Findings from Ofqual's review of apprenticeship assessment plans

Information to support Trailblazers developing assessment plans

November 2017

Ofqual/17/6313

Contents

This document has been written to support Trailblazers as they develop their assessment plans. It sets out what we have learned from our consideration of nearly 100 assessment plans to date.

Foreword

In reformed apprenticeships, employer groups known as 'Trailblazers' play a leading role in each apprenticeship Standard. They are required to select an approach for external quality assurance and Ofqual is one of the options available to them.

Where a Trailblazer asks Ofqual to provide external quality assurance, we take a keen interest in the assessment plan they produce. This assessment plan, together with the knowledge, skills and behaviours set out in the Standard, provides the basis for organisations to develop end-point assessments. This makes it very important.

We review the Trailblazer's draft assessment plan, providing advice and working with them, to ensure it will support the development of sound end-point assessments – that these will assess apprentices reliably and consistently, on the right things, at the right level, wherever and whenever they take the assessment. Once we have this assurance, we can act as the external quality assurance provider.

In other words, our advice to Trailblazers on assessment plans seeks to ensure that the end-point assessments developed should give employers what they want.

Having assurance about assessments and their comparability, including over time, is crucial. It is important in cases where a single end-point assessment organisation is involved – so that assessments are comparable over time and across locations – and it applies equally where there are a number of different end-point assessment organisations for a Standard.

We have looked at a large number of assessment plans, with a range of different assessment types, with these issues in mind – almost 100 of them in around 18 months. Our aim in this document is to set out approaches we have seen to date on key aspects of assessment design, and to share our reflections to help Trailblazers ensure their assessment plans, and the end-point assessments they lead to, are fit for purpose.

The importance of assessment expertise

1. Assessment design is a technical and complex discipline. Some of the concepts it must address, such as reliability and consistency, are relatively simple to identify; but they can also be challenging to secure, or maximise, in practice. Validity is at the heart of the assessment design process and it is also at the heart of our consideration of assessment plans. The central question we have in mind when we consider them is:

   "Will this assessment plan support the development of sufficiently valid end-point assessments?"

2. In this context, we approach validity as:

   The degree to which an assessment plan is likely to produce end-point assessments that will effectively ***measure*** *what the Trailblazer has indicated* ***needs*** *to be measured, by implementing the assessment* ***methods****.*

3. Validity is an overarching quality: it stems from ***all*** the specific features and processes for a given assessment in combination. A sufficiently valid assessment is measuring the ***right things***, in the ***right way***, to produce ***accurate*** and ***useful*** assessment results.

4. This makes validity a very important consideration in assessments – in many ways, the most important consideration of all. It is also vital to the end-users of assessments, such as employers.

5. This is because a valid assessment means that those taking it have been assessed on what was intended and is needed; that this has been done in a reliable and consistent way; and, therefore, that the results generated by the assessment can be trusted.

6. Assessment validity comes from all the stages of the assessment cycle:

   - determining the content to be taught and assessed
   - designing tasks to assess that content
   - assessing learners' responses
   - determining learners' results
   - interpreting those results.

7. This makes it important that validity is 'designed into' an assessment at the outset and then maximised at every stage of delivering it. If there are fundamental issues with the validity of an assessment, these cannot usually be rectified without redesigning it from first principles. For example:

    - If the content itself is not right – is not what users need – then it will not matter how well designed the tasks are, or how accurately they are assessed: the results will not tell users what they want to know about.

    - Alternatively, if the content and tasks do cover what users need, but the assessment criteria do not support reliable judgements, producing results will not help: users will not be able to have confidence in them.

8. Because validity is affected by many issues, at many points, it always tends to involve trade-offs; to entail striking the best balance between things, and deciding what is most important and therefore what to prioritise. For example:

    - Multiple-choice tests can be highly reliable, if they are well designed. They can cover a large amount of content, and can be marked quickly and accurately. But they may be less effective than other question types at assessing some things – such as how and why areas are linked, themes across issues, and so on. A different question type (like an essay) may be more suitable in such cases – or even a different method of assessment altogether, like a report, or a presentation.

    - As a rule, the longer an assessment is, the more evidence it will generate about attainment in what it seeks to measure (this is presuming it is well targeted). This extended length should increase its reliability. Taken too far, however, it would also undermine its manageability (for instance, in terms of cost and effects on learners and centres) – this would undermine the overall effectiveness of the assessment procedure in measuring what is needed.

9. The fundamental aim in maximising assessment validity is to determine what matters most, then to use the approach that will generate the most accurate information possible about what is to be measured, with that priority in mind. This is a complex process. The considerations and decisions involved are numerous and typically are highly interlinked.

10. As such, it is beneficial for assessment expertise to be drawn on right through the development process – such as through Trailblazers engaging with potential end-point assessment organisations, and through the technical expertise we bring to bear on assessment plans.

About this document

11. The main sections of this document consider approaches we have seen in assessment plans to date on key aspects of assessment design. These sections are written with the expectation that the Trailblazer has developed an apprenticeship Standard, which clearly sets out the knowledge, skills and behaviours deemed necessary for the occupation and level, as clarity on what is to be assessed is crucial. The four areas the sections address are:

    1) *Assessment methods*

    2) *Individual assessment grading*

    3) *Generating overall results*

    4) *Approach to re-takes and re-sits*

12. The issues involved in these aspects of assessment design are complicated, so these sections address only some of their main points, at a fairly high level. In some cases, examples are provided: these are intended to help in visualising the issues, rather than seeking to advocate specific approaches. It is important to emphasise that the issues – and particularly the interconnections between them – are largely unique to each individual assessment plan, meaning that appropriate expert input, in each specific instance, is always critical.

13. The document concludes with a suggested way of thinking through the issues involved in creating an assessment plan, reflecting on what we have seen in them to date. This is not by any means the only way to do this – the aim is to suggest areas for consideration, where this may be useful, given the complex and interrelated nature of the issues involved in assessment design.

Background

14. Each reformed apprenticeship comprises the following main parts:

1) **Standard**	• This sets out the required knowledge, skills and behaviours (KSBs) for the apprenticeship. • It is produced by an employer group, known as a 'Trailblazer'.
2) **Assessment plan**	• This specifies the form the end-point assessment (EPA) for a Standard must take (eg assessment methods and grading approaches). • It also specifies expected training within the apprenticeship and any 'gateway' assessment that must be completed before the EPA. • It is produced by the Trailblazer; the requirements take a wide range of forms, varying considerably by sector and level.
3) **End-point assessment**	• This provides an indication – *wholly at the end of the apprenticeship* – of attainment with respect to the required KSBs. • It is produced by an end-point assessment organisation; and there may be multiple EPAs, developed by different end-point assessment organisations, for a given Standard.

