RESEARCH AND ANALYSIS

Designing qualifications that will be fit for their purposes

The necessity and centrality of purpose analysis

Author

This report was written by Paul E. Newton from Ofqual's Standards, Research, and Analysis Directorate.

Contents

Executive summary

This report is about qualification design and the issues that need to be considered when designing a qualification that will be fit for multiple purposes. It provides conceptual frameworks (thinking tools) that can be used to:

- distinguish between complementary perspectives on qualification purposes
- situate these perspectives within broader design-related considerations
- understand the scope of qualification design and its stages

These frameworks have been developed at Ofqual over the past decade and have been put into practice when drafting new regulations, particularly the purpose perspective framework. As such, the report has direct relevance to the UK policy context and is written with a UK audience in mind, although the principles that it develops are likely to be more widely applicable. The concept that binds these frameworks is the idea of a qualification **purpose**, which the report unpacks in detail and locates within a broader network of design-related concepts.

Context

The challenge of designing qualifications to be fit for multiple purposes has been recognised for a very long time. During the second decade of the 20th century, a national system of examining for England and Wales was established for students aged 16 and over (School Certificate) and 18 and over (Higher School Certificate). According to the report of a high-profile investigation into the HSC by the Secondary School Examinations Council (SSEC), it was intended to serve 2 main purposes:

> Those two functions are – firstly to discover and place in a just order of merit candidates to be recommended for the award of State Scholarships and Local Education Authority Scholarships; and secondly to provide a suitable test of two years' work in Secondary Schools beyond the stage of the First Examination.

(SSEC, 1939, page 8)

Unfortunately, it soon became clear that these purposes were in tension, and that the HSC was not entirely fit for both:

> A compromise has continually to be sought between the technique most suitable for examining the ordinary Higher School Certificate candidate and the very different method that would no doubt be adopted if the selection of scholars were the sole object.

(SSEC, 1939, page 31)

This idea of **tension** between the purposes we want qualifications to achieve lies at the heart of this report. Conflict of this sort is inherent to qualification design, which means qualification design is always a matter of trade-off and compromise.

To help explain what a qualification designer is responsible for designing, the report proposes that a qualification is a procedure for certifying attainment in an education or training context, which includes specifications for syllabus outcomes, qualification structure, and an assessment procedure. The report develops the principle that qualification design decisions (related to these 3 facets) should flow directly from the purposes that have been prioritised for a new qualification. Importantly, these purposes should never be presumed and should be considered anew for each new qualification. We describe this process in detail and refer to it as **purpose analysis**.

Academic papers in this space – particularly those focused on assessment – often dwell on the potential for tension: explaining, for instance, why an assessment designed to satisfy one purpose cannot necessarily be expected to satisfy a different purpose equally well, if at all (Shepard, 1979; Newton, 2007a; Newton 2007b). Papers like these often present long lists of potential purposes – as though attempting to convey a sense of shock and awe concerning the complexity of the task at hand – explaining how this typically necessitates uncomfortable decisions concerning which of these purposes to prioritise and which, inevitably, to de-prioritise or simply dismiss (Newton, 2023a).

Although purposes have challenged qualification designers for well over a century, they remain remarkably poorly theorised, which presents a threat to effective policy and practice. The most problematic aspect of the literature on qualification purposes is that it has developed in silos. Thus, assessment purposes tend to be analysed by assessment scholars who tend to publish in outlets targeted at assessment specialists, while curriculum purposes tend to be analysed by curriculum scholars who tend to publish in outlets targeted at curriculum specialists.[1] This is a problem because qualification design requires both perspectives to be considered simultaneously, alongside others too. The present report explains why this is the case and offers suggestions for how to undertake a more **integrated** analysis.

Concepts

One of the challenges Ofqual routinely grapples with is how to proceed when different perspectives on how best to design a qualification recommend conflicting design decisions (see Newton, 2017a). Responding to this challenge has involved distinguishing between the following 3 perspectives, each of which is critical to qualification design:

[1] Cedefop (2010) provides a rare example of a far broader analysis of qualification purposes.

- the **expertise** perspective – concerning the body of learning that needs to be acquired for the award of a certificate
- the **engagement** perspective – concerning the engagement that needs to be exercised (by learners and by teachers) to achieve the required standard
- the **information** perspective – concerning the information that needs to be provided by a qualification certificate

We refer to these as distinct perspectives because they each answer the question of what a qualification needs to do from a different vantage point, to establish that:

- the body of learning that is acquired while studying for the qualification is as valuable as possible (for learners, for the economy, and for society) – expertise
- the features and processes that comprise the qualification actively facilitate uptake, learning, completion, continuation (and so on) – engagement
- the information that qualification results provide is of the right kind and of sufficient accuracy (to support effective decision-making) – information

These perspectives are not necessarily or inherently in tension with each other, in terms of their implications for qualification design. However, they certainly can come into conflict with each other and they often do. For instance, a qualification being designed for a group of largely disengaged learners might build in a substantial amount of optionality, to help personalise its content to the interests and motivations of individual learners (an engagement-driven design decision). Yet the body of learning developed through a qualification like this might have limited market value, particularly if its relevance and utility to employers remained unclear, owing to a lack of clarity concerning what students had actually learnt (a concern from both the expertise and information perspectives). So, this might raise the question of whether – when push comes to shove – it is more important for the qualification to play a role in reengaging disaffected learners, or more important that learners who ultimately pass the qualification have acquired a body of learning with very clear relevance and utility for employers (even if this meant that far fewer students opted for the qualification in the first place, or passed it at the end of their course of study). Decisions like this are not at all easy to take. But it is better for them to be taken explicitly, rather than for them to be taken implicitly, through qualification design decisions based purely on traditions of custom and practice or gut instinct.

It is not simply the overarching perspectives on qualification purposes that can conflict with each other in terms of their implications for qualification design. There are multiple categories of purpose within each perspective, and multiple purposes within those categories. It is likely that a qualification designer will want to build their qualification to satisfy multiple purposes within each perspective, and these purposes may also be in tension with each other. For instance, a designer might want to design an end-of-year assessment to provide information that can be used

both to gauge the overall level of attainment of students in a particular subject area (to place them in an appropriate set next year) and to identify specific gaps in their learning (to guide subsequent instructional interventions). Yet an assessment for drawing valid inferences concerning gaps in learning would need to be very much longer than an assessment for drawing valid inferences concerning overall levels of learning. Purpose analysis requires designers to work through potentially conflicting implications like these, and to revisit their prioritisation of intended purposes when they encounter problems.

Although these purpose perspectives – and the many different purposes subsumed within each one – provide a focus for purpose analysis, they are not the only considerations that need to be borne in mind. Qualification design also needs to be tailored to the composition of the target **cohort** for the new qualification and the anticipated circumstances of their teaching, learning, and assessment.

It also needs to be tailored to broader **systemic requirements** related to the integrity of the system within which individual qualifications play their part. The report distinguishes between:

- **structural** integrity – to secure the system itself – including the need for qualifications to relate to each other coherently, and to be resilient to large-scale external crises (pandemics, cyber attacks, and suchlike)
- **cognitive** integrity – to secure public understanding and participation – including the need for qualification practices to be transparent and plausible
- **ethical** integrity – to secure public confidence and consent – including the need for processes and products to be robust (that is, resistant to subversion) and just (that is, fair and seen to be fair)

Of course, qualification design also needs to respect various **pragmatic constraints**, including costs and resource availability, as well as constraints related to wider political, legal, and commercial concerns.

This report describes each of these factors and their potential implications for qualification design. With such a wide array of factors at work, it should not be surprising that qualification design is always a matter of **trade-off** and **compromise**.

A theme that becomes increasingly apparent throughout this report is the fact that qualification design is a highly **technical** business, which demands technical expertise from its practitioners, including expertise in purpose analysis. There is no single correct approach to qualification design. But there are undoubtedly incorrect approaches, for instance, where decisions are made without sufficient consideration of alternative perspectives, or where decisions are made purely based on traditions of custom and practice with insufficient attention to their likely consequences in new configurations, contexts, and circumstances.

The following account is not a manual for qualification design. But it does help to articulate the sorts of considerations that need to be taken into account whenever a new qualification or qualification system is being planned.

Finally, the frameworks discussed in this report should be considered work-in-progress, as much remains to be done to determine how best to utilise them to support policy and practice in this area.

