area of professional practice that may be trusted to a learner to execute unsupervised, once he or she has demonstrated the required competence. An examination is an example of an individual assessment test.
- 11 Assessments need to:
- identify learners who have not demonstrated the expected level of performance, attainment or achievement needed to progress in or complete training

- give learners appropriate opportunities to receive timely feedback that provides a basis for action, so that they can understand what is expected at their level of practice and provide them with evidence and guidance as to how they can act to improve their performance and continue to develop. As well as reaching minimum standards for safe competent practice, learners should be encouraged and have the opportunities to excel at all stages and levels of training.

What's new in this guidance?

Programme of assessment

- 12** A programme of assessment refers to the integrated framework of exams, assessments in the workplace and judgements made about a learner during their approved programme of training. The purpose of the programme of assessment is to robustly evidence, ensure and clearly communicate the expected levels of performance at critical progression points in, and to demonstrate satisfactory completion of training as required by the approved curriculum. The programme of assessment should include the overall assessment philosophy, the assessment strategy and practical operational aspects such as guidance to assessors and expected levels of performance.
- 13** The programme of assessment is likely to be comprised of several different individual types of assessment. These may include national examinations, summative assessments, assessments in practice and formative or developmental assessments such as supervised learning events. The choice of methods relates to what learning outcome is being assessed and why, and the consequences or importance of the assessment being performed. A range of assessments may be needed to generate the necessary evidence required for global judgements to be made about satisfactory performance, progression in, and completion of, training. All assessments, including those conducted in the workplace, must be linked to the relevant curricular learning outcomes (eg through the blueprinting of assessment system to the stated curricular outcomes).

Safe management of critical progression points in training

- 14** Critical progression points are points of increased training risk must be identified and safely managed through the requirements, assessments and guidance set out in the approved curricula and programme of assessment.
- 15** These critical progression points will include when transitioning to higher levels of professional responsibility or entering a new or specialist area of practice or when a trainee is being considered or deemed to have satisfactorily completed the programme of training.

- 16 As well as progression at critical points, ensuring that learners receive annual review of their progress and performance is also required.

Greater emphasis on validity

- 17 Validity is seen as the key consideration in current assessment theory ([Health Professional Assessment Consultancy \(HPAC\) 2016](#)). We define validity as 'interpretations and uses of tests that make sense and are supported by appropriate evidence' (adapted from Kane 2013:3). A detailed description of validity theory is provided in HPAC's report.

Assessment strategy based on the validity model



- 18 HPAC (2016:11) note this approach means that organisations that award qualifications (such as those leading to a certificate of completion of training (CCT)) are required to 'provide evidence that their awards are granted to individuals who have achieved defined recognisable standards in the various domains of competence'. Validity theory identifies three stages in this process.

- Purpose: setting out of the explicit purpose of particular assessments, clearly communicating their contribution to the wider programme of assessment and

decision making process particularly in relation to progression and satisfactory completion of training (what is being tested and why) (covered in part 1)

- Evidence: collecting five key types of evidence about the assessment or programme of assessment:
 - the choice, content and format of assessments – are the assessment methods that have been chosen appropriate for the declared outcome being assessed? (covered in part 1)
 - the practical conduct of the assessment, eg whether those assessing and being assessed understand what they are required to do or demonstrate, or whether exams are scored or conducted correctly and effectively (covered in part 2)
 - the internal structure, incorporating issues including the psychometric performance of a test – eg is an examination sufficiently reliable, for example? (covered in part 2)
 - the relationships between different assessments within the programme of assessment – do assessments that intend to test similar things do so in practice? Are they all necessary, what particular value do they add and why? (covered in part 2)
 - the consequences – describing the consequences of the assessment, and how decisions made using assessments are defensible. Decisions should be consistent, defensible and fair to doctors in training and provide appropriate assurance about the safety and quality of their practice (HPAC 2016). The consequences of a decision will affect the evidence needed to support it; more impactful decisions will need more robust evidence (Kane 2013). (covered in part 3)
- Argument: bringing the evidence together to argue that assessments are valid for their intended purpose.

19 This approach should form the basis of the planning, conduct and management of assessment, and be used to carry out quality review and improvement of the assessment. HPAC (2016:12) note that this includes the information provided to learners and the information and training provided for those assessing.

Outcomes-based curricula

20 *Excellence by design*, our standards for curricula, requires all postgraduate curricula to describe appropriate high level outcomes as generic, shared and specialty-specific professional capabilities. Outcomes-based curricula focus on what kind of capabilities doctors will have upon completion of the programme rather than the process by

which these capabilities are achieved (Harden et al 1999). Doctors are required to demonstrate complex knowledge and skills, but (good) doctors are defined by much more than this. Training programmes should ultimately aspire to assess and evidence a learner's overall performance and professional capabilities.

Incorporating *Good medical practice* and the *Generic professional capabilities framework* as generic outcomes in the programmes of assessment

21 In the standards, we require organisations to develop outcomes-based curricula containing high level generic, shared and specialty-specific outcomes. The expectations of doctors outlined in *Good medical practice* and the *Generic professional capabilities framework* must be included in all postgraduate medical curricula as minimum regulatory requirements of training programmes.

The importance of professional judgement in all assessments

22 The guidance emphasises the importance and centrality of professional judgment in making sure learners have met the learning outcomes and expected levels of performance set out in the approved curricula. Assessors must use their professional expertise and experience. Through their understanding of the expected levels of performance, they must make accountable, professional judgements as part of a valid programme of assessment. A coherent and integrated *programme of assessment* will include how professional judgements are used and collated to support decisions on progression and satisfactory completion of training.

College and faculty responsibilities for administration and conduct of assessments

23 The quality of administration and governance is part of an assessment's validity. We describe colleges' responsibilities for:

- quality managing the assessments they carry out themselves; they will also have a role in
- supporting and enabling the quality of assessments carried out locally by issuing appropriate guidance about assessment methods, decision aids and specialty-specific expected levels of performance at different stages or critical progression points in training.

Flexibility

24 Our curriculum standards allow for greater flexibility in the way assessments are designed. Those designing a programme of assessment will need to describe what informed their choices and how these considerations support the overall validity of the programme of assessment.

Part 1: Designing programmes of assessment

Setting out the purpose of the programme of assessment

Excellence by design: Theme 1, CR4.1a, CR4.1b, CR4.2, CR4.3, CR4.4

This section relates to validity theory requirements that organisations need to set out the purpose of a programme of assessment as a whole and its individual components.

Key issues in this section

Ref	Key issues for consideration
CR4.1a CR4.1b	<p>What capabilities and kind of doctors are you aiming to produce? What is their scope of practice and level of performance?</p> <p>How does the programme of assessment help to achieve or confirm these objectives?</p> <p>Where is this written down?</p> <p>What are the underlying principles and purpose of the programme of assessment?</p> <p>Does each assessment within the programme of assessment have a clear purpose in relation to the curricular outcomes? Is it clear what methods will be used to assess and when?</p> <p>How do they relate to critical progression points?</p> <p>Is the choice of methods and timing supported by a clear rationale, support, research or informed practice?</p>
CR4.3	<p>How does the programme of assessment aim to capture the doctor's professional development towards achieving curricular learning outcomes over time?</p> <p>How is longitudinal development intended to be captured in assessment?</p> <p>Where and how is global judgement of a learner's overall performance made?</p> <p>What is the format of the test (numbers/length/scoring)? How was this decided?</p>
CR4.2	<p>How is underperformance by learners identified, and what is the approach to its management? Are there decision aids or guidance frameworks?</p> <p>Can learners who aren't making progress be identified? What happens to them then?</p>
CR4.4	<p>What are the equality considerations of your choices?</p>

Ref	Considerations for approvals or quality assurance
CR4.1a	<ul style="list-style-type: none"> ■ Clear articulation of desired purpose of each element of an assessment, and indication that this has informed choice of format to ensure validity
CR4.1b	<ul style="list-style-type: none"> ■ Assessment strategy document clearly presenting the identified purpose of each assessment element in relation to one another and in the context of the wider curricular learning outcomes. Identification of the role each summative assessment plays in progression decisions and satisfactory completion of training
CR4.3	<ul style="list-style-type: none"> ■ Systematic approach to identifying each area or level of attainment required prior to critical progression points or for satisfactory completion of the training programme (documented and highlighted in a matrix/overarching blueprint) ■ Systematic approach to consistently identifying discrepancies between learners' performance, behaviour and attainment of expected levels of performance across different components of the programme of assessment

Guidance

Assessing outcomes using a programmatic approach (CR4.1b)

- 25** A programme of assessment is designed to demonstrate that a learner has met the learning outcomes of the approved curriculum. Since most outcomes are not confined to one-off assessments (Schuwirth and van der Vleuten 2011), assessments should be part of an integrated programme of assessment that gives learners multiple opportunities for feedback and development, and to demonstrate the required outcomes over the course of their training.
- 26** A programme of assessment must have a clearly stated overall purpose in relation to the curriculum and must clearly state the purpose of its individual assessment components, setting out the range of different assessments that can contribute different evidence to support overall judgements about performance and decisions about progression and satisfactory completion of training.
- 27** This means organisations will need to plan their programme of assessment (not just examinations) as an integrated, interdependent, programme to show doctors have achieved the relevant learning outcomes at appropriate stages of training. Programmes of assessment should be synoptic in design, asking learners to integrate learning from across the programme, applying 'their skills, knowledge and understanding with breadth and depth' (Plymouth University, undated).