15. Focusing particularly on the assessment plan, this document is intended to provide a clear framework for end-point assessment organisations to use in designing EPAs, such that:

    - an apprentice's result accurately reflects their level of competence in relation to the Standard;
    - the interpretation of that result by apprentices and employers will be in line with the meaning intended by the Trailblazer;
    - the requirements of end-point assessments will be consistent, across locations and over time; and

- where relevant, different end-point assessment organisations will interpret the requirements in a consistent manner.

Section 1: Assessment methods

16. An apprenticeship Standard identifies ***what*** is to be assessed in the end-point assessment (EPA). This is the stipulated 'knowledge, skills and behaviours' – or 'KSBs' in short. The related assessment plan indicates ***how*** those KSBs are to be assessed.

17. The combination of assessment methods chosen in the assessment plan needs to ensure that all the KSBs will be assessed appropriately. It should give an apprentice the optimum opportunity to demonstrate the KSBs, and should give employers confidence that the apprentice has been assessed with methods sufficiently able to determine their occupational competence.

18. There have been 2 central aspects where this issue has been approached effectively in assessment plans we have considered: i) the Trailblazer expectations have been clear and; ii) the assessment methods have been well suited to the KSBs.

Clarity of Trailblazer expectations

19. It is important that it is clear in the assessment plan which KSBs are to be targeted in each assessment method. This helps to ensure that EPAs place consistent expectations on apprentices – including, but not only, where different end-point assessment organisations are involved. An example of a clear approach to this we have seen is:

Key to assessment method identification	
IEA	Independent End Assessment activity – identifies which assessment method will be used for that section of the standard.
T	Assessment will be through the On demand test
O	Assessment will be through the observation
PD	Assessment will be through the professional discussion
O/ PD	May be covered by observation depending on which scenario selected. If not must be covered by professional discussion

	Knowledge and Understanding (Know it)	IE A	Skills (Show it)	IE A
Safety	Health and safety legislation in aviation and in relation to own role and organisation and how to monitor it	T / PD	Monitor area of responsibility to ensure compliance with aviation legislation and organisational procedures, addressing and / or reporting hazards	O
Security	Requirements for maintaining aviation security in own area of authority and action to take in the event of a breach of security	O/T	Ensure aviation security is maintained in own area of operations e.g. challenging people in restricted areas, recording and reporting of security incidents	O / PD
Compliance & Legislation	Aviation and other applicable legislation, procedures and regulations relating to an aviation environment, and monitoring procedures within own area of responsibility	T/ PD	Monitor compliance with legislation, procedures and regulations in an aviation environment within own area of responsibility	O / PD
Communication	Methods of communication to ensure effective and timely transfer of information to different audiences using relevant language and format	O / T	Communicate effectively within the aviation environment adapting methods and language to meet the situation	O
Inter-personal skills	Roles within the team and how these work together to achieve the organisation's objectives, and how to embrace equality and inclusivity in the workplace	PD	Maximise team performance and meet the objectives of the team whilst embracing equality and inclusivity	O/ PD
Aviation systems	The aviation systems used within own role and how to operate them, identify faults or errors and what remedial action to take	O/ PD	Use aviation systems effectively in own role. Take remedial action upon identification of faults or errors in a timely manner if they occur	O / T

20. Although it can be acceptable to assess a KSB using more than one assessment method – as in this example – where that is the case, each method should have specific, distinct expectations, to ensure an appropriate focus and mitigate the risk of duplication.

21. As such, the clearest assessment plans we have considered have often mapped each unique KSB to a single assessment method. This has made it highly transparent how each KSB is to be assessed. It should ensure that each KSB is assessed *somewhere*, and that individual KSBs are not unintentionally assessed *multiple* times. That could lead to them having a greater influence than intended on outcomes.

22. There are also some overarching issues in relation to assessment methods that we have seen considered effectively in assessment plans. In particular, these have included:

- a clear and unambiguous indication of where the *evidence* for each KSB will be found – in terms of which assessment method, and where within it;
- all evidence being 'counted' only *once* – this is important for fairness, and particularly in cases where one assessment method 'follows up on' another;
- no 'on-programme' assessment in the end-point assessment, in line with reformed apprenticeships policy – for example, a portfolio may *support* an end-point assessment, but should not itself be assessed;
- the *duration* of each assessment being sufficient to secure reliably the evidence that is needed, but also being manageable for all those who are involved.

23. However, we have also seen instances where assessment plans have not been so clear. An assessment plan that maps every KSB to every assessment method – so that the end-point assessment organisation determines what to assess where – is problematic. It is also problematic where an apprentice's achievement of KSBs in one assessment method determines which KSBs will be assessed in a subsequent assessment method. Both of these undermine comparability – meaning apprentices may be assessed on the same content, but in very different ways.

24. For example, it might not be appropriate if an assessment plan allowed a KSB either to be demonstrated (e.g. via a practical skills test) or to be discussed (e.g. via an interview). Those different methods might individually be legitimate, depending on the nature of the KSB targeted, but they also differ a lot in what they entail and the level of assurance they give.

Suitability of assessment methods

25. It is likely that many KSBs will lend themselves to being assessed in particular ways. The assessment methods chosen should be those best suited to assessing each KSB. This is a critical step in the design process as it lays the foundation for an effective EPA.

26. Each assessment method should be one that will enable the most reliable outcomes, for the specific KSBs it will target. It is important to consider that an assessment method that does this for one KSB may not be as suited to others.

27. Determining the optimal assessment method for a KSB mainly involves considering the nature of that KSB, what its achievement 'looks like', and the degree of assurance about it that is needed or intended. For example:

    - Is it a KSB that would have to be physically demonstrated? Or might it be asked about? And if so, could that be a 'set question', or would it need a more interactive approach?

- *These types of consideration may help to indicate if a KSB might be assessed through a practical observation, or by a written test, or in an interview/discussion.*

- Is it a KSB that can be judged in a 'binary' way – someone can do it or cannot, knows it or does not? Or is it something where *degrees* of achievement are possible and could be used by assessors?

  - *These types of consideration may help to indicate if a KSB might be assessed using a list of criteria (eg met/not met), or whether some form of scoring system could be helpful.*

- Is it a KSB where absolute confirmation is needed that an apprentice has met the requirement? Or is it something where an apprentice's strengths and weaknesses might balance each other out?