Introduction

The process of designing a qualification that will be fit for purpose might seem to be fairly straightforward – simply a matter of:

- deciding what purpose the new qualification will need to serve
- identifying design characteristics that would enable it to achieve that purpose
- building the new qualification to incorporate those characteristics, thereby rendering it fit for that purpose

In practice, and especially when done well, the process tends to be far more convoluted – more like this:

- canvass a wide range of opinions on the purposes that the new qualification will need to serve
- reach some kind of consensus over the variety of purposes that the new qualification should ideally serve
- identify design characteristics that should, in theory, enable the new qualification to achieve purposes of that sort
- where certain of those intended purposes lead to conflicting design decisions, recognise the need to prioritise between them (and do so)
- build the new qualification to incorporate characteristics that seem likely to render it fit enough for the highest priority purposes (and, where possible, not entirely unfit for any remaining prioritised purposes)

In other words:

- qualification design is more complicated that it is often assumed to be
- qualifications always need to be designed with multiple purposes in mind
- it is normal for intended purposes to recommend at least somewhat conflicting design decisions, which means that qualification design is always a matter of trade-off and compromise
- qualifications may never be perfectly fit for purpose – not even for their highest priority purpose – but they ought to be fit enough for their most important purposes (even when they end up being entirely unfit for other purposes)

We have introduced the term **purpose analysis** to describe the thinking, debating, and consensus-building that needs to occur before finalising any design decisions. The product that purpose analysis culminates in – which is essentially a profile of prioritised purposes – provides a single point of reference for all subsequent

qualification design and development activity (as well as for qualification review, further down the line).

The following 2 sections explain why concepts such as purpose and qualification are not as easy to define as we might assume. Subsequent sections explain the many different meanings that can be attached to qualification purposes, the differing implications alternative meanings hold for qualification design, and therefore the importance of prioritising between qualification purposes.

Purposes

Part of what makes purpose analysis complicated is the fact there are all sorts of purposes that need to be considered when designing a new qualification. It is not simply that there are different purposes. There are different categories of purpose, and there are quite different perspectives on qualification purposes (such that purposes cluster within categories within perspectives). To make matters worse, there are no generally agreed taxonomies or classification schemes for qualification purposes, and they have been described in different ways by different people at different times (see Newton, 2023a).

Within Ofqual's General Conditions of Recognition (GCR), fitness for purpose requirements feature within Conditions D and E. Specifically, 'fitness for purpose' is Ofqual's overarching technical requirement for regulated qualifications, as set out in Condition D1:

> D1.1 An awarding organisation must ensure that each qualification which it makes available is fit for purpose.

Condition E1 refers separately to qualification 'objectives' which overlap with the idea of a qualification purpose:

> E1.1 An awarding organisation must ensure that each qualification which it makes available or proposes to make available – (a) has a clear objective in accordance with this condition, and (b) meets that objective.
>
> E1.2 The objective of a qualification must be such as to lead to a benefit for Learners who have reached a specified level of attainment, and may include – (a) preparing Learners to progress to a qualification in the same subject area but at a higher level or requiring more specific knowledge, skills and understanding, (b) preparing Learners to progress to a qualification in another subject area, (c) meeting relevant programmes of learning, (d) preparing Learners for employment, (e) supporting a role in the workplace, or (f) giving Learners personal growth and engagement in learning.

For certain qualifications, Ofqual publishes Qualification Level Conditions (QLC), which include additional regulations beyond the GCR, such as the general purposes that the type of qualification in question will need to serve. For instance, the purposes that were specified in 2014 for the new GCSE (9-1) QLCs stated that:

> We have developed these requirements with the intention that GCSE qualifications (graded from 9 to 1) should fulfil the following purposes:
>
> To provide evidence of students' achievements against demanding and fulfilling content;
>
> To provide a strong foundation for further academic and vocational study and for employment; and
>
> To provide (if required) a basis for schools and colleges to be held accountable for the performance of all of their students.

Notice how these regulations seem to invoke quite different kinds of purpose. The first seems to relate to assessment: to provide adequate and appropriate assessment evidence. The second seems to relate to learning: to equip learners for progression opportunities, which chimes with the idea of a qualification objective from GCR Condition E1. The third seems to relate to a use that qualification results might be put to: to hold schools to account based on aggregated qualification results.

Notice how the first and third of these purposes seem to relate to **information** provided by qualification results – its production and use respectively. The second seems to relate to a body of learning – the **expertise** acquired while studying for the qualification. Shortly, we will see how these statements reflect 2 different perspectives on the purposes that qualifications are supposed to serve, and these are not the only perspectives.

As already noted, purpose analysis is important because it builds the foundations for qualification design. In other words, the decisions that we make when designing a new qualification are – or, at least, or ought to be – guided directly by the purposes we want it to achieve. But what does it actually mean to design a qualification?

Qualifications

More specifically, what is it that needs to be designed, when we design a new qualification? Although this might sound like a straightforward question, it is surprisingly tricky to answer, which is partly because the qualifications industry lacks a generally accepted definition of its core concept. This, for instance, is a rare example of a technical definition of 'qualification' from the annex to a European Recommendation on the validation of non-formal and informal learning (EU, 2012):

> a qualification means a formal outcome of an assessment and validation process which is obtained when a competent body determines that an individual has achieved learning outcomes to given standards

Although this is a precise and considered formulation, it has its limitations. From the perspective of this report, for instance, it provides little if any insight into exactly what it is that needs to be designed. Is it the outcome, or the process for determining the outcome, or both, or something else entirely?

Perhaps non-technical definitions might help here? Among a variety of alternative everyday definitions, Dictionary.com (British) explains that 'qualification' can refer to:

- "an official **record** of achievement" – the validation of an achievement, as in: she was awarded a qualification
- "an **ability**, quality, or attribute" – the achievement itself, as in: he has the right qualification to succeed in that role
- "the **act** of qualifying" – the process of securing an achievement, as in: qualification will not be easy

In following sections, we will see how each of these everyday definitions is relevant to the process of designing a qualification, as they can be mapped on to 3 distinct perspectives on the purposes that qualifications need to be designed to achieve:

- the **information** perspective – concerning the information that needs to be provided by a qualification certificate
- the **expertise** perspective – concerning the body of learning that needs to be acquired for the award of a certificate
- the **engagement** perspective – concerning the engagement that needs to be exercised (by learners and by teachers) to achieve the required standard

Unfortunately, this does not bring us much closer to understanding what needs to be designed when designing a qualification. So, perhaps we need to consider the activities that people who might reasonably call themselves 'qualification designers' appear to be doing when they design a qualification. Yet even this is ambiguous, as responsibilities for qualification design are often shared across a variety of stakeholders, with no obvious 'chief' design architect. Consider, for example, England's subject-based GCSE qualifications:

- their core syllabus outcomes are specified by the Department for Education (DfE)
- rules governing the development of their assessment procedures are specified by Ofqual

- further details concerning syllabus outcomes and assessment procedures are specified by exam boards[2]

Each of these stakeholders – the DfE, Ofqual, and the exam boards – plays a critical role in the qualification design process for GCSEs in England. And they are supported by other key stakeholders too – such as curriculum experts, subject domain experts, and those (including employers) who use qualification certificates – each with slightly different perspectives on the qualification design process, as well as different vested interests.

What seems to be clear, though, is that qualification design involves specifying:

- syllabus outcomes (which characterise the domain of learning)
- an assessment procedure (which governs how assessments will be developed and delivered)
- the overarching structure of the qualification (for example, whether modular or linear, with implications for syllabus outcomes and assessment procedure).[3]

In other words, qualification design is partly an exercise in **curriculum** design – or more specifically **syllabus** design – and it is partly an exercise in **assessment** design. Yet qualification design also requires careful attention to the potential impact of syllabus and assessment decisions on **pedagogy**. Indeed, it is often the case that syllabus and assessment decisions are made specifically with pedagogical considerations in mind, as will become clear later. So, it is important to recognise that qualification design lies right at the intersection between curriculum, pedagogy, and assessment (see Figure 1).[4]