Designing a strategy as a basis for a programme of assessment (CR4.1b)

- 28** An organisation should take a structured, coordinated approach to the design and development of its programme of assessment and communicate these in the form of an assessment strategy. This will be required to demonstrate that the programme of assessment meets the standards set out in *Excellence by design*. The fundamental principles and practical considerations must be assembled into a coherent strategy that demonstrates how judgements made throughout the programme of assessment can be used to justify that learners can be deemed to have safely and competently completed the approved programme of training.
- 29** Specifically, the overall assessment strategy should set out how organisations have made sure:
- assessments contribute to enabling safe, high quality care for patients
 - all learners have opportunities to develop and improve their performance from feedback to achieve the approved learning outcomes
 - the assessment approaches adopted afford all groups of learners a fair opportunity to develop or demonstrate they have achieved the required learning outcomes (subject to patient safety considerations)
 - learners who have not met curricular outcomes are identified and there is a clear approach as to how they should be managed.
- 30** In adhering to these high-level principles, the assessment strategy document should clearly articulate:
- the purpose of each assessment and how this is ensured in the selection, development and validation of the format chosen
 - the combination of assessment methods that are to be used to test each part of the curriculum, and why they are appropriate and proportionate to what is being tested
 - standard setting principles and approaches in the context of expected levels of performance
 - the way assessments provide, in conjunction with one another, the required information to contribute appropriately to important decisions regarding progression within, and completion of, training.
- 31** The assessment strategy document does not require exacting detail for each assessment, but should give an overview of the programme as a whole and where and how critical progression point decisions should be made, highlighting key principles such as safety, expected levels of performance and scope of practice and

responsibility. We anticipate a wide range of approaches will be acceptable if supported by appropriate evidence, and a clear compelling narrative.

Making decisions at critical progression points within the programme of assessment (CR4.3)

- 32** The decisions made at critical progression points and upon completion of training should be clear and defensible. They must be fair and robust and make use of evidence from a range of assessments, potentially including exams and observations in practice or reflection on behaviour by those who have appropriate expertise or experience. They can also incorporate commentary or reports from longitudinal observations, such as from supervisors or formative assessments demonstrating progress over time (see also AoMRC 2016).
- 33** Periodic (at least annual) review should be used to collate and systematically review evidence about a doctor's performance and progress in a holistic way and make decisions about their progression in training. Current annual reviews of progression (ARCP) processes can support the collation and integration of evidence to make decisions about the achievement of expected outcomes.
- 34** Assessments such as entrustable professional activities-type formats also involve looking across a range of different skills and behaviours to make global decisions about a learner's suitability to take on particular responsibilities or tasks, as do decisions about the satisfactory completion of modules or defined areas of practice within curricula.
- 35** To put this into practice, organisations will need to:
- provide clear performance criteria and acceptable evidence at each critical progression point against which summative decisions and judgements can be made. Decision aids or flow diagrams may assist this process as will guidance describing expected or inadequate levels of performance
 - match assessments to appropriate points in the curriculum. In structuring programmes of assessment organisations should consider:
 - identifying how learners are expected to progress through the curriculum; we anticipate the latter stages of the programme of learning will include assessments that integrate complex evidence to reflect the increasing complexity and capability of those in training. The importance of making global judgements about overall performance and safe progression should be the guiding principle
 - describing what outcomes learners should have already demonstrated and to what level, and with what degree of confidence or reliability

- the overall balance and assessment load placed on learners, faculty, organisations and patients.

36 Some examples of approaches to structuring assessments organisations may find helpful to use, adapt or combine are:

- HPAC's programmatic assessment case study (2016)
- the portfolio approach to a curriculum module described in Roberts et al (2014)
- the approach developed by the Association of American Medical Colleges (2014) which is concerned with a learner's trustworthiness to complete clinical tasks that integrate different skills and knowledge independently (ten Cate 2013).

37 The choice is not limited to these approaches but decisions and choices should be set out and clearly justified in the assessment strategy and supporting documentation.

Linking curriculum content and assessment

Excellence by design: CR4.1d-e

This section discusses elements of the first domain of evidence set out in validity theory, which looks at evidence in relation to content and format decisions about assessments.

Key issues in this section

Ref	Key issues for consideration
CR4.1d	<p>How do you make sure those completing training meet our regulatory requirements in relation to demonstrating the generic outcomes? How do you embed <i>Good medical practice</i> and the generic professional capabilities throughout your curriculum or assessment programme?</p> <p>Can the college or faculty demonstrate that the learning outcomes and assessments have been clearly linked through blueprinting to the approved curriculum, and is the blueprint used as a basis for the programme of assessment?</p> <p>How are the expected levels of performance determined and what should they be at critical progression points in and for satisfactory completion of training?</p> <p>Examinations</p> <ul style="list-style-type: none"> ■ How do you map the programme of assessment to the learning outcomes? ■ How do you choose and develop items relevant to assessment blueprint domains? (HPAC 2A.1-3) <p>Assessments at work</p> <ul style="list-style-type: none"> ■ How do you map the skills required by the assessment task to learning outcomes? (HPAC 2B.1) ■ How are the results intended to be used?
CR4.1e	<p>With respect to the curricular outcomes, how are the assessments of skills, knowledge and capabilities balanced and demonstrated at the appropriate level throughout the programme of training?</p> <p>What are the critical progression points in training?</p> <p>What will ARCPs be required to demonstrate?</p>

Ref	Key issues for consideration
CR4.1d CR4.1e	<p>How do you make sure the programme of assessment continues to reflect the curricular purpose and outcomes?</p> <p>How are decisions about assessment made with appropriate oversight and input from those responsible for the curriculum?</p> <p>What structures make sure that those responsible for curriculum design and development have appropriate oversight and input into all aspects of the programme of assessment?</p>
Ref	Considerations for approvals or quality assurance
CR4.1d	<ul style="list-style-type: none"> ■ Blueprinting grid for each assessment element reflecting the organisation of the relevant syllabus (informed by the organisation of the curriculum), with sample population showing the division of labour between different formats used where applicable ■ Multi-dimensional approach to blueprinting, demonstrating how different domains within each area or module of the curriculum are covered by a given assessment (single clinical topic assessed with regard to basic science, management, investigations, communication etc) ■ Overarching assessment blueprint identifying what is assessed by which method in the context of the curriculum modules; clear identification of what is required by way of attainment in such a context for satisfactory progression or completion of training
CR4.1e	<ul style="list-style-type: none"> ■ Syllabus for each assessment clearly organised with reference to the organisation of the curriculum and linked to learning outcomes at each critical progression point so expected content or performance and standard is clear and progress can be monitored

Guidance

38 Our standards require programmes of assessment to be based on the curriculum, and must reflect the themes, duties and responsibilities described in *Good medical practice* and the *Generic professional capabilities framework*. As such organisations need to make sure that:

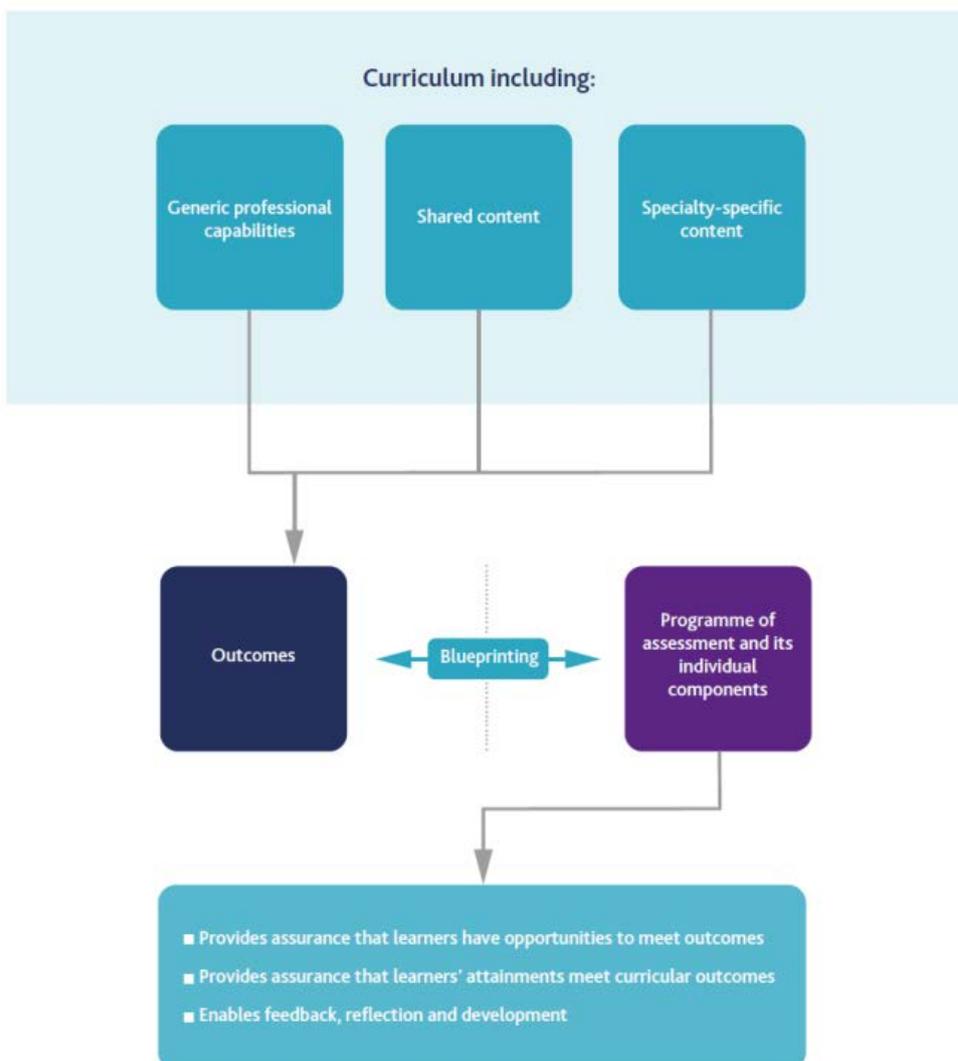
- curricula describe the generic, shared and specialty-specific learning outcomes required for satisfactory completion of the programme of training and that must be demonstrated at critical progression points within that training