  - *These types of consideration may help to indicate if a KSB might be assessed using a method that gives certainty on individual aspects, or where the emphasis is more on overall attainment.*

28. The best assessment method varies by KSB – assessing every KSB optimally therefore tends to require a range of methods. The key point in this is that ***what*** is being assessed should suit ***how*** it is being assessed. Technically, this quality is sometimes referred to as 'fidelity' – the extent to which an assessment is faithful to the real world, such as in reflecting accurately the relevant workplace requirements.

29. The table below summarises some particularly effective combinations of assessment methods and KSBs we have seen in assessment plans. It sets out the type of KSBs particular assessment methods have targeted and key design considerations that have been explained in the related assessment plan.

Assessment method	Type of KSBs	Key design considerations
Knowledge test	• Factual knowledge • Application of knowledge • Analysis and evaluation	• Number of questions per test • Balance of 'constrained' questions (such as multiple choice) and open questions • Size of question bank to mitigate test predictability • Test availability (e.g. scheduled or on-demand) • Proportion of KSBs to be sampled per test
Practical observation	• Physical demonstration of skills • Underpinning knowledge drawn on by skills • Behaviours	• Scheduling assessments • Implications in terms of location and equipment needed • Any supporting evidence to be captured (e.g. video footage for monitoring)
Portfolio	• Various – emphasis is combination it allows and extended scope possible • Can be less time bound • Can involve real examples drawn from everyday workplace, such as for behaviours	• Clarity regarding scope and evidence requirements • Need for separate interaction between assessor and apprentice to ensure material is authentic • Form(s) of supporting assessment used – as the EPA itself should not entail on-programme assessment
Project	• Various – emphasis is combination it allows and extended scope possible • Can be less time bound	• Clarity regarding scope and evidence requirements • Need for separate interaction between assessor and apprentice to ensure material is authentic

Assessment method	Type of KSBs	Key design considerations
Presentation	• Various – emphasis is combination it allows • Scope for planned and ad hoc material	• Clarity regarding scope and evidence requirements • Balance between planned and ad hoc material • Potential for anxiety to affect results • Must address KSBs, not general 'presentation skills'
Interview	• 'Soft skills' • In-depth and multi-step exploration of knowledge	• Balance between consistency and tailoring questions • Potential for anxiety to affect results • Must address KSBs, not general 'interview skills'

30. In terms of less effective practice, there have been some assessment plans we have considered where the assessment methods proposed initially have not been well suited to the KSBs to be targeted. This has been particularly evident where the proposal has been that apprentices will be asked – for example, in a professional discussion or interview – about aspects such as their behaviours, IT skills and report writing; but with no requirement for them to give evidence.

31. In such circumstances, it is likely that an apprentice's account of their own abilities would differ from an assessor's judgement of their abilities based on reviewing evidence. This would have meant the assessment, if taking the form initially proposed, would not have been very reliable. As a result, its usefulness to end-users, such as employers, would have been limited.

Section 2: Individual assessment grading

32. Each assessment method in an assessment plan brings together a number of KSBs. The grade for each assessment method needs to reflect the level of competency that an apprentice has achieved in relation to the KSBs targeted by that method.

Degree of assurance

33. A central issue in determining the pass requirements (and any other grading requirements) for an individual assessment method is the degree of assurance that is intended. It may be that ***absolute confirmation*** is intended that an apprentice has (or has not) demonstrated the KSBs targeted. Technically, this approach is often termed 'mastery'.

34. Alternatively, it may be viewed as acceptable for apprentices to demonstrate a ***balance*** of strengths and weaknesses across the various KSBs targeted – for example, if the assessment method is addressing a number of KSBs, and obtaining a general sense of attainment across them is the intention. Technically, this approach is often termed 'compensatory'.

35. This decision between absolute 'mastery' assurance and general 'compensatory' assurance can also be made within, rather than across, KSBs – in those cases where they are multi-faceted. For example, if a KSB comprised three distinct strands, it may be that all three of them would have to be demonstrated ('mastery'); alternatively that a balance across them would be acceptable ('compensatory').

36. The 2 approaches are very different:

    a) 'mastery' means ***all the KSBs*** targeted by the assessment (and sometimes all aspects within the KSBs) have to be demonstrated to achieve the grade – they are <u>all</u> what are termed 'hurdles';

    b) 'compensatory' means ***some of the KSBs*** targeted by the assessment (and sometimes some aspects within the KSBs) have to be demonstrated to achieve the grade – there are no <u>specific</u> 'hurdles'; though it will be necessary to demonstrate a satisfactory performance overall – for example, by demonstrating sufficient competency in any seven KSBs where 10 of them are assessed.

37. In a sense, these approaches represent the opposite ends of a scale. The 'mastery' approach gives greater certainty that all the individual aspects have been demonstrated, but has a potential cost in terms of overall measurement, since a single 'blip' in performance can lead to a fail even if performance

everywhere else is exemplary. By contrast, the 'compensatory' approach can potentially provide a more accurate measurement overall, but it gives less clarity that any individual aspect has been demonstrated. Which of these approaches is optimal depends on the degree of assurance that is intended.

38. It is also important to emphasise that an assessment plan can use a *combination* of 'mastery' and 'compensatory' approaches. What is intended and possible in one assessment method – owing to the KSBs it targets and the nature of the method itself – may not be viable in another assessment method within the same assessment plan. For example, a written test focused on knowledge may work effectively taking a 'compensatory' approach – if the intention is that apprentices have a strong knowledge base across the piece, rather than 'perfect' knowledge throughout. However, a practical observation focused on skills, within the same assessment plan, could lend itself to a 'mastery' approach – if the skills are all absolute prerequisites.

39. There are also further 'mixed' possibilities between these 2 extremes. In particular, some KSBs within an assessment method may be stipulated as requirements ('hurdles') for a given grade, but then strengths and weaknesses may be allowed to balance each other out across the other KSBs targeted. For example, it might be intended to make any particularly critical aspects (such as those related to health and safety) into specific requirements, but to accept an overall balance elsewhere.

40. Where we have seen the intended degree of assurance approached effectively in assessment plans, it has been clear which (if any) KSBs are to be absolute requirements – or 'hurdles' – and which (if any) KSBs are to be allowed to balance each other out. This is important, as it should help to ensure that end-point assessment organisations accurately reflect the Trailblazer's intentions, making for confidence in what the related results mean. For example:

Annex C – Requirements for core exam – Multiple choice

There are three elements to the exams contained within the bus and coach engineering independent end assessment. There are two multiple choice exams, the first of which has a 'hurdle' section where apprentices must achieve 100% on seven safety critical questions. These will represent 25% of the exam marks available. The remainder of the questions must form a representative sample of the remainder of the criteria detailed below and in Annex D.