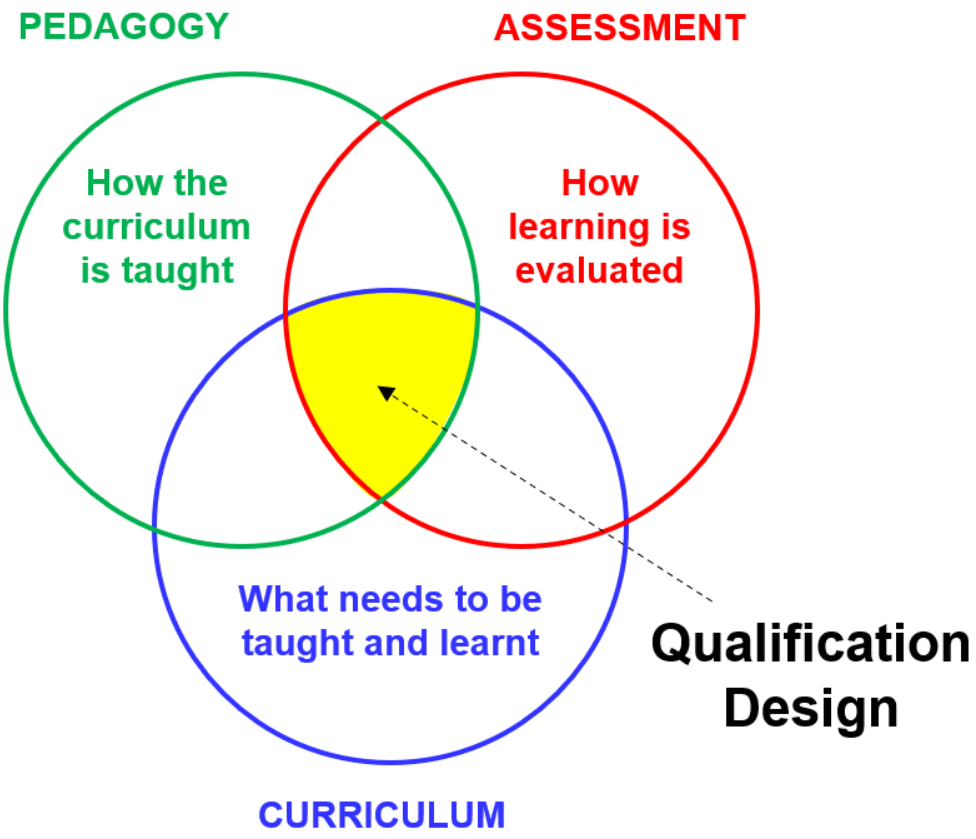
---

[2] The boards then publish qualification syllabuses (which tend now to be known as specifications), develop and deliver assessments, process assessment results, and award qualification certificates. Incidentally, it is worth noting that England is unusual, internationally speaking, in supporting a qualifications market, which permits the same (or equivalent) qualifications to be awarded by multiple bodies. National certification systems are normally administered by national bodies, rather than responsibilities being shared across multiple levels of administration.

[3] These terms and their referents will be explained in detail shortly.

[4] Note that the 3 perspectives are not labelled 'curriculum' (expertise), 'pedagogy' (engagement), and 'assessment' (information) even though there is a certain amount of overlap here. There is a large amount of overlap for curriculum (and expertise) and assessment (and information), but less so for pedagogy. In particular, the idea of engagement seems broader than just pedagogy: later, we distinguish between motivational and pedagogical aspects of engagement. In addition, it is not unreasonable to say that curriculum and assessment issues are more fundamental to qualification design than pedagogical ones, even though implications for pedagogy should never be overlooked when designing a qualification.

Figure 1. Qualification design at the intersection



This analysis suggests the following definition of a qualification, which explains what a qualification designer is responsible for designing:

> A qualification is a procedure for certifying attainment in an education or training context, which includes specifications for syllabus outcomes, qualification structure, and an assessment procedure.[5] These specifications have direct intended implications for what students ought to learn, aspects of teaching and learning processes, and how to interpret qualification results.

One final observation concerns the level of analysis that we are referring to. For instance, we might be referring to the design of:

- a free-standing **qualification** (such as the Extended Project Qualification)

[5] In recent years, 'specification' has come to replace 'syllabus' as the technical term for a document that is produced by an awarding body to describe a particular qualification. Ofqual now requires awarding organisations to publish a specification for each of the qualifications that it makes available, which sets out a variety of critical details. In the above definition, the term 'specification' is intended less formally, simply to mean a detailed description of, say, syllabus outcomes.

- a qualification **type** (such as the General Certificate of Secondary Education, which spans a variety of subject-based syllabuses)
- a qualification **framework** (such as the Qualifications and Credit Framework, to which many but not all regulated qualifications were accredited during the 2010s)

Purpose analysis is relevant to the design of all 3 levels. For the sake of this report, however, it is perhaps easiest to start by thinking of purpose analysis in relation to the design of free-standing qualifications, or qualification types, before thinking about the role they play within broader frameworks.

In this report, we will sometimes refer to a qualification **system** as distinct from an individual qualification. When we consider the purposes that qualifications need to serve – within the section headed 'Purpose perspectives' – we will tend to focus at the level of individual qualifications, or qualification types. Yet the subsequent section on systemic requirements explains why qualification design sometimes needs to be driven by higher-level considerations related to the system within which an individual qualification operates. Having said that, we will use the term 'system' fluidly – at times referring to a qualification type (for example, the GCSE system), at other times referring to a qualification framework (for example, the QCF system), and at other times referring to the entire landscape of regulated qualifications.

Qualification design

Qualification design decisions should never be arbitrary, nor unthinkingly default to longstanding traditions of custom and practice. Instead, they ought to be tailored consciously to a bespoke, prioritised profile of intended purposes.[6] This is why purpose analysis is both necessary for and central to qualification design.

Idealised framework

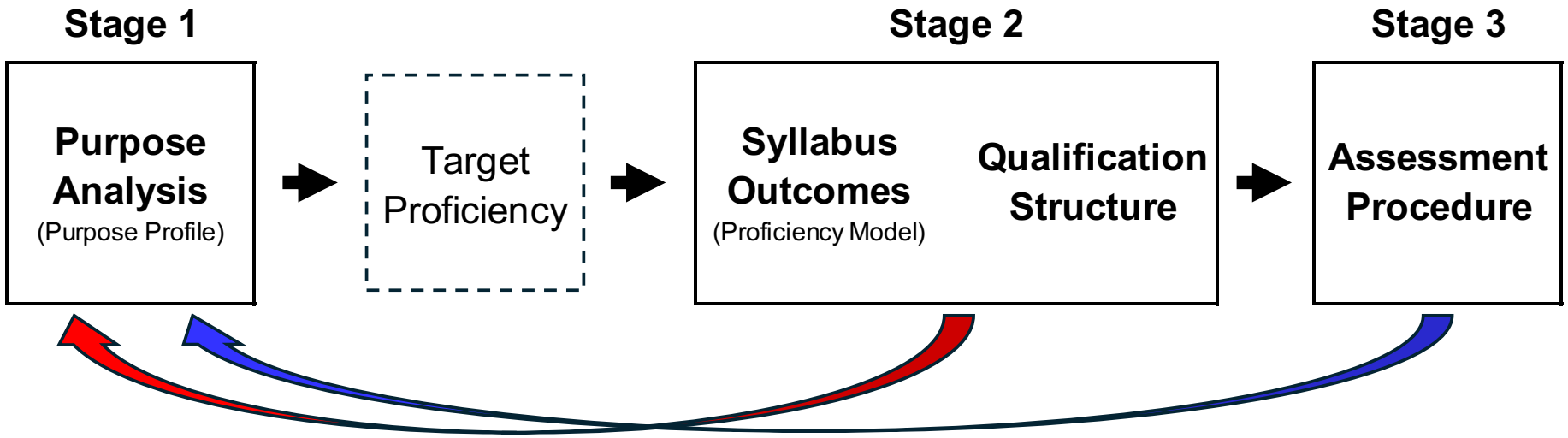
In the final report of a recent research programme, we proposed that qualification design involves 3 critical stages (Newton, Curcin, Clarke, & Brylka, 2024a). These stages comprise the following activities and products:

1. **purpose analysis** – we identify the full range of purposes (or goals) that explain why the proposed new qualification is deemed to be necessary for a particular cohort of learners in a particular set of circumstances (which elucidate the anticipated contexts of teaching, learning, and assessment), and then we prioritise between those purposes
2. **syllabus outcomes and qualification structure** – we construct a model of the learning that will need to be certified in order to maximise the likelihood of satisfying our profile of prioritised purposes, which also involves decisions concerning the overarching structure of the qualification
3. **assessment procedure** – we design a procedure to govern the development of assessment materials, the delivery of assessment events, the processing of assessment evidence, and the award of qualification certificates

Figure 2 illustrates this idealised framework for qualification design. The blue and red arrows indicate that – although there is a natural order to the qualification design process – it needs to be understood (and practised) iteratively, as challenges encountered during stages 2 and 3 may sometimes necessitate a fundamental re-evaluation of decisions reached during stage 1.

[6] It is not that traditions and customs have no place in qualification design, development, or delivery. Quite the reverse: they can often be very important for supporting public understanding and instilling public confidence. The point is simply that qualification design needs to be purpose-driven, not custom-driven, and where purposes change so too may practices, even very longstanding ones.