- generic, shared and specialty-specific learning outcomes are linked to and demonstrated through evidence gathered throughout the approved programme of assessment
- the programme of assessments, the individual items within it and the content and format are derived exclusively from the content of the approved curriculum.

39 To make sure this is the case, those that are responsible for the curriculum should be closely involved in the development of the programme of assessment and those involved in assessment should be familiar with the curriculum; assessment developments should not be carried out independently of the curriculum requirements.

40 A blueprint is a template, table or matrix that provides the evidence that learners are judged against the stated learning outcomes of the approved curriculum. The generic, shared and specialty-specific outcomes of the curriculum provide the framework for the design, planning and evaluation of the programme of assessment.

Assessment should link to curriculum through a coherent strategy and blueprint



- 41** Blueprinting is a key exercise for developing the evidence to support the validity of an assessment (Coderre et al 2009). A blueprint specifies which assessment method is used to assess each learning outcome; it is also used to describe how the content of an exam should be chosen. It may also show the stage of training or critical progression points at which outcomes are assessed, and how content may be sampled across different assessments over time. Blueprinting of assessments against the curricular learning outcomes is essential in taking a systematic approach to the design of the programme of assessment; this is sometimes referred to as mapping.
- 42** As well as this overarching blueprint for the programme of assessment, individual tests should also have a blueprint showing how curricular content will be covered and sampled in the individual assessment.
- 43** Organisations may also design systems to collate and aggregate various low stakes assessment (including assessments at work) and provide evidence of coverage of outcomes (for example, see Maastricht case study in HPAC 2016). To do this, organisations need to provide a range of assessments mapped to curricular outcomes which can be completed, captured and used to demonstrate the spread of assessments completed by the learner, and provide a process by which the spread of assessments completed is reviewed against the criteria required for progression or satisfactory completion.
- 44** Failing to reflect changes in the curriculum within the programme of assessment will compromise the purpose and validity of the assessment system. For this reason, changes to the format, content, rules, standard or structure of assessments should not be made without appropriate oversight, involvement and agreement of those responsible for the curricula and vice versa. Similarly, blueprints must be kept up to date and reflect the approved curricular content and learning outcomes. Organisations should be able to describe why any standalone changes to the assessment system are required and be able to describe how the programme of assessment continues to meet the needs of the curriculum (or why it will be unaffected by the change, if the change is purely a technical assessment issue).

Content decisions: formative and summative

Excellence by design: CR2.4, CR4.1h, CR4.2, CR4.8

This section discusses elements of the first domain of evidence set out in validity theory, which looks at evidence in relation to content and format decisions about assessments.

Key issues in this section

Ref	Key issues for consideration
CR4.1h	<p>Where are summative decisions made about whether learners have achieved the stated outcomes?</p> <p>How are the different kinds of assessment reflected in blueprints and assessment strategies? What are the formative or developmental assessments?</p>
CR4.2 CR4.8	<p>How are feedback, improvement and remediation incorporated into your programme of assessments?</p> <p>What are the expectations regarding feedback for those conducting assessments locally?</p>
Ref	Considerations for approvals or quality assurance
CR4.8	<ul style="list-style-type: none"> ■ Feedback integrated into assessments at work, with sufficient prominence afforded to this in the case of formative assessments ■ Routine feedback provision with summative assessment results to enable unsuccessful candidates to identify and target specific areas of development prior to their next attempt at that examination or assessment ■ Routine feedback provision with summative assessment results to enable successful candidates to improve their performance ■ Routine feedback provision with summative assessment results to enable successful candidates to improve their performance
CR4.1h	<ul style="list-style-type: none"> ■ Assessment clearly framed as summative by explicit contextualisation with regard to a critical progression point or stage of training and inability to progress without successful completion; role of such decisions in relevant ARCP processes identified. Mandatory nature of expected levels of performance to progress and relevant limits on number of attempts clearly identified ■ Clear identification of formative and summative elements, along with how each part of the programme of assessment contributes to decisions regarding progression with training

Ref	Considerations for approvals or quality assurance
CR4.2	<ul style="list-style-type: none"> ■ Routine feedback provision to enable unsuccessful candidates to identify and target specific areas of weakness prior to their next attempt at the examination ■ Clear information for trainees regarding sources of guidance and support where difficulties are encountered in passing relevant assessments ■ Clear information regarding rationale and detail of the relevant attempts limit and any requirements or conditions for extra attempt(s) beyond this
CR4.2	<ul style="list-style-type: none"> ■ Process in place for examiners to identify potentially dangerous practice demonstrated by candidates within the context of an assessment ■ Feedback sufficiently instructive to identify occasions on which an unsuccessful candidate is some way below the minimum standard required, either in specific domains/syllabus areas, or in the assessment overall to ensure appropriate and timely remediation
CR2.4	<ul style="list-style-type: none"> ■ Clear information regarding the overall assessment system and its constituent parts available in the public domain

Guidance

45 Programmes of assessment combine several functions ([see paragraph 26 above](#)). Individual types of assessments within the programme may have a range of purposes, such as:

- identifying or developing an individual's strengths and weaknesses (CS4.6, CR4.3) to plan future learning, career guidance, remediation and professional development
- providing opportunities for reflection
- enabling key capabilities to be developed further through formative or developmental assessments particularly when outlining expected levels of performance or in the promotion of excellence
- demonstrating (partly or wholly) the achievement of curricular outcomes at critical progression points and preventing the progression of those who have not achieved them
- demonstrating achievement of the expected level of performance and in determining satisfactory completion of training.

- 46** The purpose of an assessment in relation to both its immediate objective and the role it plays more generally within the programme of assessment must be clear to learners and assessors. A lack of clarity regarding the purpose of an assessment can serve to undermine its validity by compromising the extent to which this purpose is understood and achieved.
- 47** All assessments should, as far as possible:
- generate effective feedback, from assessors with the right expertise and/or experience, and with appropriate training where required
 - prompt the learner to consider their own performance and development needs
 - result in the learner taking action and provide evidence of that action (eg through a further formative assessment or in the course of supervision).
- 48** As part of the requirement for curricula to make sure learners get appropriate developmental feedback, organisations should provide assessments at work that are primarily formative in nature. Dedicated advice on improving the quality of feedback is given by AoMRC (2017).
- 49** Formative assessments should:
- require and enable interaction between learners, assessors, teams and patients
 - be chosen or led by the learner or a trainer to gain experience of, feedback about or insight into one or more areas of the learner's performance.
- 50** Feedback from these formative assessments can be used to identify issues of engagement, professional development or serious underperformance, which can be communicated to those responsible for training programmes. In prioritising learning and feedback, such assessments should not require learners to demonstrate that they can progress; rather, the standard required to progress may be used as a benchmark to guide discussions, comment upon attainment and plan future learning and development.
- 51** Feedback from numerous small formative assessments can be monitored, collated and reviewed periodically to give a rounded view of the learner's performance and improvement over time. This can in turn enable reliable holistic judgement to be made about suitability for progression, eg at annual review (ARCP) (see HPAC 2016 case study, van der Vleuten et al 2012).
- 52** Organisations should also provide summative assessments and processes which:
- demonstrates that the learner has acquired (and maintained) knowledge and skills as required by the approved curriculum (eg knowledge or clinical exams, logbooks, assessments of practice)

- enables management of learners who do not demonstrate expected levels of performance, including providing further assessments to gain further evidence (for more information, see the AoMRC's guidance on implementing generic professional capabilities)
- give information on a learner's ability to practice safely within a defined area or aspects of it with progressive levels of supervision and accountability (multi-source feedback (MSF), supervision reports, portfolios, entrustable professional activities (EPAs), some assessments of practice)
- synthesises the range of assessment data at critical progression points to consider the overall performance of the learner and to make summative decisions and judgements about whether they are performing to the level expected/whether they are making acceptable progress towards achieving curricular outcomes. This is consistent with the approach in which assessments are neither formative nor summative but have different stakes attached (see Maastricht [case study in HPAC 2016](#)). Underpinning this should be:
 - clear information on the use of assessments for learners and assessors
 - a defensible process for collation and synthesis of this evidence, supported by guidance and decision aids to enable synthesis and decision making; currently this is provided through ARCP.