It is expected that the pass mark will be set at 70% with an expected pass range of 70-89%. The pass mark must not be below 70%, but minor flexibility to increase the pass mark and adapt the grade range to allow for effective test design by assessment organisations is permitted.

41. Where this issue has not been addressed effectively in assessment plans, it has often been fundamentally unclear whether a 'mastery' or a 'compensatory'

approach has been intended. Alternatively – or sometimes in addition – it has been unclear how the approach is intended to operate. For example, an indication that apprentices must 'achieve 50%' in an assessment method could mean a number of things – half of the targeted KSBs, half on every KSB targeted, half the marks (if marks are to be used) from any or all of the targeted KSBs, and so on.

42. There is nothing inherently problematic with any of those possibilities – it depends on the Trailblazer's intentions as to what degree of assurance is needed. However, these possibilities would give very different levels of assurance about what specifically an apprentice had demonstrated to achieve a grade in that assessment method – which KSBs and to what extent. Such flexibility would not make for confidence in the related results.

Describing the expectations – 'Mastery'

43. Where the intention has been for assurance that all the KSBs targeted by an assessment method have been demonstrated, or that all aspects within KSBs have been demonstrated, assessment plans have often included ***grade descriptors*** (see next table for an example). These can be used to indicate what each grade 'looks like' – in terms of what attainment would comprise – across all (or within all) the KSBs targeted by the assessment method in question.

44. An alternative – with much the same aim – has been a ***criteria-referenced*** approach. With this, the pass requirements (and any other grading requirements) are described in turn for each of the individual KSBs targeted. Assessors then judge how far each of them has been demonstrated (or not). This tends to entail a binary judgement – 'yes' or 'no', 'met' or 'not met'.

45. These types of approach have benefits in terms of transparency – it should be reasonably clear which KSBs are required and what has to be demonstrated. However, there are also limitations. In particular, it can be challenging to articulate the requirements in such a highly specific way that they do not involve *any* degree of interpretation whatsoever.

46. Approaches to grade descriptors that have been effective in assessment plans we have seen have tended to be method-specific, setting out what the evidence would look like for the targeted KSBs in that particular assessment context. This includes where a given KSB is targeted, in different ways and with different evidence in mind, by different assessment methods.

47. Generic descriptors we have seen – those that do not vary by assessment method – have tended to be less effective. This is principally because they have not been sufficiently detailed and specific to support accurate and consistent

interpretation by end-point assessment organisations in relation to particular assessment methods. Lack of detail has also been a weakness of method-specific grade descriptors, as has an excessive degree of scope for interpretation. It is important that grade descriptors can be clearly and consistently understood by assessors.

48. Where grade descriptors are used, it is important that it is straightforward to distinguish in the assessment plan between the KSB itself, and the evidencing of that KSB – the way in which an assessor can judge that the KSB has been 'achieved'. For example:

<table>
<tr><th>Module</th><th>Standard</th><th>What do I need to KNOW</th><th>Pass Criteria</th><th>Distinction Criteria</th><th>Method of Assessment</th></tr>
<tr><td rowspan="5">Meeting regulations and legislation</td><td>1. Know the appropriate legislation and regulatory requirements that affect your business</td><td>1.1 Identify appropriate legislation and regulation and how this effects your organisation</td><td>Explain how the relevant legislation and regulations affect the organisation's customer service provision.</td><td rowspan="2">Explain the potential impact on the organisation if it fails to adhere to each of the relevant legislation and regulations.</td><td rowspan="5">Apprentice Showcase</td></tr>
<tr><td rowspan="4">2. Know your responsibility in relation to this and how to apply it when delivering service</td><td>2.1 Why is it important to keep information confidential within an organisation</td><td rowspan="2">State their responsibilities for keeping information confidential in the organisation.</td></tr>
<tr><td>2.2 What information needs to be kept and remain confidential within your organisation?</td><td rowspan="3">Explain how a code of practice or ethical standards affects customer service.</td></tr>
<tr><td>2.3 What are the responsibilities of the employee under the health and safety at work act?</td><td rowspan="2">State the responsibilities of employees and employers under the Health and Safety at Work Act.</td></tr>
<tr><td>2.4 What are the responsibilities of the employer under the health and safety at work act?</td></tr>
</table>

49. Where this issue has been approached well in an assessment plan, an apprentice or employer can take confidence that a pass (or other grade) for an assessment method means what was intended – the KSBs that have been evidenced, and the extent to which this has been done. Where it has not been done well, then a pass (or other grade) may not carry enough assurance about what has been evidenced and how.

50. The example below clearly stipulates the KSBs that have to be evidenced to achieve a pass in the Professional Discussion element. There are additional criteria in the assessment plan for the other grades. The statement and tables have been drawn from different parts of the assessment plan, but taken together they clearly state what is required for the pass grade in this assessment method. In this example, the assessment plan usefully articulated both what is part of a pass grade and how that could be judged.

Professional discussion	Pass / Distinction / Fail Ascertained through assessment criteria, articulated in Annex G

In order to achieve the professional discussion, the apprentice will demonstrate their competence against each of the assessment criteria below

(Annex G)

In order to pass an apprentice will:	• Clearly articulate examples from the workplace relevant to evidencing competence across the standard • Explain why it is essential to instil the importance of following procedures to staff • Provide examples of how staff are managed effectively, including motivation and development of teams and individual staff members • Provide an overview of how the engineering operation meets the needs of the business • Provide reasoned examples of how the operation operates efficiently • Explain the importance of keeping up to date with current industry trends and provide examples of how this has been achieved • Provide evidence to show they have been part of the budgeting and cost control in the organisation • Describe how the engineering operation meets environmental regulatory needs

	• Provide an effective evaluation of own performance, including behaviours, identifying where opportunities for improvement have been taken and results thereof evaluated • Demonstrate how feedback has been sought from managers and customers and how this has been effectively dealt with

Describing the expectations – 'Compensatory'

51. Where the intention has been for a balance of strengths and weaknesses across (or within) the KSBs targeted, assessment methods have tended to involve either:

    - ***marking*** the KSBs individually, or marking whatever tasks are being used to target the KSBs, then adding those scores together to form a total; or
    - a single, ***best-fit*** judgement being made across all the KSBs targeted by that assessment method.

52. Both of these approaches can be effective. It largely depends on whether the KSBs lend themselves to fine-grained, point-by-point judgements which can then be combined, or if making a single judgement while keeping all the KSBs in mind is more appropriate.