Figure 2. An idealised framework for qualification design



Purpose analysis

The foundation for effective qualification design is a prioritised profile of intended purposes. Assembling and prioritising a purpose profile for a new qualification requires us to consider:

- characteristics of its target cohort, including the circumstances under which they are likely to be taught, to learn, and to be assessed
- a variety of perspectives on the purposes that the qualification will need to serve
- requirements that relate to the overarching qualification system
- constraints that set parameters for the qualification design process

These cohorts, perspectives, requirements, and constraints are explained in detail in the next main section of this report (headed 'Drivers and constraints').[7]

Purpose analysis is essentially a **policymaking** process, through which the various purpose perspectives – and the specific purposes those perspectives subsume – are prioritised, with due regard to characteristics of the target cohort, systemic requirements, and anticipated constraints. Its rationale is to ensure that all subsequent qualification design decisions – related to syllabus outcomes, qualification structure, and assessment procedure – are made cogently and coherently, not arbitrarily or unthinkingly.[8]

Although it is only the first stage in the qualification design process, decisions made during this stage will (by definition) have implications for decisions made subsequently. Purpose analysis lays the groundwork for these subsequent decisions, with a particular focus on ensuring that the resultant profile of prioritised purposes is both coherent and viable. This involves anticipating and interrogating the design implications, and their likely consequences, that follow from alternative

[7] So, just to be clear, purpose analysis extends beyond the analysis of intended purposes, or goals, to consider a full range of potential design drivers and constraints.

[8] It seems natural to distinguish between syllabus outcomes, qualification structure, and assessment procedure, and it seems reasonable to assume that they collectively exhaust all the critical design decisions that need to be taken when building a qualification. Having said that, it may not always be obvious where a particular design characteristic is best located. For instance, the aggregation approach tends to be thought of as an assessment-related decision, although the requirement to pass a particular component or unit (for the award of the qualification) could equally be considered a structure-related decision, despite the fact that both of these decisions ought really to be determined by the nature of the target proficiency itself, which would be an outcome-related decision. So, there is likely to be a certain amount of overlap between these 3 facets.

purpose-related aspirations, to ensure that the purposes within the emerging profile are:

- mutually compatible
- suitable for learners from the target cohort and their circumstances
- compatible with broader systemic requirements
- practically viable in relation to wider constraints

Where any of these criteria seem not to be satisfied, the profile of prioritised purposes will need to be revisited.

It will not be possible to anticipate and interrogate each and every potential design decision, but the more thorough the analysis that occurs at this stage, the smoother the progress that will be possible during subsequent stages, and the lower the risk of having to return to the drawing board a long way into the design process.

Once the profile of prioritised purposes has been assembled, it will have direct (logical) implications for what will need to be taught, learnt, and assessed. We refer to whatever needs to be taught, learnt, and assessed as the **target proficiency** for the new qualification.[9] The profile of prioritised purposes provides a single point of reference for all subsequent qualification design activity.

Syllabus outcomes

Stage 2 involves translating the target proficiency that is implicit in the profile of prioritised purposes into something far more explicit. There is no universally agreed approach to undertaking this activity, and different approaches are adopted for different qualifications.[10] At one extreme, the process revolves around constructing a **syllabus**, which is essentially a list of topics for teaching. At the other extreme, the process revolves around constructing a set of **outcomes**, which is essentially a list of objectives for learning. We have introduced the term ‘syllabus outcomes’ for this report as a label that aims to capture the essence of both approaches.

[9] At the end of the first stage, the target proficiency will still be implicit, which is indicated in Figure 2 by a dotted line. It is explicated during the second stage.

[10] Unfortunately, there is no coherent, overarching literature on specifying syllabus outcomes, which makes it very hard to direct the interested reader to relevant resources. For instance, there are important insights to be gleaned from scholarship in fields as disparate as educational assessment (for example, Perie & Huff, 2016; Raymond, 2016), instructional design (for example, Mager, 1997; Gronlund & Brookhart, 2009), textbook design (for example, Cambridge Assessment, 2018), and so on. The resources just mentioned are for illustration, and many others could have been identified.

The activity of explicating a target proficiency results in a **proficiency model** of one sort or another (Newton, Curcin, et al, 2024a). This model might be constructed in different ways for different qualifications, but it is likely to involve:

- labelling the proficiency (for example, 'the ability to assemble bicycles')
- specifying what the proficiency includes and excludes (for example, the basic skills of assembling such as torquing, the sorts of bicycles that an assembler would or would not be expected to assemble, whether they would be expected to be able to fabricate bespoke parts, whether they would be expected to manage team members, and so on)
- distinguishing between core and peripheral abilities (for example, bicycle assembly versus working with customers versus understanding the bicycle market)
- exemplifying different levels of proficiency (for example, related to speed of assembly or to autonomy of working)

This final bullet point is important to emphasise. Ideally, the proficiency model should include more than just a list of topics to be taught – it should make some attempt to characterise the anticipated learning itself. This should include information concerning what it means to have achieved a satisfactory level of learning within the domain, that is, how broad and deep the learning needs to be for a learner to be judged to have passed the qualification (and to achieve higher grades, if awarded).

Proficiency modelling is likely to involve written descriptions. But it might well incorporate a variety of representational formats, especially when the proficiency is exemplified via performance evidence (for example, using audio or video recordings). Exemplification is important for filling in the gaps that remain when relying heavily upon written statements.

It is fair to say that proficiency modelling is poorly theorised as a component of qualification design (Newton, Curcin, et al, 2024a).[11] The following sections illustrate 2 quite distinct traditions, the second having emerged as an alternative to the first.[12]

---

[11] Part of the reason why it is poorly theorised may relate to the fact that proficiency modelling is very demanding, in a multidisciplinary sense. As illustrated in FAB (2017, page 97), it is partly a philosophical activity (conceptualising the general structure of the target proficiency), it is partly a psychological activity (identifying the particular features of the target proficiency), and it is partly an exercise in semiotics (communicating the essential characteristics of the target proficiency).

[12] More detailed accounts of these traditions can be found in Newton, Curcin, Clarke, & Brylka (2024b) and Newton, Clarke, Curcin, & Brylka (2024). Over the past decade or so, there has been a notable international 'shift to learning outcomes' (Cedefop, 2009). In fact, Cedefop has recently observed that: "almost all EU Member States are now actively using learning outcomes (or competence) statements to define, review and refine their qualifications, VET curricula and programmes." (Cedefop, 2025, p.14). Some countries have attempted to bolster their traditional

Syllabus content specifications

Traditionally, certification requirements for public exams in England – O levels and A levels, for instance – have been articulated within syllabus documents. For many decades, until at least the 1970s, these documents were minimally specified, primarily comprising lists of topic areas to be covered by the teacher.[13]

Figure 3. Cambridge 1957 O level physics syllabus

Syllabus	Notes
1. Measurement of length and of volume.	Both f.p.s. and c.g.s. systems are expected. Candidates will not be asked to *describe* a vernier or a screw-gauge, but may be expected to use them in the practical examination.
2. Measurement of time by use of the simple pendulum.	A knowledge of the formula relating periodic time to length of the pendulum will not be expected; if required in the practical examination, it will be given.
3. Densities of solids and liquids.	Experimental determination of densities, e.g. by density bottle or by weighing and use of a measuring cylinder, is expected.
4. Pressure in liquids and gases; transmission of fluid pressure; the hydraulic press.	Quantitative formulae required.
5. Boyle's Law.	Experimental demonstration for air is included.

'syllabus content' approach by adding detailed 'learning outcome' statements. Debate continues over the extent to which the 2 approaches are intrinsically complementary or in tension with each other (see Winch, 2023). Attempts to bolster traditional approaches by adding outcome statements have not always been successful. This area deserves to be theorised far better than it currently is.

[13] These documents tend to be written in far greater detail nowadays (see, for example, Annex 1 from Newton, Curcin, et al, 2024b).

Figure 3 reproduces an extract (listing just 5 items) from the Cambridge 1957 O level physics syllabus, which was 9 pages long, listing 79 items of content in total alongside notes on the scope of each item (typically a sentence for each one).[14]

For a traditional exam syllabus like this, there would have been no explicit proficiency model. Instead, the target proficiency would have remained implicit: partly within the syllabus content list, and partly within each successive exam paper (although textbooks also played an important role). This made it hard for teachers to be confident that their implicit understanding of the target proficiency was the same as the examiner's, particularly during the early years of a new syllabus when there would have been few if any past papers to consult.

Things began to change during the 1980s, making qualification requirements more transparent for teachers and students. Syllabuses began to incorporate aims, assessment objectives (with allocated mark proportions), descriptions of standards at key grades, and so on (Kingdon & Stobart, 1988). These new features began to make GCSE and A level target proficiencies far more explicit than they had been in previous decades.