Evidencing format decisions: acceptability, feasibility, cost effectiveness

Excellence by design: CR2.5a-f, CR4.1b-c, CR4.4

This section discusses elements of the first domain of evidence set out in validity theory, which looks at evidence in relation to content and format decisions about assessments.

Key issues in this section

Ref	Key issues for consideration
CR2.5 a-f	How are the stakeholders of an assessment included or represented in the design process?
CR4.4	<p>What do you know about the experiences of different stakeholders, and especially different groups of learners?</p> <p>How were learners, supervisors, deaneries/Health Education England (HEE) local offices and local education providers, patients and the public involved in the development of, or change to, assessments (where appropriate)?</p> <p>What equality and diversity considerations were identified? How did they influence the outcome?</p>
CR4.1b CR4.1c	<p>How are you assured that assessments will function as intended in practice?</p> <p>Do new or modified assessments require piloting to determine their feasibility or reliability in practice?</p> <p>If so, what does piloting show?</p> <p>What other information did you gather about putting assessments into practice?</p> <p>What resources and guidance are required to implement the programme of assessment?</p> <p>How are patient safety and quality of care prioritised in assessment design?</p> <p>How do you decide that an assessment is safe to use?</p> <p>What rules or processes ensure safety?</p>

Ref	Considerations for approvals or quality assurance
CR4.1b	<ul style="list-style-type: none"> ■ Assessment strategy document clearly presents the identified purpose of each element in relation to one another and in the context of the assessment's outcome. Identification of the role each summative assessment plays in progression decisions within the broader training programme
CR4.1c	<ul style="list-style-type: none"> ■ Details of pilot structure and outcomes (include metrics where appropriate) for proposed new assessments ■ Details of pilot participants (prospective candidates or past candidates of the live assessment) ■ Evidence of routine monitoring in the inclusion of new questions whilst ensuring that reliability of the assessment is not compromised
CR4.4	<ul style="list-style-type: none"> ■ Equality and diversity considerations in assessment material development

Guidance

53 Assessments should be developed with input from those responsible for carrying them out and subject to them (R2.1–3). Ensuring feasibility and acceptability to assessors and doctors in training is a priority in assessments at work (AoMRC 2009, 2016), as is ensuring the environment has the capacity to deliver the assessment. Organisations could demonstrate this in a number of ways:

- Using (or establishing) groups which involve trainers, employers, learners and patients to understand what is likely to work in practice or not. This should be proportionate, and many organisations will already have structures for this.
 - We recommend including the service in the development of the assessment to ensure feasibility in the learning and working environment.
 - It is desirable to include a diverse range of stakeholders who share protected characteristics in such groups; different groups of learners are likely to have different experiences of undertaking assessments, which can affect the outcome. Taking steps to understand these experiences and include them in the process of producing these assessments may help organisations to understand and address the possible impacts of assessments.
- Identifying the resource required and the capacity of the environment to deliver them.
- Planning the programme to be cost effective and efficient in sampling and evidencing the approved curricular learning outcomes.

- Piloting or trialling new developments if feasibility is questionable or if practical experience of delivery is required.

54 Organisations should gather evidence that equality and diversity issues have been properly considered and have influenced the outcome. Organisations and institutions need to be able to demonstrate that, wherever it was relevant to do so, they considered equality and diversity issues pertinent to the work they undertook (eg through data collection, impact assessments and equality analyses).

Part 2: Managing programmes of assessment

Evidencing content quality: setting up structures to ensure the quality of assessment

Excellence by design: CR2.5a–f, CR4.1g, CR4.5, CR4.6

This section discusses establishing quality management and improvement structures and processes to collect and analyse performance data against all the evidence domains described in the model of validity theory. Specifically it discusses in detail:

- elements of the second domain of evidence in validity theory, which is concerned with the conduct and administration of assessments, including supporting information to all those involved
- elements of the fourth domain of evidence in validity theory, which is concerned with the relationships between assessments testing similar things.

Key issues in this section

Ref	Key issues for consideration
Quality structures	
CR2.5 a-f	How are stakeholders, including learners, patients and the public, involved in the oversight of assessment?
CR4.5	<p>How is the programme of assessment reviewed and continuously improved?</p> <p>How do structures make sure the review and management of assessment is carried out with appropriate input from or links to the curriculum?</p> <p>How do structures make sure curricula changes are reflected in assessment approaches and the wider programme of assessment?</p> <p>How do you make sure the quality of assessments and items produced? What checks and review of assessments are made?</p>
CR4.6	<p>Does comparing assessments with other assessments that purport to measure similar things assure you that assessments are measuring what they intend to measure?</p> <p>What does this tell you about the validity and balance of assessments within the programme of assessment?</p>

Ref	Key issues for consideration
Managing exams and assessments in the learning environment	
CR4.1g CR4.5	<p>How do you quality control all stages of the production of assessments and assessment items, and review and manage the quality of exams?</p> <p>How do you check the marking or moderation of the exam for accuracy and quality?</p> <p>What safeguards make sure responses are accurate?</p> <p>How do you systematically analyse learner responses to review and check assessment question phrasing and brief to learners?</p>
CR4.1g CR4.5	<p>How do you support the local conduct of assessments or organisations that carry out assessments locally?</p> <p>What training, guidance, rules, decision aids or other resources do you give to support deaneries/HEE local offices and assessors to conduct assessments, understand the expected level of performance or make decisions?</p> <p>What information or processes can you use to support the identification of issues and improvement of local practice?</p> <p>How do you identify where design changes and additional support or information are necessary?</p>
Ref	Considerations for approvals or quality assurance
CR4.1g	<ul style="list-style-type: none"> ■ Evidence of quality assurance infrastructure, with processes drawing on appropriate expertise to identify and manage issues and resolve these appropriately
CR4.5	<ul style="list-style-type: none"> ■ Systematic monitoring of assessment performance metrics including: reliability coefficients, standard errors of measurement (SEMs); pass rates; examiner marking behaviour; ARCP outcomes with regard to assessment outcomes ■ Periodic review of guidance material in the public domain; feedback from relevant stakeholder groups regarding need for necessary updates
CR4.6	<ul style="list-style-type: none"> ■ Regular reporting through annual specialty return and the GMC quality assurance framework

Guidance

Quality structures

- 55** Organisations must subject their programme of assessment to systems and processes that continually monitor and improve the validity (including reliability) of their assessments. Organisations must have processes to quality control all stages of assessments they deliver themselves (such as national knowledge based exams) and specific guidance and expectations to support those conducted by others.
- 56** A common issue across medical education is difficulty collecting equality and diversity data about learners. Given the importance of collecting equality data as part of the management of assessment, organisations may wish to consider what information they can provide (in addition to legal notices) to explain to learners the importance of equality and diversity data in managing the programme of assessment, citing examples of its use where appropriate.

Managing the ongoing quality of assessment designed and delivered by organisations (usually exams)

- 57** The validity of assessments depends on their practical conduct and management as well as their design. Assessments designed and delivered entirely by organisations, such as national exams, require appropriate quality management at all stages. Organisations:
- are obliged to carry out quality control of the scoring and judgements of examiners
 - must make sure assessments have been scored and reported accurately and fairly
 - must have processes to check for and identify errors in administration
 - should take steps to prevent and detect instances of potential or actual malpractice.
- 58** Software packages to process assessment data are widely available and can help to minimise clerical and processing errors; we encourage organisations to use technology to manage assessment data for this reason. Staff should have sufficient expertise to use this software appropriately and be able to identify where errors in the use of, or calculations made by, this software have occurred.
- 59** Organisations should also:
- publish information about the performance of their exams (see appendix 1 for a suggested publication scheme) (CR4.8)

- routinely analyse (and publish) them for trends related to protected (and other relevant characteristics) (see appendix 1 for suggested analysis)
 - use information from routine analysis to continuously improve their assessments; an example may include conducting item analysis to improve reliability of exams (eg Auewarakul et al 2005)
 - investigate anomalies and act to address risks to fairness or safety identified
 - report any such concerns to the regulator if they are serious or systemic.
- 60** Bodies within the organisations should routinely review and report on the overall patterns shown in assessment and consider what issues these raise. This should include analysis of results for equality and diversity issues (Coombes et al 2016), and the identification of performance differentials between different groups.