53. Where point-by-point marking has been the intention, assessment plans have tended to propose lists of the relevant requirements. For example, what an assessor would expect demonstrating as part of a practical observation, or the mark scheme for a written test.

54. Where a single best-fit judgement has been the intention, this has tended to require a description of what, overall, a given grade should 'look like'. This is similar to that in the 'mastery' approach described above, but with the emphasis on an overall balance in relation to what is described, rather than complete and consistent coverage of it. It represents a more holistic judgement of performance within the assessment method.

55. Assessment methods that have placed an emphasis on overall attainment – rather than achievement of specific KSBs – have often used ***grade thresholds***. In cases where assessment methods have used marks, these have generally been expressed as a proportion of the total. For example, 60 or more marks out of 100 is a pass, 75 or more is a merit, and so on. Where this has been done effectively, the thresholds have been clearly specified, with a sense that they are in approximately 'the right place' – that is, a sufficiently large proportion of

marks that they should provide a sound base of evidence about attainment, but not so large as to be potentially inaccessible.

56. In some cases, grade thresholds like these have also been accompanied by high-level grading indications – setting out what, broadly, the attainment of a given grade would have to entail. This can be important for designing assessments – particularly (but not only) if multiple end-point assessment organisations are involved. (For example, if a pass is for, say, 60 marks out of 100, to indicate what achieving that 60 marks should mean, roughly speaking, an apprentice has had to demonstrate.)

57. More effective examples of this have also provided, in addition to the grade threshold and high-level grading indication, an idea of the intended distribution of the available marks across the KSBs being assessed. This should help ensure consistent use of the available marks by different assessors and end-point assessment organisations. The example below combines high-level descriptors, a distribution of available marks and a pass threshold for each of the high-level skill areas.

Annex A – Summary of Assessment Method and Grading

PA = Project Assignment (Total marks available = 100)

PD = Professional Discussion (Total marks available = 100)

This table details which Knowledge, Skills and Behaviours are tested in each of the assessment methods within the EPA. It also details the individual elements of each high level Knowledge and Skill that are detailed in the Standard.

The awarding organisation will develop a detailed marking scheme based on this table which will be held in the Assessment Tools

					Method	
High-level Skill	**Skills tested**	**Fail**	**Pass**	**Distinction**	**PA marks**	**PD marks**
Business Development	• Identify, progress and convert sales leads into new clients, candidates and placements as required • Proactively and consistently strive to identify and obtain new business opportunities • Source suitable vacancies in line with company policies and sales procedures • Manage and profitably develop client relationships	• Unable to show any depth of business acumen • Requires support to build new relationships • Waits for tasks to be assigned	• Demonstrates a sound understanding of commercial priorities • Independently seeks and secures new relationships • Proactively initiates and completes tasks	• Maximises opportunities to deliver profitable new business • Proactively contributes to sales activity outside of own specialism • Takes ownership of tasks in a proactively and timely manner		✓ **20 marks** **(min 4 marks = pass)**
Candidate Management	• Identify and attract candidates using all appropriate methods to fill jobs • Monitor responses/applications received and make sure that candidate applications are processed efficiently • Shortlist and present suitably qualified applicants against defined job vacancies • Manage the recruitment and selection processes by	• Employs ineffective sourcing methods • Shows inaccurate interpretation of CV content • Frequently makes poor decisions	• Consistently sources relevant candidates for current vacancies • Accurate assessment of candidate relevancy • Decisions are thought through, using a range of information or techniques	• Builds candidate pools and networks for current and future vacancies • Accurate and rapid assessment of candidate skills, knowledge and motivations • Decisions are fully evidenced and justified • Adapts decision making to each situation.	✓ **20 marks** **(min 4 marks = pass)**	

	effectively liaising with the client, candidate and internal teams • Successfully place suitable candidates with clients					

Describing the expectations – 'Mixed'

58. In some cases, the intention has been that some of the KSBs within an assessment method are requirements ('hurdles') for a given grade, but then strengths and weaknesses are allowed to balance each other out across the other KSBs targeted. Where this type of approach has been addressed effectively in assessment plans, this has generally been owing to clarity – a clear and unambiguous indication of what is a 'hurdle' and what is not.

59. In its simplest form, this type of approach might indicate, say, that there are ten KSBs, and that in order to pass, the first three must be achieved, as must any other three of the remaining seven. This can be effective if there is a fairly clear 'split' in the KSBs between those that are absolute requirements and those where a balance across them may be acceptable.

60. It is also possible to combine a 'numerical threshold' with specific expectations – for example, to say that a pass is for 60 or more marks out of 100, and that within that, an apprentice must have met KSBs 'a', 'b' and 'c'. However, this does create complexities for setting and marking assessments. For example, the 'specifics' must always be 'designed in' so that the assessment covers them. Also, the assessor (or whatever process determines the result) has to 'know' that it is not only the overall score that counts, but also the achievement of the specific expectations. This can be challenging to ensure.

61. A different type of 'mixed' approach that can be effective is where each KSB is individually scored – say, on a scale of 1 to 5 – and, in order to pass, an apprentice has to get at least, say, 2 on every KSB. The grades available for the assessment method then come from their *overall* score. So, if there were ten KSBs, it might be that a pass would be for 20 marks (as that would be the outcome of achieving 2 marks on each of the ten) with merit and distinction at 30 and 40 marks respectively. An approach like this can usefully blend minimum ('mastery') expectations with a degree of ('compensatory') balance.

62. That type of blended approach can also be used between different grades. For example, there might be a 'mastery' approach to achievement of the pass grade, by requiring all the relevant expectations to be demonstrated, but a degree of 'compensation' for the higher grade or grades, by indicating that an

amount (rather than all) of those additional expectations must be demonstrated. In such cases, it is important that the amount needed, in the compensatory part, is as clear as possible – for example, 'all', 'half', or 'x number of' would likely be interpreted more consistently than terms like 'many' and 'much'.

Differentiation between grades

63. Most apprenticeship end-point assessments include a grade (or grades) above pass. It is important that the expectations around these additional grades are clear and appropriate. Where we have seen this issue approached effectively in assessment plans, setting out what constitutes achieving such grades has tended to involve either:

    a) evidencing ***additional KSBs*** to those required for a pass

    b) evidencing some/all of the KSBs for a pass, ***but to a higher level***

    c) a ***combination*** of these approaches.

64. All of these can be appropriate – it depends on what most characterises higher levels of attainment in the apprenticeship in question. That is, whether it is greater 'breadth', greater 'depth', or a mixture of these.