Learning outcome specifications

Outcome-based qualification design was introduced as an antidote to the ambiguity of the syllabus content tradition. Instead of leaving the target proficiency largely implicit, the principle underpinning this new tradition recommended comprehensive explication, in terms of both breadth and depth. As such, outcome-based qualifications typically incorporate a more explicit proficiency model that specifies learning outcomes for each unit (which explicate the breadth of the target proficiency) and assessment criteria for each learning outcome (which explicate its depth).

[14] It also provided a description of the practical exam, which would 'test whether the candidates have worked through a satisfactory course in the laboratory and are capable of handling simple apparatus' (UCLES, undated, page 37).

Figure 4. Example of how units are specified within CASLO qualifications (criteria only displayed for outcomes 2 and 5)

Learning Outcomes	**Assessment Criteria**
The learner will:	**The learner can:**
1. Interpret the given information relating to the work and resources when applying finishing plaster to prepared surfaces.	
2. Know how to comply with relevant legislation and official guidance when applying finishing plaster to prepared surfaces.	2.1. Describe their responsibilities regarding potential accidents, health hazards and the environment whilst working: in the workplace, below ground level, in confined spaces, at height, with tools and equipment, with materials and substances, with movement/storage of materials and by manual handling and mechanical lifting. 2.2. Describe the organisational security procedures for tools, equipment and personal belongings in relation to site, workplace, company and operative. 2.3. Explain what the accident reporting procedures are and who is responsible for making reports.
3. Maintain safe and healthy working practices when applying finishing plaster to background prepared surfaces.	
4. Select the required quantity and quality of resources for the methods of work to apply finishing plaster to prepared surfaces.	
5. Minimise the risk of damage to the work and surrounding area when applying finishing plaster to prepared surfaces.	5.1. Protect the work and its surrounding area from damage in accordance with safe working practices and organisational procedures. 5.2. Minimise damage and maintain a clean work space. 5.3. Dispose of waste in accordance with current legislation. 5.4. Describe how to protect work from damage and the purpose of protection in relation to general workplace activities, other occupations and adverse weather conditions.

Learning Outcomes	**Assessment Criteria**
The learner will:	**The learner can:**
	5.5. Explain why the disposal of waste should be carried out safely in accordance with environmental responsibilities, organisational procedures, manufacturers' information, statutory regulations and official guidance.
6. Complete the work within the allocated time when applying finishing plaster to prepared surfaces.	
7. Comply with the given contract information to apply finishing plaster to prepared surfaces to the required specification.	

Outcome-based qualifications began to take root in the landscape of vocational and technical qualifications in England during the 1970s and into the 1980s (Newton, Clarke, et al, 2024). The approach that became widespread in England stated that qualifications should be specified in terms of units with both learning outcomes and assessment criteria, plus a requirement that students would need to achieve all specified learning outcomes to pass a unit, and pass all units to pass the qualification. We have referred to this as the CASLO approach, because these qualifications all confirm the acquisition of specified learning outcomes (Newton, Curcin, Clarke, & Brylka, 2024b). An example of a CASLO unit specification is provided in Figure 4.[15]

The fundamental goal of an outcome-based approach is alignment: to ensure that plans for teaching, learning, and assessment are derived from the same specification of intended learning outcomes.[16] The detailed specification of learning outcomes helps to ensure that:

- assessors assess everything that they need to assess
- teachers teach everything that they need to teach
- learners learn everything that they need to learn, and that
- assessment, teaching, and learning are aligned to the same proficiency model

Qualification structure

Decisions concerning the overarching qualification structure will tend to be taken alongside decisions concerning syllabus outcomes. Perhaps the most fundamental structural decision concerns whether the qualification is going to follow more in the syllabus content tradition, more in the learning outcome tradition, or somewhere in between. As just explained, this will have direct consequences for the nature of the proficiency model, but it will also have direct consequences for the nature of the assessment procedure. For instance, if it were decided to adopt the CASLO approach, then this would necessitate exhaustive assessment, that is, it would require the assessment of all specified learning outcomes. This, in turn, would most

[15] The details of this example were taken from an actual regulated qualification (circa 2024): the NOCN_Cskills Awards Level 2 NVQ Diploma in Plastering (Construction) – Solid (603/2368/1). Its Total Qualification Time was specified as 870 hours (which means that a learner might reasonably be expected to complete the qualification in this much time from a plausible baseline of prior attainment).

[16] Ironically, one of the classic criticisms of outcome-based qualifications is that this detailed explication of learning outcomes tempts teachers and assessors to think that there is no further need to plan teaching, learning, and assessment, as though the unit specifications already constituted those plans (which is almost the exact opposite of what this approach is meant to convey).

likely mean devolving substantial responsibility for assessment to teachers or trainers working locally, within schools, colleges, or in the workplace (Newton & Lockyer, 2022).

Important structural decisions might include:

- whether to adopt a modular (unitised) or linear approach to teaching, learning, and assessment
- whether to award credit for partial completion, to support credit accumulation, credit transfer, accreditation of prior learning, or suchlike
- whether to provide a nested qualification structure, which would allow students to begin studying for a smaller qualification with the potential to scale it up to a larger one by adding units (or to begin studying for a larger qualification with the potential to scale it down to a smaller one if needs be)
- the need to establish structural relationships with other qualifications, particularly where the qualification is intended to be part of a formal progression route
- whether to tailor the design of the qualification to a particular instructional philosophy (such as learner-centred education), or to a particular instructional approach (such as problem-based learning)
- the amount of time that will be available to the typical learner for teaching, learning, and assessment (guided learning hours, total qualification time)
- expectations (or requirements) concerning the breadth and depth of prior learning

Any decision related to qualification structure would, of course, follow directly or indirectly from prior (stage 1) decisions and the prioritised purpose profile.

Assessment procedure

Just as for decisions related to syllabus outcomes and qualification structure, decisions related to the assessment procedure ought to follow directly or indirectly from prior (stage 1) decisions and the prioritised purpose profile, with due regard to systemic requirements and anticipated pragmatic constraints. All sorts of decisions will need to be taken concerning, for example, the:

- elicitation of assessment evidence (evidence requirements, approaches to sampling and coverage, location, authentication, task formats, task controls, response formats, reasonable adjustments, evidence recording, recognition of prior learning)
- evaluation of assessment evidence (assessor credentials, assessor training, approach to judging evidence, assessment criteria, location, quality assurance of judgements, judgement recording)

- combination of assessment judgements (aggregation approach, standardisation techniques, result categories)
- communication of assessment results (result labels, certificate design, qualification guides)
- rectification of issues arising (appeal processes, special consideration, re-assessment opportunities)

Fitness for purpose

The idea of purpose analysis emphasises the importance of adopting an expansive view of fitness for purpose. We have already noted that fitness for purpose is Ofqual's overarching technical requirement for regulated qualifications. Condition D1 of our General Conditions of Recognition goes on to explain that:

> D1.2 A qualification will only be fit for purpose if that qualification, as far as possible, secures the requirements of – (a) Validity, (b) Reliability, (c) Comparability, (d) Manageability, and (e) Minimising Bias.

Although, nowadays, many scholars and practitioners would be happy to use the term 'validity' as an umbrella concept – an overarching framework for organising a full range of technical criteria such as reliability, comparability, bias, and so on – there is no consensus concerning exactly how encompassing this framework ought to be. More specifically, there are passionately held views – both for and against – concerning the wisdom of stretching the concept of validity beyond merely technical considerations to embrace wider societal ones related to the acceptability of running the qualification or qualification system, which would take into account a full range of intended and unintended consequences (see Newton & Baird, 2016).

In previous reports, Ofqual has suggested that it is helpful to think of qualification validity as something like: assessing the right thing, in the right way, to provide accurate and useful assessment results (Newton, 2017b, page 6). This is a reasonably broad conception, which stretches to incorporate the specification of syllabus outcomes, qualification structure, and assessment procedure, recognising that these aspects are closely interlinked. Indeed, the specification of syllabus outcomes is essentially the first step in specifying an assessment procedure (Newton, 2017b, page 21). In other words, we need to establish exactly what it is that we want to measure before designing a procedure for measuring it.