Managing the ongoing quality of assessment – conducted in the learning environment

- 61** Assessments conducted in the learning environment, particularly formative or developmental assessments, need an environment and culture that values and supports education and training. The quality and conduct of these assessments are the responsibility of organisations providing training. These organisations are subject to our *Promoting excellence* standards. Monitoring and maintaining the quality of these assessments, when conducted, should be an explicit consideration in their design and development (see *Promoting excellence* standard 1.2). Organisations and deaneries and HEE local offices need to work together to make sure curricular and professional standards are maintained. This may include gathering information to:
- take steps to provide or improve guidance for those using and providing assessments
 - engage with those using and providing assessments
 - improve the design of assessments conducted in the workplace
 - support deaneries' and HEE local offices' quality management, helping them in continuously improving their conduct of assessment (where appropriate and possible).

Managing the programme of assessment as a whole

- 62** Collating the results of different assessments together enables integrated judgements about a trainee's overall performance. Organisations should:
- where possible, provide evidence that assessments are valid by reference to other assessments that assess similar things

- optimise and minimise the assessment load by establishing the relationships between assessments and eliminating unnecessary repetition of testing, balancing this with the need to triangulate judgements
- where possible, make this information available to learners and supervisors so that a longitudinal view of an individual's strengths and weaknesses can be formed.

63 A benefit of using a programme of assessment is that a balanced view can be taken of the need for testing across the length of training. Evidence of validity of individual assessments can also be collected through looking at the relationships between results of different assessments that aim to test similar things (HPAC 2016); for example showing 'a strong positive correlation with some other measure of the same achievement or ability' and no/negative correlation with assessments of different outcomes (Downing 2003:835). High level outcomes are likely to be demonstrated by a range of assessments over time, so some degree of congruence between different assessments of similar or the same outcomes may be useful to show that the outcome has been demonstrated by the trainee.

Setting out expectations: learners

Excellence by design: CR2.4, CR4.1g, CR4.2, CR4.4

This section discusses elements of the second domain of evidence in validity theory, which is concerned with the conduct and administration of assessments, including supporting information to all those involved.

Key issues in this section

Ref	Key issues for consideration
<p>CR4.1g</p> <p>CR4.2</p>	<p>How do you make sure educators and learners have enough information about assessment to respond to it in the way intended in the design?</p> <p>How are learners informed about the role of formative or developmental assessments in the programme of assessment?</p> <p>What information is given to learners about the intended purpose of specific assessments?</p> <p>How can communication to learners around what is expected be improved?</p> <p>How are learners familiarised with assessment formats?</p> <p>Is the grading or mark sheet (or a variation of it) or expected level of performance shared with learners prior to the completion of the assessment? If not, why not?</p> <p>How are learners informed about the feedback and development they can expect to receive from all parts of the programme of assessment?</p>
<p>CR4.4</p>	<p>What kind of support do you offer to different groups of learners (eg international graduates, those in need of reasonable adjustments)?</p> <p>Do you mandate, support or suggest particular actions for those delivering assessments on your behalf? Do you provide clear guidance material for assessors?</p> <p>How are reasonable adjustments and appeals dealt with? What is the process for handling complaints?</p> <p>How is this information communicated to learners?</p>

Ref	Considerations for approvals or quality assurance
CR4.1g	<ul style="list-style-type: none"> ■ Evidence of quality assurance infrastructure, with structures and processes drawing on appropriate expertise to identify and manage issues and resolve these appropriately. Presentation of this infrastructure to all stakeholders to offer appropriate transparency of processes, including the ability to appeal outcomes ■ Clear identification of assessment context, content (syllabus) and standard against which candidates will be assessed, with reference to the relevant critical progression point where applicable ■ Sample questions for each format available in the public domain
CR4.2	<ul style="list-style-type: none"> ■ Clear information for trainees regarding sources of guidance and support where difficulties are encountered in passing relevant assessments
CR2.4	<ul style="list-style-type: none"> ■ Clear information regarding the overall assessment system and its constituent parts available in the public domain
CR4.4	<ul style="list-style-type: none"> ■ Reasonable adjustments policy in the public domain

Guidance

64 An effective programme of assessment depends on learners being aware of what is expected from them. Learners should have information to help them understand:

- how they can use the programme of assessment to drive and plan their own learning and development
- what feedback they can expect to receive from their assessments
- why they are being assessed
- what skills, knowledge, behaviours and capabilities they are expected to develop and demonstrate to satisfactorily complete training
- the relationship between assessments
- what are the critical progression points and expected levels of performance at different phases of training
- what assessments are summative or enable progression and which are formative or developmental, and how critical progression decisions are made

- the processes for appeals in summative exams, reasonable adjustments and similar should be clear and transparent.
- 65** The responsibility for communicating about these issues with learners is shared with deaneries/HEE local offices, as part of their responsibility under *Promoting excellence*. Organisations will need to make resources available to communicate the purpose, format, rules and decision making process to learners depending on the individual assessment and how it impacts upon the programme of assessment. Organisations designing assessments at work should work with those putting them into practice to give all relevant information to learners.

Setting out expectations: examiners and assessors

Excellence by design: CR4.1g, CR4.4, CR4.9, CR4.10, CR4.11, CR4.12, CR4.13

This section discusses elements of the second domain of evidence in validity theory, which is concerned with the conduct and administration of assessments, including supporting information to all those involved.

Key issues in this section

Ref	Key issues for consideration
CR4.1g CR4.9	<p>What are the organisation's expectations for assessors and examiners in each assessment?</p> <p>What information is given to assessors and examiners about assessments they work within?</p> <p>How can communication with assessors and examiners around what is expected be improved?</p> <p>How is understanding of the standards required developed across all assessors and examiners involved? (HPAC 2B.5)</p>
	<p>Who can act as an assessor or examiner in what circumstances?</p> <p>What are the particular requirements for particular roles? Where is this set out? Which of these roles requires formal selection and how is this done?</p>
	<p>Which roles require particular training or experience, and why? What is the content of the training?</p> <p>How is this training managed, organised and delivered?</p>
CR4.4 CR4.10 CR4.11 CR4.12 CR4.13	<p>How are assessors or examiners able to use their judgement and experience in decision making and providing feedback?</p> <p>Do they understand their equality and diversity responsibilities?</p> <p>What monitoring, appraisal, feedback and support do your assessors receive?</p>

Ref	Considerations for approvals or quality assurance
CR4.1g	<ul style="list-style-type: none"> ■ Appropriate guidance provided to examiners about their role within assessment material in addition to provision through briefing and subsequent calibration processes
CR4.13	<ul style="list-style-type: none"> ■ Equality and diversity training requirement for examiners, and, where appropriate, other assessors, outline content of training
CR4.10	<ul style="list-style-type: none"> ■ Role description and person specification for assessors and examiners respectively ■ Selection policy available for the appointment of new examiners ■ Outline content of examiner training programme ■ Clear communication of annual time commitment expected of examiners
CR4.11	<ul style="list-style-type: none"> ■ Calibration of examiners integrated into routine examination schedule ■ Calibration includes material-specific discussion where a station is marked by more than one examiner, in addition to overview of expected standard, with reference to the role played by summative assessment within the training programme, undertaken with all examiners prior to each examination diet ■ Feedback provided to examiners in the form of peer observation or marking data
CR4.12	<ul style="list-style-type: none"> ■ Guidance for examiners included in marking scheme or, where global judgements are applied to generic domains, station-specific guidance included in the assessment material to inform judgements
CR4.13	<ul style="list-style-type: none"> ■ The need for initial training (and subsequent training, where applicable) stipulated and includes equality and diversity training

Guidance

66 Organisations are already required to set out requirements for those who work directly for them in their exams (examiners), and should take steps to support the conduct of assessment in the workplace through setting out clear requirements and providing support, guidance, training and resources as appropriate. These should be provided for every stage in which assessors or examiners are involved, including in item writing, standard setting and the conduct of specific assessments.

Professional judgements

67 Assessment literature sees professional judgement of appropriately trained, expert assessors as a key aspect of the validity of assessment and a defensible way of forming global judgements of professional performance (HPAC 2016, Street 2015). Assessment tools are available that can capture and integrate a range of professional judgements from different groups including colleagues, supervisors and non-medical staff. HPAC (2016) advises that expert judgement can be applied to decisions about levels of supervision or entrustment and that the concept of trust can be helpful in supporting assessor decisions and feedback. Methods that can give this evidence include:

- multi-source feedback
- entrustable professional activities and other assessments at work
- supervision reports.