65. Where we have seen this issue approached effectively in assessment plans, the expectations have been clear and progressive. For instance, where grade descriptors have been used, these have clearly set out what typifies attainment above the pass level. As a rule, this has tended to be most effective where the expectations are substantively different in terms of their content and detail – differentiating solely through word choice (for example, 'generally' for pass and 'consistently' for merit) has not tended to be enough. The example below clearly sets out (moving from left to right) distinctive, high-level criteria for each of pass, merit and distinction.

SKILLS	Pass	Merit	Distinction
Control circuit application Electrical and electronic control systems setting, testing and fault finding; and their integration with system-associated communication networks	Sets, tests and fault finds in existing control circuits to the correct industry standard.	Develops control circuits for a range of applications	Explains how to integrate control circuits with associated communication networks.
Mechanical operations Positioning, fixing, jointing and testing of pipework, electrical circuits and water circuits where relevant.	Carries out the mechanical operations for pipework, electrical and water (or other) circuits where relevant.	Manages variations to pipework, electrical and water (or other) circuits where relevant.	Recognises and rectify faults in pipework, electrical and water (or other) circuits where relevant.
Application of mathematical principles Determining heating and cooling loads and selecting and balancing appropriate components and systems for maximum performance and efficiency.	Applies formulae to basic mathematical problems.	Manipulates formulae to determine a broad range of RACHP heat loads.	Uses mathematical principals to justify the selection and balance of appropriate equipment for a broad range of RACHP heat loads or applications

66. Similarly, thresholds for grades above pass have been a meaningful 'distance' from lower thresholds – so that their achievement will represent a genuinely different level of attainment. Specifying such expectations helps to illustrate a clear *progression*, for individual assessments, between a pass grade and any additional, higher, grades.

Section 3: Generating overall results

67. The overall grade for an apprenticeship needs to reflect the level of competency that an apprentice has achieved in relation to the specified KSBs. This is done by combining (often termed 'aggregating') their results from the individual assessment methods. There are a number of ways of doing this and it involves a number of considerations. This process is crucial in ensuring apprenticeship grades provide a clear and reliable indication, to apprentices and employers, of attainment in the KSBs that the Trailblazer has specified.

68. Clarity and simplicity have tended to be hallmarks of effective approaches we have seen to generating overall results. These characteristics should help ensure that, where different end-point assessment organisations are involved, they have a consistent and accurate understanding of the Trailblazer's intentions. They should also help ensure that employers and apprentices are able to understand what results are 'telling them' – in a way that more complicated schemes may not.

Degree of assurance

69. A fundamental consideration in aggregation – as with assessment methods and individual assessment grading – is whether assurance is intended about *specific aspects*, or the emphasis is on *overall attainment*, so a balance of strengths and weaknesses is accepted.

70. Where the intention has been for specific assurance, effective approaches in assessment plans have tended to entail straightforward tabulations of the overall grades that result from different combinations of the grades for individual assessment methods. For example:

Assessment 1	*Pass*	*Pass*	*Merit*	*Merit*
Assessment 2	*Pass*	*Merit*	*Pass*	*Merit*
Overall result	***Pass***	***Pass***	***Pass***	***Merit***

71. By contrast, assessment plans with effective approaches to emphasising overall attainment – that is, where a degree of balance is permitted – have tended to specify overall scores required, which broadly stem from the scores available in individual assessment methods. For example, an overall pass might be 120 out of 200, where there are 2 assessments, each of which is marked out of 100.

72. There are different options within such an approach. An *entirely* compensatory model would mean, using the example quoted, that an apprentice who scored 100 in one assessment and 20 in the other assessment would pass overall. A *partially* compensatory model might say that an apprentice must score at least 120 overall – but that, within that, they must score at least 40 in each assessment method. So an emphasis on overall attainment, but with a degree of specific assurance as well.

73. Approaches to aggregation often use 'like for like' combination. For example, where the intention is for specific assurance, the overall grades that result from different combinations of individual grades such as those shown in the table above. Alternatively, where the emphasis is on overall attainment, the overall grades that come when different numerical scores are added together.

74. It is important to note that numerical scores and grades are not expressed using a common scale, so they cannot be added together. Where the intention is for them to be combined, this has been addressed mostly effectively where assessment plans have included tabulations of how results from the individual assessment methods combine to give overall grades. For example:

Assessment 1	*Pass (or better)*	*Pass (or better)*	*Merit (or better)*
Assessment 2	*60+ (out of 100)*	*75+ (out of 100)*	*90+ (out of 100)*
Overall result	***Pass***	***Merit***	***Distinction***

Assessment weightings

75. In aggregating the results from individual assessment methods, it is important that each of them contributes the importance it is intended to have to the overall result. This importance is often implicitly built into assessment methods through the KSBs they target – that is, an assessment method targeting more KSBs than another has a greater notional weighting in terms of the amount of the Standard it covers.

76. Assessment plans that have addressed the issue of importance effectively have often done so by treating it as an implicit consideration, resulting from where and how KSBs are targeted, rather than by seeking to set out explicitly any notional 'weightings'. For example, indicating for 2 assessment methods, that one is to be worth 60% and one worth 40%.

77. The indication of such weightings is often not meaningful – if they are already built into the targeting of KSBs. It can also be difficult to put such weightings into practice. For example, if there are 2 assessment methods, both of which set out detailed pass criteria for each individual KSB they target, and the overall pass comes simply from achieving a pass in each of the assessment methods, then suggesting the assessment methods have weightings (particularly different weightings) does not follow.

78. The use of quantitative weightings has been most effective in assessment plans only in cases where: the assessment methods (all) use numerical scores; the overall grade is from the aggregation of those scores; and the allocation of KSBs to the different methods does not, in itself, reflect the amount they are intended to contribute to the overall result.

79. Where that has been the case, it has sometimes been useful to give assessment methods a particular weighting in the overall result by designing them in such a way that the marks they contribute reflect their intended value. For example, if assessment X is intended (broadly speaking) to be 'worth' twice as much as assessment Y, giving it a mark total that is twice as large – 120 marks for X versus 60 marks for Y, or whatever.

80. Where any intended weightings have not been 'designed in' to assessment plans, the approach used has involved 'scaling'. For example, if there are 2 assessments, both marked out of 100, but the intention is that assessment A

should be 'worth' twice as much as assessment B, then the scores for assessment A can be 'scaled' upwards. In this case, that would mean multiplying those scores by 2 – so an apprentice with a score of 65 out of 100 marks for assessment A would end up with a *scaled score* of 130 out of 200 marks for that assessment. This would then be added to their score for the non-scaled assessment B to give a total out of 300 marks.