While this is a reasonably broad conception, it still leaves space for considerations that are critical to purposeful qualification design but that lie beyond the scope of validity, including the consideration of constraints related to financial costs, political commitments, laws, and so on (Newton, 2017b, page 65). So, while D1.2 foregrounds the more technical aspects of fitness for purpose, any comprehensive

evaluation would need also to incorporate more expansive judgements. The remainder of this report provides a more comprehensive account of the sorts of considerations that need to be borne in mind whenever a new qualification or qualification system is being planned.[17]

Purposeful design

The underlying principle at work here is that qualification design decisions should never be arbitrary, nor unthinkingly default to longstanding traditions of custom and practice, but must always be consciously tailored to a bespoke, prioritised profile of intended purposes. In other words, qualification design must always be purposeful. Consequently – before anything related to qualification structure, syllabus outcomes, or the assessment procedure is officially decided – we need to scrutinise and unpack implications related to:

- the targeted cohort (and their circumstances)
- intended purposes
- systemic requirements
- anticipated constraints

This comprises the first stage of qualification design: purpose analysis. The following sections explain these 4 components of purpose analysis in turn.

[17] While this report explains the sorts of judgements that are in scope, it does not go into detail concerning the nature of the evaluative criteria that would need to be brought into play when making such judgements.

Drivers and constraints

Qualification design is a complex process: driven by a convoluted network of purposes and requirements, bounded by a convoluted network of constraints, and always needing to be targeted at a specific cohort of learners. The following sections will explain and illustrate the sorts of cohorts, purposes, requirements, and constraints that need to be considered when designing a qualification. This will involve distinguishing between perspectives on qualification purposes (on the one hand) and systemic requirements for qualification design (on the other), each of which will be subdivided into various categories and sometimes sub-categories. These perspectives and requirements need to be considered simultaneously when designing a qualification within a broader qualification system.

It is important to consider **purpose perspectives** to ensure that a new qualification will function effectively as a qualification. These considerations are framed primarily in terms of its target cohort. Respectively, they concern the:

- **expertise** that needs to be acquired (by learners) to help prepare them for the demands of life and work
- **engagement** that needs to be exercised (by learners but also by their teachers) to help maximise the quality and quantity of the learning that occurs
- **information** (about learners) that qualification results need to provide to help those who use qualification results to make better decisions, that is, better decisions than they would be able to make without that information[18]

We might even say that it is of the nature of a (worthwhile) qualification that it:

- directs the acquisition of (valuable) expertise by learners
- secures (sustained and effective) engagement by learners and by their teachers
- generates (accurate) information about learners' attainments

We refer to these as distinct 'perspectives' on qualification purposes because they each answer the question of what a qualification needs to do from a different vantage point, to establish that the:

- body of learning that is acquired while studying for the qualification is as valuable as possible (for learners, for the economy, and for society) – expertise

[18] Qualification results are used in many ways by many different users: teachers, students, administrators, selectors, and so on. The important point is that these results are used to make decisions, and different decisions often require different kinds of information.

- features and processes that comprise the qualification actively facilitate uptake, learning, completion, continuation (and so on) – engagement
- information that qualification results provide is of the right kind and of sufficient accuracy (to support effective decision-making) – information

Each of these perspectives is important to the design of any qualification. Yet their design implications sometimes conflict with each other, forcing us to choose between them. Hence the importance of prioritising them from the outset. To animate this idea, it can be helpful to think of these perspectives being championed by different varieties of technical expert, for example:

- an expert in the subject area – whose bottom line is the value of acquired learning outcomes – who focuses on ensuring that syllabus outcomes specify exactly what they need to (at least, exactly what the subject expert thinks they need to)
- an expert in teaching practices – whose bottom line is the quality and quantity of learning – who focuses on ensuring that students are motivated to choose the qualification, put the required effort into studying it, and subsequently remain positively disposed to continue on a learning journey, as well as on ensuring that teachers teach as effectively as they ought to
- an expert in assessment – whose bottom line is the accuracy of measurement – who focuses on ensuring that the information that results provide is as accurate as it needs to be to guarantee the effectiveness and defensibility of (result-based) decision-making

Debates concerning how best to design a new qualification sometimes do involve these exact stakeholders arguing from very different perspectives. For instance, a subject expert might insist that the assessment procedure must incorporate all the valued learning outcomes – to ensure they actually get taught and learnt. Whereas an assessment expert might respond by insisting certain critical outcomes must not be assessed – where they demonstrably cannot be measured with sufficient accuracy. The point is simply that their respective professional concerns may orient them towards different design characteristics and bottom lines.

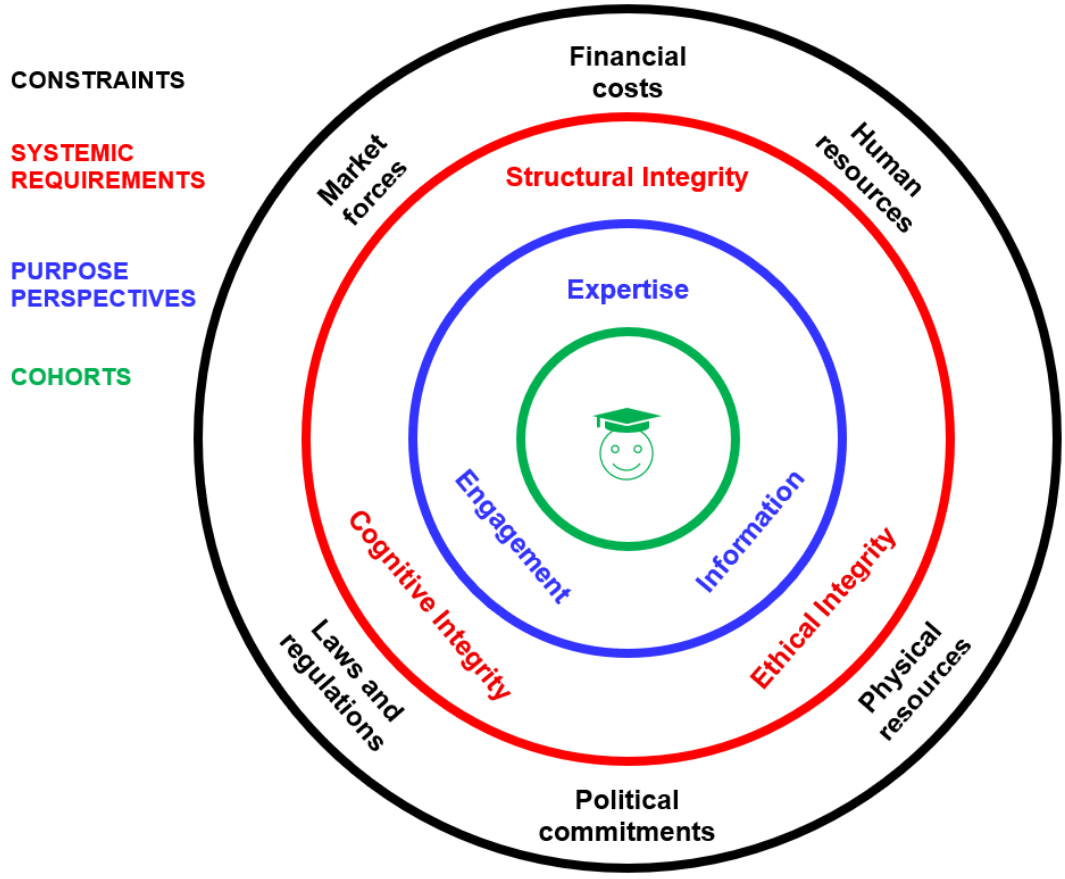
Beyond these purpose perspectives, it is also important to consider **systemic requirements**, which are important to the effective operation of the qualification system as a mechanism for organising society itself. These considerations are framed in terms of the integrity of the system:

- **structural** integrity – to secure the system itself – including the need for qualifications to relate to each other coherently, and to be resilient to large-scale external crises (pandemics, cyber attacks, and suchlike)

- **cognitive** integrity – to secure public understanding and participation – including the need for qualification practices to be transparent and plausible
- **ethical** integrity – to secure public confidence and consent – including the need for processes and products to be robust (that is, resistant to subversion) and just (that is, fair and seen to be fair)

Systemic requirements embody a generic societal perspective on how the system and its qualifications ought to operate (as a mechanism for organising society). This recognises that qualifications play a fundamental role in distributing job opportunities (when students complete their formal education and transition into the workforce) as well as in distributing opportunities for increasingly specialised education or training (before they transition into the workforce). Because qualifications play such a fundamental role in organising society, they need to have integrity, and they need to be seen to have integrity. These requirements are viewed most clearly, or vividly, through the eyes of those at the centre of the system: those studying for qualifications and (particularly for younger learners) their parents too.

Figure 5. Cohorts, perspectives, requirements, and constraints



In the following sections, we will work outwards from the target cohort to consider purpose perspectives, systemic requirements, and constraints, respectively. These 4 categories are illustrated in Figure 5.