Supporting professional judgement in examinations and assessments at work

68 Research on differential attainment has found that interpersonal interactions and local context can potentially put some groups at a disadvantage (Regan de Beere et al 2015, Woolf et al 2016). Professional judgements are made in this context so it is important that assessments take place fairly and reliably. Organisations can support this by:

- the design of assessments, specifically:
 - designing assessment programmes that collate multiple judgements and assessors when using professional judgements
 - setting out clearly what criteria professionals should judge against
 - using formats that require or encourage assessors to record evidence and provide reasons for their judgements where appropriate.
- the choice of assessors, specifically:
 - defining what professional expertise is needed for in each assessment and when or if particular training is required. This should not unduly restrict the range of assessors that can be used, but where particular professional qualifications, experience, credibility or training is necessary, this should be clear
 - encouraging diversity amongst decision makers (Woolf et al 2016); for example attempting to recruit a diverse cohort of examiners and standard setters

- considering how patient feedback can be used eg:
 - giving opportunities for learners to obtain formative feedback from patients
 - considering the evidence base for using (simulated) patients in summative assessments practical examinations.
- training and supporting assessors, including:
 - providing information and training for assessors where appropriate, including their responsibilities in safeguarding patients and the public, and refreshing training periodically
 - providing resources such as decision aids to those who make such judgements (eg considering when and what training or guidance is needed) to make sure they understand their role and how to keep their judgements fair
 - increasing trainers' and assessors' understanding of the barriers faced by specific groups of doctors: for example, Woolf et al (2016) found that, while all doctors in training faced challenges, those from UK minority backgrounds or who trained overseas were vulnerable in particular ways that could result in poorer outcomes for these groups – such as in poorer perceptions by trainers and lack of opportunities to demonstrate outcomes
 - providing, where possible and appropriate, training in the consistent application of the mark criteria and standards, and providing regular calibration opportunities for high stakes tests.
- monitoring and appraising their performance, including:
 - producing information about their examiners' performance (eg leniency vs harshness)
 - acting to remediate or remove their assessors who consistently fail to assess candidates in line with assessment rules.

69 For assessors they manage directly (usually their examiners), guidance on the recruitment and management of these assessors is already set out by [the AoMRC](#).

Assessors working within deaneries/HEE local offices/delivering assessments at work

70 Postgraduate deaneries and HEE local offices make sure that educators have the necessary knowledge and skills, support and resources they need for their role. Organisations do not have any specific obligations for the quality or appraisal of these assessors. But they should:

- set out clear expectations for assessors, including who is an appropriate assessor in a particular task and what professional expertise, experience, credibility or training (if any) is required
- provide resources to support deaneries and HEE local teams to fulfil their role in ensuring the fair and correct conduct of assessment – this may include providing guidance or training
- use local networks, where possible, to support assessors (eg college tutors)
- support local quality management of assessors. Examples of this support might include feeding the results of portfolio audits to deaneries/HEE local teams or providing targeted training to support the conduct of a particular assessment
- gather and use information on supervisor and assessor engagement with appropriate quality management and feedback processes where possible
- consider providing information to support the educational aspects of appraisal by the deanery/HEE local team where possible and appropriate; eg enabling supervisors to review the feedback they have given to learners in their appraisal where possible.

Evidencing decisions about assessment structure: standard setting

Excellence by design: CR4.1f, CR4.11, CR4.12

This section discusses elements of the third domain of evidence in validity theory, which is concerned with the internal structure of assessments.

Key issues in this section

Ref	Key issues for consideration
<p>CR4.1f</p>	<p>How do you use standard setting to ensure safety, fairness to learners and promote excellent patient care? How does this provide assurance to stakeholders?</p> <p>What standard setting method do you use to determine the passing standard? Why do you use this method?</p> <p>What do you do to manage uncertainty around borderline candidates? Why?</p> <p>How do you ensure consistency in the standard between diets?</p> <p>What arrangements exist for review of the standard itself and the standard setting process? How are the range of stakeholders, including patients and the public and learners themselves, involved in this?</p>
<p>CR4.11 CR4.12</p>	<p>What experience do you need to set standards?</p> <p>How are standard setters recruited, trained and managed?</p>

Ref	Considerations for approvals or quality assurance
CR4.1f	<ul style="list-style-type: none"> ■ Clear identification of standard against which trainees are being assessment, evident in syllabus and with reference to relevant critical progression point within training programme ■ Criterion-referenced approach in the standard setting of assessments; compromise methods used in addition to this by way of triangulation but not as principal method to identify pass marks ■ Details of standard setting approach used for each individual component of an assessment. If test equating is applied to standard set assessments, indication of review process and frequency of periodical revisiting of question material ■ Details of training for standard setters and calibration measures prior to each exercise ■ Details of whether a compensatory or conjunctive approach is taken for each element of the assessment to inform the overall pass/fail status ■ Details of application of SEM in deciding upon the final pass mark of an assessment
CR4.11	<ul style="list-style-type: none"> ■ Role description and person specification (standard setters)
CR4.12	<ul style="list-style-type: none"> ■ The need for initial training (and subsequent training, where applicable) stipulated and includes equality and diversity training (standard setters)

Guidance

- 71** Standard setting should reference the purpose of the assessment in explicitly considering what the consequence of passing the assessment will be in providing assurance about the safety of training, professional practice, patients and the public.
- 72** Decisions with significant consequences (eg GMC specialist registration) must not use norm-referenced standards, by which passing or failing learners are defined relative to the performance of other learners. Aside from this, organisations should select the most appropriate method to ensure professional standards and fairness to learners are maintained. The [AoMRC notes](#) the importance of considering the purpose of an assessment in deciding the standard setting method.
- 73** Organisations should attempt to make sure the identified standard is maintained with each diet of a summative exam (unless there is a reason to modify the standard). This means making sure that, through a criterion-referenced approach, the chosen standard setting technique enables the identified standard to be applied consistently to assessment material in each examination diet. Fluctuations in pass marks or pass

rates and data gathered from statistical analysis might for example prompt consideration of examiner behaviour, standard setting or item quality.

- 74** Organisations need to show that an appropriate standard has been set to pass a summative assessment. They also need to:
- make sure the chosen approach to standard setting is suitable for the format of the examination and put into practice in a way that follows evidence and best practice. A typology and explanation of methods is set out in guidance from the AoMRC (2015:3,11–18)
 - make sure the standard setting process is informed by the expected scope of practice and level of performance of doctors training in programmes leading to the award of a UK CCT only
 - establish how borderline candidates are to be treated*
 - consider guidance from the [AoMRC \(2015\)](#).
- 75** Where pass rates are unstable, low or otherwise of cause for concern, organisations should investigate to determine whether this is caused by defects in the assessment itself, or whether this stems from other causes (eg changes in the performance of the learners taking the test). They should take action where this analysis shows the quality of the assessment is an issue. Organisations should consider:
- comparison of each assessment diet with others
 - reviewing assessment material
 - reviewing standard setting approaches
 - looking over a period of years to identify pass rate trends and investigating possible factors accordingly
 - reviewing the curriculum and the linking of the curriculum and learning to assessment.
- 76** Organisations should periodically review both the standard and the standard setting process to ensure the standard set and methodology around it continues to be appropriate.

* Advice on the choices available about the calculation of SEM when it is used to adjust pass-marks can be found in McManus (2012).

Evidence about assessment structure: statistical analysis

Excellence by design: CR4.1a, CR4.5, CR4.6, CR4.7

This section discusses elements of the third domain of evidence in validity theory, which is concerned with the internal structure of assessments.

Key issues in this section

Ref	Key issues for consideration
CR4.1a CR4.5 CR4.6 CR4.7	<p>How do statistics show your exams or assessments are of good quality and set appropriately to consistent standards?</p> <p>How can/do you use this information to strengthen the validity (including the reliability) of your assessments?</p> <p>What psychometric analyses do you perform on your test or test response data?</p> <ul style="list-style-type: none"> ■ Item level data (eg discrimination) ■ Test level data (eg reliability, SEM, generalisability) <p>How does this show the test is valid?</p> <p>What's your approach to ensuring reliability in exams where cohorts are too small to calculate reliability or SEM?</p> <p>If an assessment is too small to make reliability calculations, what steps have you followed to ensure reliability?</p>
CR4.6	<p>What does assessment data say about the performance of different groups of learners?</p> <p>What is your review process for checking quality of items using the data from the test? How are poorly performing items managed?</p> <p>How is data used to set standards for, or trigger review of items or whole assessments?</p> <p>Can you identify or quantify the main sources of error in assessments?</p> <p>Who reviews and interprets this data, and what actions are taken as a result?</p> <p>Which correlations with other relevant variables such as other in-course assessments or other summative assessment do you investigate?</p> <p>Is the correlation with similar tests, or dissimilar tests?</p>

Ref	Considerations for approvals or quality assurance
CR4.1a	<ul style="list-style-type: none"> ■ Clear articulation of how the question sampling (number of questions included in each element of an assessment) fulfils requirements of appropriate syllabus coverage and reliability (internal consistency)
CR4.5	<ul style="list-style-type: none"> ■ Systematic monitoring of assessment performance metrics including: reliability coefficients, SEMs, pass rates, examiner marking behaviour, ARCP outcomes with regard to assessment outcomes
CR4.6 (CR4.7)	<ul style="list-style-type: none"> ■ Report of examination quality, considering and explaining key sources of validity evidence including psychometric properties of individual elements of assessments (item and test level metrics, especially reliability and SEM) and contribution to the programme as a whole including (eg through congruence with similar or related tests)

Guidance

77 An important source of evidence about the validity of an assessment (or set of assessments) is its internal structure, ie psychometric properties (Sullivan 2011:119). These properties can help to understand the quality of the pre-test planning, design and the quality of the assessment's conduct. It can help to identify key quality concerns and provide evidence about whether decisions are fair and defensible, and provide information that enables the validity of assessment approaches to be strengthened. Psychometric evidence can also help to identify and investigate questions about fairness and variations between different groups.