81. It is important to note that where scaling is to be used, this should always involve a 'positive factor' – that is, multiplying the original scores by a value greater than one, whether this is a whole number or a decimal number. The use of 'negative scaling' – multiplying original scores by a value less than one – undermines reliability. In particular, it compresses the distribution of scores, so that they are less spread out than they were originally: this makes any differences in performance less clear. Also, because there are fewer possible scaled scores than original scores, it inherently means some different original scores end up as the same scaled score: this is also undesirable.

82. Although scaling can be useful, when properly applied, it has often been more effective when intended weightings have been 'designed in' to assessment plans. This has meant scaling is not needed, since the degree of importance attached to each assessment method is already built in. For instance, in the example above, if assessment A were marked out of 200 to begin with, or assessment B were marked out of 50, no scaling would have been necessary.

Making decisions about aggregation

83. As a final point on aggregation, it is important to emphasise that the degree to which it entails a choice depends, in part, on the assessment methods used and the approach to individual assessment grading. The type of information they provide is what the aggregation will be able to use. Again, this underlines the inter-connected nature of the different intentions and considerations in designing an assessment plan.

84. For example, if the assessment methods indicate detailed pass criteria for each KSB, and the overall pass comes from achieving a pass in each of them, then an approach to overall results that emphasises overall attainment – that is, where a degree of balance is permitted – is not viable. Similarly, it would not be meaningful to suggest notional weightings for those assessments – their weighting would already be built into them through the KSBs they target.

Section 4: Approach to re-takes and re-sits

85. In cases where an apprentice needs to re-take or re-sit their end-point assessment, there should be the facility for them to do this. In the context of reformed apprenticeships, the term 're-take' is used to mean that further learning is required, whereas the term 're-sit' means no further learning is required. As such, re-takes have funding implications that re-sits may not.

86. It is important that assessment plans make clear any expectations the Trailblazer has about re-takes and re-sits – particularly to help ensure that if multiple end-point assessment organisations are involved, they do not take different approaches on aspects where the Trailblazer has specific views. There should be a 'level playing field' when it comes to these.

87. Two considerations we have found it helpful for assessment plans to set out on re-takes/re-sits are:

    - the number of them permitted (if there is intended to be any restriction on this); and
    - the time period that must elapse from the initial attempt (if there is any intention to prescribe this).

88. A further point that we have seen addressed effectively in assessment plans – again, where there is a particular intention about it – relates to the re-use of assessment evidence by the apprentice. Basically, whether an apprentice may use again any evidence that was deemed satisfactory (or better) at their initial attempt; or alternatively if the re-take/re-sit must draw entirely on fresh material.

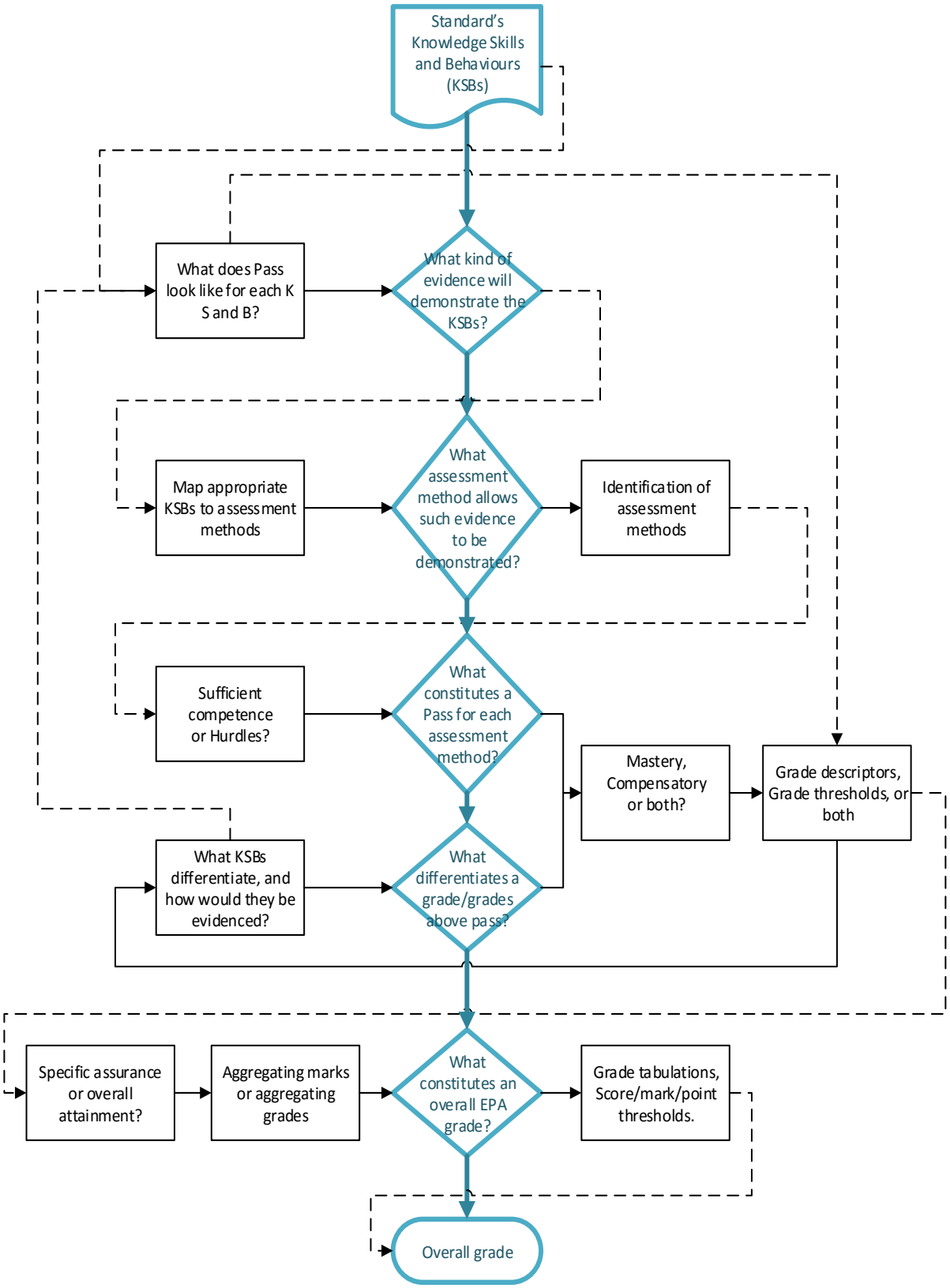
89. Perhaps the most substantive issue in relation to re-takes/re-sits is the grades available. Ordinarily, someone taking an assessment would tend to be able to achieve any of the grades – their grade would be based on how well they performed. However, in some cases, we have seen assessment plans that 'cap' the grades available at a re-take/re-sit – say, to no more than 'pass'.