1. Cohorts (within circumstances)

A qualification that is designed to be fit for one particular group of learners (and the circumstances of their teaching, learning, and assessment) will not necessarily be fit for another group (and their circumstances). For example, a school-leaving certificate in Welsh that has been designed for first-language speakers is not going to be ideal for second-language speakers. This is why the WJEC exam board offers 2 separate qualifications: a GCSE in Welsh Language, and a GCSE in Welsh Second Language. Similarly, a technical qualification that has been designed for 16- to 19-year-olds studying full-time in college is not going to be ideal for 16- to 19-year-old apprentices who have full-time jobs. This is why the T Level programme has been developed independently of the apprenticeship route.

Unfortunately, decisions sometimes fail to take sufficient account of the needs of particular cohorts (in their particular circumstances). A good example of this was the Qualifications and Credit Framework (QCF), which operated from 2008 to 2015. It was heavily prescriptive, such that, to be accredited to this framework, all qualification units had to be designed in essentially the same way. Moreover, the intention was for the vast majority of regulated vocational and technical qualifications to be accredited to this framework by 2010. Concerns were raised from the outset concerning the suitability of a one-size-fits-all approach, and when Ofqual reviewed its operation, it agreed that the logic of this framework was not universally applicable (Ofqual, 2014). The framework did make a lot of sense for adults who wished to return to learning, often with highly divergent needs and low confidence.[19] Yet it made far less sense for other cohorts and the circumstances of their learning. So, we withdrew the framework, and we replaced it with a far less prescriptive one, the Regulated Qualifications Framework (RQF).

Learners differ in many ways, so we cannot assume that the target cohort for any new qualification will be entirely homogeneous. However, we can often assume a certain amount of homogeneity, sufficient to justify tailoring its design characteristics accordingly. For instance, cohorts might be sufficiently similar in terms of their:

[19] According to Wilson (2010), the origins of the credit system that inspired the QCF sprang directly from the needs of adult learners, which were best served by qualifications that could be achieved in small steps, unit-by-unit, with those achievements formally recognised and celebrated through the award of credits. This assumed that gradual reengagement with learning was best facilitated by establishing a mix-and-match approach to assembling qualifications from units, enabling learners to personalise the trajectory of their learning to their personal interests and needs.

- motivation (from highly motivated to completely disaffected)
- self-esteem (from very high to very low)
- current interests or ambitions (including whether these are largely academic or tending towards vocational)
- education or work status (such as mainstream 14-16 education, mainstream 16-19 education, 16-19 training, 19+ unemployed, 19+ employed, home educated, prison educated, correspondence)
- prior attainment (especially in relation to cohort-level prior assessments, such as national curriculum assessments, or GCSE maths and English)
- special educational needs or disabilities (from few to many and varied)
- duration of study (days, weeks, months, years)
- mode of study (such as full-time, part-time, block release, sandwich)
- location of study (such as school, further education college, university, workplace, combination of institution and workplace, evening school, club)
- funding status (government-funded, employer-funded, self-funded)
- nationality (home student, overseas student)
- socio-economic background (based on factors like family income, parental qualifications, or suchlike)
- personal commitments (such as concurrent employment, caring for dependents)

The implication here is that a qualification designed specifically for (say) low-attaining, disaffected, prison educated students is likely to have quite different characteristics from a qualification that is designed specifically for high-attaining, highly motivated, mainstream 16 to 19 students. And both might look quite different from a qualification that is designed specifically for middle-achieving, averagely motivated, part-time, self-funded, 19+ employed, evening school students.

It is particularly important to distinguish between target cohorts in terms of their motivations, as differently motivated learners will need to be engaged in different ways and to differing extents. For instance, we might distinguish between:

- compliant learners – where the engagement goal might be to facilitate and sustain an even higher level of enthusiasm for (and autonomy over) their learning
- indifferent learners – where the goal might be to ignite a desire to engage proactively with their course
- disengaged learners – where the goal might be to help them to overcome a significant obstacle to engaging with their course

- fragile learners – where the goal might simply be to create and sustain conditions that give them a reasonable chance of being able to complete their course

When it comes to disengaged learners, there is a significant problem to be overcome, and it behoves the qualification designer to consider whether it would be profitable to attempt to help address this via qualification design features (as we will discuss in more detail below). The engagement needs of fragile learners would seem to be of a different order again. They are nicely encapsulated within this quotation from a Community Learning Provider, in a report on the benefits of Personal and Social Development (PSD) provision:

> We have chaotic learners – the idea of getting out of bed in the morning and putting themselves together is a fundamental achievement. The only time they come out the house is just to go to the jobcentre and our course. We've got lots of case studies where they have begun to turn up consecutively and started talking to others. It may be drugs, alcohol, crime; we break the cycle. In terms of value for money, it stops people going to the GP twice a week. How much is that actually worth?

(Gray & Lockhart-Smith, 2015, page 46)

Gray & Lockhart-Smith also explained the value of the PSD certificate itself. This is not simply its motivational value – bearing in mind that some of these learners may never before have aspired to achieve a national qualification – but also its value in signalling that the learner has been able to demonstrate sufficient commitment to be able to complete the course successfully. Recognising that PSD learners have very different needs, ambitions, and problems, providers in their study highlighted flexibility (in the delivery of PSD units) as an important feature of PSD qualifications.

2. Purpose perspectives

As noted earlier, it is important to consider purpose perspectives to ensure that the new qualification will function effectively as a qualification, which concerns its ability to:

- direct the acquisition of (valuable) expertise by learners
- secure (sustained and effective) engagement by learners and by their teachers
- generate (accurate) information about learners' attainments to support decision-making

All 3 perspectives are relevant to the design of all qualifications, but certain perspectives are likely to be more important than others in relation to certain ones. Because different purposes often hold somewhat different implications for qualification design, it is important to prioritise between them from the outset.

Consider, for instance, a distance swimming certificate, which is awarded for being able to swim 600 metres. Which of the 3 perspectives – information, expertise, or engagement – ought to be deemed the most and least important in this context?

Well, to some extent, it must be important that the information provided by the certificate is accurate. In other words, it should only be awarded to learners who have genuinely demonstrated their ability to swim the certificated distance. Having said that, distance swimming certificates are rarely (if ever) used to make important real-world decisions. So, the information perspective is unlikely to be the highest priority for a distance swimming award. In other words, if the assessment process were to be a little error prone – for example, if busy examiners were prone to occasional lap-counting errors – then this would probably not matter too much.

What really seems to matter the most for a distance swimming award like this is that the learners who train for the award end up acquiring a body of learning with genuine real-world value, that is, the ability to swim a significant distance.[20] So, the expertise perspective must surely be a high priority for a distance swimming award.

Of course, learning to swim requires dedication, and part of the rationale for establishing a distance swimming award (at all) is to provide a goal for learners to aim at. This invokes the engagement perspective. Indeed, in this example, the engagement perspective might well be judged to have a higher priority than the information perspective. The fact that distance swimming certificates are frequently accompanied by material badges – which are worn with pride on tracksuits and swimsuits – is evidence of the engagement perspective at work, and the desire to motivate young people to learn how to swim.

The following sections provide a comprehensive analysis of these 3 purpose perspectives and their implications for qualification design. Before beginning this analysis, it is worth revisiting the sorts of design decisions that we are talking about, which relate either to the syllabus outcomes or to the assessment procedure or to the overarching qualification model (with implications for both syllabus and assessment). For instance, from the engagement perspective, we might aspire to maximise student motivation for a school-based qualification by specifying content that would seem to be interesting for a particular cohort of students (a syllabus-related decision), or by specifying a grading scale that provides stretch and challenge for high achievers (an assessment-related decision), or by dividing the qualification into separately certificated units that can be achieved in steps (a model-related decision).

---

[20] This example helpfully illustrates that the expertise perspective concerns a trajectory of proficiency from novice through to expert, which means that the body of learning associated with a low-level qualification is inevitably going to be low level. In this example, it refers simply to the ability to swim a specified distance, regardless of stylistic elegance, time taken, or suchlike.

It is tempting to assume that information-related design decisions relate purely to the assessment procedure, while expertise-related decisions relate purely to syllabus outcomes. In fact, the situation is slightly more complicated than this. For instance, we know that, whatever the syllabus might say, the nature of the assessment procedure will also direct the acquisition of expertise. We tend to refer to this as the **backwash impact** of assessment on the curriculum, which can result in a difference between the enacted syllabus and the intended one. This would generally be seen as a negative impact, resulting from **construct underrepresentation** in the assessment procedure. In other words, if the assessment procedure consistently fails to assess certain elements of the construct (from one administration to the next) then this risks teachers no longer teaching those elements and students no longer learning them.[21]

Expertise-related purposes

In order for the economy and for society to flourish, it needs its young to develop and hone certain abilities and tendencies – knowledge, skills, competencies, dispositions, attitudes, and suchlike – some of which will be required by all of its young (as a foundation for everyday life and work) and some by only a subset (as a foundation for specific occupational pathways). We call this the expertise perspective because it relates to an anticipated trajectory of learning – from having less expertise to having more of it – which is typically associated with studying an education or training course (or studying a succession of courses over an extended period).