78 This information cannot be produced or used without a critical understanding of the different measures. Organisations need to:

- make sure they have secured sufficient access to expertise to analyse and understand the data, and to act where it shows action is required
- carry out psychometric review and investigation into assessments at the level of:
 - items within exams. Examples of metrics that are likely to be appropriate for this purpose and which organisations should consider using as appropriate include:
 - item/test characteristic curves (ICCs/TCCs)
 - inter-item correlations
 - item-total correlations.

- the properties of tests as a whole. In particular, reliability is an important part of the demonstration of overall validity (Downing 2004). Examples of metrics likely to be appropriate for this purpose and which organisations should consider using as appropriate include:
 - the reliability
 - generalisability
 - SEM
 - item factor analysis
 - differential item functioning.
- the programme as a whole (eg through the congruence of similar tests or analysis of the outcomes of those completing training).

79 Psychometric analysis should take a broad view of the quality of assessments and aim to produce information that enables the quality of all aspects of assessment to be understood and continuously improved. The exact choice of metrics may vary with the test in question but the range of metrics should be justified in reporting about tests.

Assessments at work and reliability

80 The purpose of assessments at work will generally be formative supervised learning events, where feedback and engagement in the learning process is key; we do not require these assessments to meet reliability criteria and caution it may be undesirable to try and reduce the rich information these formats can give to something that can be demonstrably reliable (see HPAC 2016:37,44).

81 If the purpose of assessment in the workplace is summative, then judgements about knowledge, skills or performance need to be made reliably. Organisations may wish to consider:

- using tools or formats shown to be reliable elsewhere (provided they are transferable)
- making use of expert judgement, and doing so over multiple assessors and occasions (AoMRC 2016:4). This can be:
 - within the individual assessment (with some research supporting the psychometric reliability of an MSF (Moonen-van Loon et al 2015)) or
 - as part of a programmatic approach in which a summative process collates and synthesises formative elements (see examples cited above).

- training assessors in their roles and declaring expected standards.

- 82** HPAC (2016:44) also suggest that entrustment formats may have advantages in terms of authenticity and rigour. Furthermore asking assessors to entrust responsibilities for patient care may link the assessor's judgement more closely to their own duties to uphold standards. This is important because research indicates such considerations can enable assessors, who might otherwise feel unwilling or unable, to fail underperforming or unsafe learners (Yepes-Rios et al 2016).
- 83** We do not mandate a particular approach but organisations should show that, where assessments at work are used for summative purposes, the judgement produced will be defensible, and supported by consideration of reliability and fairness issues.

Small cohorts

- 84** Some cohorts of learners may be too small to produce meaningful statistics about reliability in summative assessments. Organisations still need to design and deliver assessments that have overall validity for their intended purpose, including appropriate reliability, even if this cannot be demonstrated as a coefficient. Organisations can address some of these difficulties by:
- using tools or formats shown to be reliable elsewhere (provided they are transferable)
 - carrying out assessment design, conduct and quality management to a high standard
 - trying to achieve reliable overall results by appropriate combinations and correlations of numerous assessments taken over time with different assessors
 - comparing or correlating results with tests assessing similar things and which are known to be reliable (concurrent or predictive validity).
- 85** Approaches that use assessments at work in combination to make judgements are acceptable, as long as a programme of assessment was appropriately designed.

Part 3: The impact of a decision

Evidencing the impact of assessments

Excellence by design: CR4.2, CR4.3, CR4.4, CR4.5, CR4.6

This section discusses elements of the fifth domain of evidence in validity theory, which is concerned with the consequences of assessment decisions for all those involved.

Key issues in this section

Ref	Key issues for consideration
CR4.3 CR4.5	<p>How can the evidence about the design, delivery and analysis of assessment be summarised to justify why the assessment has strong validity? How can it be improved against all the domains of validity theory? Where is this reported and set out?</p> <p>How do you manage the impact of failing on learners and feed back to employers and training providers?</p> <p>What arrangements are there for remediation and support?</p> <p>How is potential or actual malpractice by learners managed?</p>
CR4.2 CR4.4	<p>How do decisions ensure fairness to learners, patient safety and support learners to achieve excellence?</p> <p>What kind of feedback (score reporting and qualitative information) is given on the assessment?</p> <p>How does the programme of assessment enable the identification and management of learners who are not (yet) safe to practice at critical progressions points or at the point of completion of training?</p>
CR4.5 CR4.6	<p>What monitoring of assessment outcomes do you carry out?</p> <p>What impact do the results of the assessment have on:</p> <ul style="list-style-type: none"> ■ curricular outcomes ■ the programme of learning ■ the future design and conduct of assessments? <p>What external review is undertaken of the programme of assessment?</p> <p>What impact do the results of the assessment have on the design and development of the curriculum and assessments and the programme of assessment in general? How can data about assessments help you improve quality?</p>

Ref	Considerations for approvals or quality assurance
CR4.2	<ul style="list-style-type: none"> ■ Process in place for assessors to identify (and manage) potentially dangerous learners
CR4.3	<ul style="list-style-type: none"> ■ Systematic approach to identifying each area required prior to progression at critical progression points of the training programme (documented and highlighted in a matrix or overarching blueprint for the programme of assessment as a whole)
CR4.4	<ul style="list-style-type: none"> ■ Conscious action to make sure assessment decisions and decision aids are made without bias
CR4.5	<ul style="list-style-type: none"> ■ Systematic monitoring of assessment performance metrics including: reliability coefficients, SEMs, pass rates, examiner marking behaviour, ARCP outcomes with regard to assessment outcomes ■ Periodic review of guidance material in the public domain; feedback from relevant stakeholder groups regarding need for necessary updates
CR4.6	<ul style="list-style-type: none"> ■ Regular reporting through annual specialty return on assessment quality

Guidance

- 86** The successful completion of many postgraduate training programmes is linked to the ability to practise as a consultant or GP in the NHS – so concerns about validity of assessment have their roots in the concerns of the wider public (Kane 2013:2). Organisations need to make sure decisions about progression and actions in respect of learners who do not (yet) meet standards reflect this. Guidance on the maximum number of attempts at examinations is provided in [our position statement on this issue](#).
- 87** Organisations should recognise that assessments support processes in the wider healthcare and training system. Organisations and deaneries or HEE local offices should work collaboratively to ensure and provide assurance about the quality of learners completing the programme by providing evidence against the domains of validity (see above 4). Organisations should consider how deaneries or HEE local offices would need to use assessment data, eg in the planning and management of service and training. They should make sure they provide data in a sufficiently timely way to enable deaneries/HEE local offices to use the information to plan learners' progression.

Assuring patients, learners and others

88 Groups who have an interest in assessment decisions include:

- patients and the public, employers and colleagues, all of whom can expect a certain standard of performance from a doctor; patient safety is the first priority
- the doctor, who can expect to be treated fairly, assessed objectively and to have reasonable opportunities to remediate and develop in areas of weakness
- the organisation, along with relevant local training organisation, who are accountable for conducting assessments and making key decisions in the form of critical progression and satisfactory completion. They are also responsible for deciding what is taught and, to some extent, how. Organisations should:
 - use data to continuously improve the quality and performance of their assessments
 - use data to support curricula and outcomes review.

89 Organisations should report on how the programme of assessment has provided appropriate assurance about those successfully completing training, and how its validity can be strengthened against across all the areas of evidence identified by validity theory:

- the content and design or format of the assessment
- the conduct of the assessment, including the information and training provided to all stakeholders about the assessment
- psychometric evidence and other aspects of the internal structure of individual assessments
- the relationships between the different assessments within the programme
- whether decisions made about progression in and completion of training are supported by satisfactory evidence to protect patients and ensure fairness to learners.

90 Establishing formal mechanisms with external organisations and peers to carry out periodic reviews of their programmes of assessment may help to improve quality and share practice.

Appendix 1: Annual publication of exam data

CR4.7 of our standards requires organisations to publish the quality performance metrics of high-stakes summative or progression assessments to promote transparency and openness.

We suggest the following as a minimum set of information or a template for publication about individual major national exams which provide assessment against approved curricula.

Exam name	
Exam format	Please describe the type of assessment, type and number of items
Number of candidates* and pass rates	Please state the number attempting the exam in year or each diet within year. We recommend reporting those in UK training as a distinct group (see footnote).
Breakdown of candidates and passing candidates by: <ul style="list-style-type: none"> ■ demographic groups/protected characteristics ■ place of qualification ■ attempt number 	

* We recommend colleges report data on the basis of those candidates to whom the whole programme of assessment to which a test is part of applies; for example where an examination forms part of a programme leading (eventually) to specialist/GP registration, we recommend reporting explicitly on those candidates who are in UK training and might be expected to eventually achieve entry to the specialist/GP registers (and not those pursuing the exam as an independent qualification or for purposes overseas).