90. This issue of the grades available in re-takes/re-sits is not straightforward. It can be viewed to be less fair, on those re-taking/re-sitting, if grades *are* capped – as the grade they achieve may not reflect what they have demonstrated. Alternatively, it can be viewed to be less fair, on those not re-taking/re-sitting, if re-take/re-sit grades are *not* capped – as the grades achieved for them would not show that they are the result of an apprentice having a 'further go'. This issue is basically a matter of choice – there is no 'right' or 'wrong' answer. It requires careful consideration, and a judgement, by the Trailblazer.

91. The fundamental point here – key to achieving fairness whichever approach is determined – is that the stipulations must be clear and unambiguous. This is crucial to ensuring, as with the rest of the assessment plan, that all those involved – whether the apprentices taking assessments, the organisations administering them, or the employers using the results – have a clear and consistent understanding of the Trailblazer's intentions.

An approach to thinking through assessment design

92. The diagram overleaf shows one possible way of thinking through the issues involved in designing an assessment plan. The central 'decision flow' column suggests some of the main development stages, the processes on the left are aspects that might inform a decision, and those on the right are the outputs that might lead from a decision. Again, it is important to emphasise that the considerations involved tend to be heavily interlinked, with those at different points both affecting and informing each other.


Standard's Knowledge Skills and Behaviours (KSBs)
What does Pass look like for each K S and B?
What kind of evidence will demonstrate the KSBs?
Map appropriate KSBs to assessment methods
What assessment method allows such evidence to be demonstrated?
Identification of assessment methods
Sufficient competence or Hurdles?
What constitutes a Pass for each assessment method?
Mastery, Compensatory or both?
Grade descriptors, Grade thresholds, or both
What KSBs differentiate, and how would they be evidenced?
What differentiates a grade/grades above pass?
Specific assurance or overall attainment?
Aggregating marks or aggregating grades
What constitutes an overall EPA grade?
Grade tabulations, Score/mark/point thresholds.
Overall grade

References

The following assessment plans, considered by the Technical Advisory Group for end-point assessments at Ofqual, have informed the thinking in this document.

Apprenticeship Title	**Level**
Actuarial Technician	Level 4
Advanced Baker	Level 3
Advanced Butcher	Level 3
Advanced Credit Controller/Debt Collection Specialist	Level 3
Advanced Dairy Technologist	Level 5
Airside Operator	Level 2
Arborist	Level 2
Assistant Accountant	Level 3
Assistant Technical Director (visual effects)	Level 4
Auto-care technician	Level 2
Aviation Ground Operative	Level 2
Aviation Ground Specialist	Level 3
Aviation Operations Manager	Level 4
Baker	Level 2
Beauty Professional	Level 2
Bid and Proposal Coordinator	Level 3
Building Services Engineering Craftsperson	Level 3
Building Services Engineering Installer	Level 2
Bus and Coach Engineering Manager	Level 4
Bus and Coach Engineering Technician	Level 3
Butcher	Level 2
Chartered Manager Degree Apprenticeship	Level 6

Children, Young People and Families - Manager	Level 5
Children, Young People and Families - Practitioner	Level 4
Conveyancing Technician	Level 4
Creative Venue Technician	Level 3
Credit Controller/Collector	Level 2
Customer Service Practitioner	Level 2
Customer Service Specialist	Level 3
Dental Laboratory Assistant	Level 3
Dental Nurse	Level 3
Dental Practice Manager	Level 4
Dental Technician	Level 5
Electrical, electronic product service and installation engineer	Level 3
Engineering Operative	Level 2
Equine Groom	Level 2
Facilities Management Supervisor	Level 4
Facilities Manager	Level 4
Fishmonger	Level 2
Food And Drink Advanced Process Operator	Level 3
Food and Drink Maintenance Engineer	Level 3
Food and Drink Process Operator	Level 2
Food Technologist	Level 3
Forest Operative	Level 2
Further Education Learning and Skills Teacher	Level 5
Golf Greenkeeper	Level 2
Hair Professional	Level 2
Healthcare Assistant Practitioner	Level 5

Healthcare Support Worker	Level 2
Heavy Vehicle Service and Maintenance Technician	Level 3
High Speed Rail Engineering Advanced Technician	Level 4
Higher/Further Education - Assessor-Coach	Level 4
Higher/Further Education - Learning Mentor	Level 3
Highway Electrical Maintenance and Installation Operative	Level 2
Highway Electrician / Service Operative	Level 3
Horticulture and Landscape Supervisor	Level 3
Horticulture and Landscape Operative	Level 2
Installation Electrician/Maintenance Electrician	Level 3
Junior Management Consultant	Level 4
Junior Journalist	Level 3
Land-based Service Engineer	Level 2
Land-based Service Engineering Technician	Level 3
Learning and Development Consultant	Level 5
Learning and Development Practitioner	Level 3
Licensed Conveyancer	Level 6
Mineral Processing Mobile and Static Plant Operator	Level 2
Mineral Processing Weighbridge Operator	Level 2
Motorcycle Technician (Repair and Maintenance)	Level 3
Nursing Associate	Level 5
Operations/Departmental Manager	Level 5
Paralegal	Level 3
Policy Officer	Level 4
Poultry Technician	Level 3
Poultry Worker	Level 2

Probate Technician	Level 4
Procurement & Supply	Level 4
Public Relations Assistant	Level 4
Recruitment Consultant	Level 3
Recruitment Resourcer	Level 2
Refrigeration Air Conditioning and Heat Pump Engineering Technician	Level 3
Safety, Health and Environment Technician	Level 3
Security First Line Management	Level 3
Senior Equine Groom	Level 3
Senior Healthcare Support Worker	Level 3
Senior Leader Masters Degree Apprenticeship	Level 7
Supply Chain Practitioner (Fast Moving Consumer Goods)	Level 3
Teaching Assistant	Level 3
Team Leader/Supervisor	Level 3
Thermal Insulation Operative	Level 2
Thermal Insulation Technician	Level 3
Trade Supplier	Level 2
Visual Effects (VFX) Junior 2D Artist	Level 4

We wish to make our publications widely accessible. Please contact us at publications@ofqual.gov.uk if you have any specific accessibility requirements.

This publication is available at www.gov.uk/ofqual.

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation

Spring Place
Coventry Business Park
Herald Avenue
Coventry CV5 6UB

Telephone 0300 303 3344
Textphone 0300 303 3345
Helpline 0300 303 3346