<p>Standard setting method</p>	<p>Please describe:</p> <ul style="list-style-type: none"> ■ the standard setting method and reasons for the choice or reference to where this is set out in other documents eg approvals documentation ■ the frequency with which standard setting is carried out ■ additional measures to ensure the safety of the standard set, eg minimum numbers of stations to pass or adjustments for error.
<p>Most recent report to oversight body</p>	<p>A report describing in a comprehensive and holistic manner:</p> <ul style="list-style-type: none"> ■ the quality of the assessment using a range of metrics supported by appropriate explanation and interpretation <ul style="list-style-type: none"> ■ this should include discussion of the reliability/internal consistency of the assessment for those in UK training using an appropriate choice of metrics ■ interpretations of this information including discussion of quality management activity which is required or desirable in respect of the assessment ■ any other action appropriate action in respect of the wider functions of the organisation setting the assessment ■ the quality activity undertaken in respect of the examination since the last report. <p>We suggest reporting on a regular (annual) basis, to considering the diets of the assessment carried out that year.</p>

Key terms used in this guidance

Term	Meaning
Assessment	We define assessment as all activity aimed at judging a learner's attainment of curriculum outcomes, whether for summative purposes (determining progress or completion) or formative purposes (giving feedback). An examination is an example of an individual assessment test.
Assessor	An assessor provides an assessment and is responsible for interpreting the learner's performance in that assessment. Assessors should be appropriately trained and should normally be competent (preferably expert) in the area that is being assessed. It includes examiners as a specific type of assessor.
Critical progression point	A point in a curriculum where a learner transitions to higher levels of professional responsibility or enters a new or specialist area of practice or experiences significant changes in the level of supervision or trust. Satisfactory completion of training is a critical progression point.
Examiner	An examiner is a category of assessor working within the context of a formal, summative exam.
Learners	Learners are medical students receiving education leading to a primary medical qualification and doctors in postgraduate training leading to a certificate of completion of training (CCT) or doctors completing a regulated credential.
Learning outcomes	An outcome can be defined as a level of performance or behaviour that a trainee is expected to achieve as part of their development according to their stage of training within their specialty curriculum. This can include an area of professional practice that may be trusted to a learner to execute unsupervised, once he or she has demonstrated the required competence.
Organisation	In this guidance, organisation refers to a body, expected to be a college or faculty (or a combination of colleges and faculties), with responsibility for design and maintenance of an approved curriculum and programme of assessment or a part of it. It does not include HEE local offices or deaneries which may have responsibilities for the quality of the conduct of some assessments locally.

Bibliography

- Academy of Medical Royal Colleges (AOMRC) (2016), *Improving Assessment: Further Guidance and Recommendations*, <http://www.aomrc.org.uk/publications/reports-guidance/improving-assessment-guidance-recommendations/>
- AoMRC (2015), *Guidance for standard setting: A framework for high stakes postgraduate competency-based examinations*, http://www.aomrc.org.uk/wp-content/uploads/2016/05/Standard_setting_framework_postgrad_exams_1015.pdf
- AoMRC (Forthcoming 2017), *Improving Feedback and Reflection to Improve Learning: a practical guide for trainees and trainers*, <http://www.aomrc.org.uk/all-publications/reports-guidance/>
- Auewarakul C, Downing SM, Praditsuwan R, Jaturatamrong U (2005) Item Analysis to Improve Reliability for an Internal Medicine Undergraduate OSCE. *Advances in Health Sciences Education*. 10(2), 105–113
- Coderre S, Woloschuk W, McLaughlin K (2009) Twelve tips for blueprinting. *Medical Teacher*. 31(4), 322–324
- Coombes, L, Roberts M, Zahra D, Burr S (2016) Twelve tips for assessment psychometrics. *Medical Teacher*, 38(3), 250–254
- Downing, SM (2003) 'Validity: on the meaningful interpretation of assessment data' *Medical Education* 2003;37:830–837
- Downing, SM (2004) 'Reliability: on the reproducibility of assessment data' *Medical Education*, Vol. 38 Issue 9, September 2004 Pages 915–1018
- GMC (2011) *Assessment in undergraduate medical education Advice supplementary to Tomorrow's Doctors (2009)*, GMC, London http://www.gmc-uk.org/Assessment_in_undergraduate_medical_education_guidance_0815.pdf 56439668.pdf
- Harden, RM , Crosby, JR, Davis, MH (1999) 'An introduction to outcome-based education' in ed Lilley, P, *Outcome-based education AMEE Medical Education Guide No 14*, Dundee, 1999 [https://amee.org/getattachment/AMEE-Initiatives/ESME-Courses/AMEE-ESME-Face-to-Face-Courses/ESME/ESME-Online-Resources-China-Dec-2015-\(1\)/Session-1-Guide-14-Outcome-based-education.pdf](https://amee.org/getattachment/AMEE-Initiatives/ESME-Courses/AMEE-ESME-Face-to-Face-Courses/ESME/ESME-Online-Resources-China-Dec-2015-(1)/Session-1-Guide-14-Outcome-based-education.pdf)
- Health Professional Assessment Consultancy (HPAC), (2016), Final Report for the provision of identifying key principles for consistency and reliability in curricula and assessment frameworks Prepared for the General Medical Council (GMC331), http://www.gmc-uk.org/Key_principles_for_curricula_and_assessment_HPAC_final_report_Nov_2016.pdf 69345344.pdf

Kane, MT (2013) 'Validating the Interpretations and Uses of Test Scores', *Journal of Educational Measurement*, Special Issue: Validity Volume 50, Issue 1, pages 1–73, Spring 2013

[McManus, IC \(2012\)](#), 'The misinterpretation of the standard error of measurement in medical education: a primer on the problems, pitfalls and peculiarities of the three different standard errors of measurement.' *Med Teach*. 2012; 34(7):569–76.

Moonen-van Loon JM, Overeem K, Govaerts MJ, Verhoeven BH, van der Vleuten CP, Driessen EW (2015) The reliability of multisource feedback in competency-based assessment programs: the effects of multiple occasions and assessor groups, *Acad Med*. 2015 Aug; 90(8):1093-9.

Plymouth University (Undated), *Glossary of assessment terms*, www.plymouth.ac.uk/your-university/teaching-and-learning/guidance-and-resources/glossary-of-assessment-terms accessed 21 April 2017

Regan de Bere, S, Suzanne Nunn, Mona Nasser (2015), *Understanding differential attainment across medical training pathways: A rapid review of the literature Final report prepared for The General Medical Council*, GMC commissioned literature review www.gmc-uk.org/GMC_Understanding_Differential_Attainment.pdf_63533431.pdf

Roberts, C, N Shadbolt, T Clark and P Simpson (2014), 'The reliability and validity of a portfolio designed as a programmatic assessment of performance in an integrated clinical placement' *BMC Medical Education* 2014, 14:197

Schuwirth LWT & CPM Van der Vleuten (2011), Programmatic assessment: from assessment of learning to assessment for learning, *Medical Teacher*, 33:6, 478–485

Sullivan, GM (2011) 'A Primer on the Validity of Assessment Instruments' *Journal of Graduate Medical Education*, June 2011, pp119–120

Street, S (2015) 'The progressive challenge of assessing a GP trainee's skills in the workplace: a UK update' *Education for Primary care* (2015) 26 60:2

ten Cate, O, (2013), 'Nuts and Bolts of Entrustable Professional Activities' *J Grad Med Educ*. 2013 Mar; 5(1): 157–158. doi: 10.4300/JGME-D-12-00380.1

Tighe, J, IC McManus, NG Dewhurst, L Chis, J Mucklow (2010) 'The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations', *BMC Medical Education* 10:40

van der Vleuten, CPM (1996) 'The Assessment of Professional Competence: Developments, Research and Practical Implications' *Advances in Health Sciences Education* 1: 41-67, 1996

Woolf, C, Rich, A, Viney, R, Rigby, M, Needleman, S, Griffin, A (2016) *Fair Training Pathways for All: Understanding Experiences of Progression Final Report*, www.gmc-uk.org/2016_04_28_FairPathwaysFinalReport.pdf_66939685.pdf

Yepes-Rios, M, Dudek, N, Duboyce, R, Curtis, J, Allard, RJ & Varpio L (2016): The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42, *Medical Teacher*, DOI: 10.1080/0142159X.2016.1215414

Email: gmc@gmc-uk.org

Website: www.gmc-uk.org

Telephone: **0161 923 6602**

General Medical Council, 3 Hardman Street, Manchester M3 3AW

Textphone: please dial the prefix **18001** then
0161 923 6602 to use the Text Relay service

Join the conversation

 [@gmcuk](https://twitter.com/gmcuk)

 facebook.com/gmcuk

 linkd.in/gmcuk

 youtube.com/gmcuktv

To ask for this publication in Welsh, or in another format or language, please call us on **0161 923 6602** or email us at publications@gmc-uk.org.

Published May 2017

© 2017 General Medical Council

The text of this document may be reproduced free of charge in any format or medium providing it is reproduced accurately and not in a misleading context. The material must be acknowledged as GMC copyright and the document title specified.

The GMC is a charity registered in England and Wales (1089278) and Scotland (SC037750).

Printed on 100% recycled paper.

Code: GMC/DMPAP/0517

**General
Medical
Council**