Decisions that stem from the expertise perspective typically concern the specification of syllabus outcomes. They aim to ensure that the qualification will direct the acquisition of a body of learning that is both intrinsically valuable as a foundation for life and work and extrinsically valuable in the sense of underpinning the purchasing power of the certificate in the job market (although qualifications will differ in terms of the extent to which they aim for both intrinsic and extrinsic value).

The process of specifying syllabus outcomes is likely to begin by reflecting on what the learner might wish to do with the body of learning that is certified by the qualification (or what society or certain organisations might wish them to do with it). For instance, they might need it as a foundation for further learning in a subject area,

---

[21] In assessment terminology, the ‘construct’ relates to the body of learning that needs to be assessed. So, the meaning of ‘construct underrepresentation’ is that the assessment provides insufficient information about learners’ attainments in relation to a subset of that learning. Where this is a feature of the assessment procedure itself – more specifically, where this is a feature of the ‘blueprint’ from which each new assessment is developed – this construct underrepresentation will be predictable from one assessment to the next, and this can cause the enacted syllabus to be narrower than the intended one (where teachers have stopped teaching those outcomes and learners have stopped learning them).

or they might need it to perform certain real-world activities. Subsequently, it would be important to reflect on the nature, content, and balance of the elements of knowledge, skill, competence, disposition, attitude, or suchlike that would be required to facilitate this. Finally, the process would involve deciding how best to represent the anticipated body of learning – for the purposes of teaching, learning, and assessment – which would include deciding whether the specification ought to be framed more in terms of syllabus content or more in terms of learning outcomes. Note that this would also require decisions concerning the degree of detail provided and the extent to which elements were explicated or implicated.[22]

The sort of person who might champion the expertise perspective is likely to be a curriculum specialist or a subject domain representative – in other words, a stakeholder with a professional or voluntary interest in what gets taught and learnt throughout the country, either generally (curriculum) or more specifically (subject domain). Ultimately, these stakeholders are advocating on behalf of individuals, organisations, and society as a whole, in order for the entire system to flourish. Having said that, this advocacy will be channelled through the lens of their own particular vision of the 'good life' or the 'good society' – and sometimes also through the lenses of visions held by professional bodies or companies that they may belong to – so we would not necessarily expect these specialists and representatives to agree with each other.

Curriculum purposes

Expertise-related purposes derive ultimately from **educational purposes**, which might also be described as **curriculum purposes**. At the highest level of analysis, differences of opinion over purposes of this sort stem from entirely different ideologies. Ideological stances can be categorised in different ways, for example, Adamson & Morris (2007) identified 6 distinct stances, contrasting their curriculum intentions as follows:

- academic rationalism – to enhance learners' intellectual capabilities and cognitive skills, and to teach them how to learn
- social and economic efficiency – to provide for the current and future human capital needs of a society
- social reconstructionism – to promote and facilitate social reform, change and criticism

[22] This presumes that there must be at least some explication and detail provided via syllabus content or learning outcomes, that is, there must be some central prescription of what to teach and learn. This is a prerequisite for any qualification that certifies attainment (as opposed to attendance).

- orthodoxy – to induct learners into a particular religious or political orthodoxy
- progressivism – to provide learners with opportunities for enhancing their personal and intellectual development
- cognitive pluralism – to provide a wide range of competencies and attitudes

Different ideologies capture different intentions regarding how learners ought to change as a result of their educational experiences. Offering a slightly different categorisation, Wiliam (2013) proposed that these intentions can normally be grouped into 1 of 4 categories of learning:

- personal empowerment – to allow young people to take greater control of their own lives
- cultural transmission – to transfer received wisdom from one generation to the next
- preparation for citizenship – to empower learners to actively participate in a democratic society
- preparation for work – to ensure that individual workers and the organisations they work in are productive and prosperous

Each of these categories is premised upon guided learning of one sort or another, involving the cultivation of knowledge, skills, competencies, dispositions, attitudes, and suchlike. Yet different stances lead to different recommendations concerning how to focus this learning. For example, an academic rationalist stance is likely to focus on the acquisition of knowledge, skills and values derived from academic disciplines. Whereas an orthodox stance is likely to focus on the acquisition of beliefs and practices associated with a particular orthodoxy (Adamson & Morris, 2007). Note that both Wiliam and Adamson & Morris emphasised how education systems are rarely based upon a single stance, and the combination of stances that supports any particular system is therefore likely to incorporate principles that are in tension with each other, which underscores the importance of trade-off and compromise (see also Williams, 1958). We will return to this challenge shortly.

Unfortunately, there is no universally agreed scheme for classifying educational or curriculum purposes. Many different schemes have been proposed in the past, and alternative schemes continue to be proposed to the present day (see, for example, Schiro, 2008; Biesta, 2010; Wheelahan & Moodie, 2025).

Figure 6. Syllabus purpose categories from Ofqual (2009)

Main purpose		Sub-purpose	Examples of possible qualifications
A. Personal growth and engagement in learning	A1.	Develop skills for life	Functional skills qualification Skills for Life qualification
	A2.	Develop knowledge and/or skills to operate independently and effectively in life, learning and work	Qualification in personal and social development Qualification in personal effectiveness
	A3.	Develop personal skills and/or knowledge	Qualification in music
	A4.	Develop employability skills and/or knowledge	Qualification in employability
B. Prepare for further learning or training and/or develop knowledge and/or skills in a subject area	B1.	Lay foundations for acquiring more specialised learning or training	General Certificate of Secondary Education Access to Higher Education diploma Pre-U qualification
	B2.	Lay foundations for acquiring deeper knowledge and/or skills in a subject area	Qualification in art and design Qualification in classical languages
C. Prepare for employment	C1.	Prepare for employment in a broad occupational area	Qualification in economics or business Qualification in performing arts or media
	C2.	Prepare for employment in a specific occupational area	Qualification for teaching assistants Qualification in hairdressing
D. Confirm occupational competence and/or 'licence to practice'	D1.	Achieve full competence in an occupational role to the standards required	National Vocational Qualification Other competence-based qualification
	D2.	Achieve requirements for a 'licence to practice' or other legal requirements made by the relevant sector, professional or industry body	Qualification in door supervision Qualification in accountancy
E. Updating and continuing professional development (CPD)	E1.	Update knowledge and/or skills relating to legal, technical, process or best practice changes/requirements	Qualification in infection control Qualification in health and safety
	E2.	Develop knowledge and/or skills in order to gain recognition at a higher level or in a different role	Qualification in operational or strategic management
	E3.	Develop knowledge and/or skills relevant to a particular specialisation within an occupation or set of occupations	Specialist qualification in the health sector Specialist qualification in food safety

Syllabus purposes

Syllabus purpose categories operate at a lower level than educational or curriculum ones, because they relate to individual qualifications, or types of qualification, rather than to an overarching ideological stance concerning the very meaning of education or training. Figure 6 illustrates a scheme for categorising syllabus purposes that was produced by Ofqual (2009) to help distinguish between different types of qualification within the Qualifications and Credit Framework.[23] Ofqual kept this scheme under review, recognising that it had limitations. Ideally, the categories within a framework of this sort ought to be as mutually exclusive and collectively exhaustive (MECE) as possible, but this seemed to fall short on both counts.

In 2015, we attempted to develop a more streamlined scheme.[24] This proposed the following 6 categories:

> Occupational proficiency: An occupational proficiency qualification is predominantly associated with acquiring the knowledge, skill and understanding that enables an individual to perform all or part of an occupational role. The defining feature of these qualifications is that they certify the ability to carry out a particular function within an occupational role, or a full range of functions associated with a role.
>
> Vocational readiness: A vocational readiness qualification is predominantly associated with acquiring the underpinning knowledge, skill, and understanding that constitutes a solid foundation for developing occupational competence. The defining feature of these qualifications is that they indicate readiness, or potential, to acquire the knowledge, skill and understanding that enables an individual to perform a particular occupational role. They will always have a strong vocational context, which enables them to function as direct 'stepping stones' into an occupation.
>
> Sector awareness: A sector awareness qualification is predominantly associated with acquiring important insights into the world of work. The defining feature of these qualifications is that they recognise awareness of the roles and functions associated with occupations within sectors. They will always have a strong vocational context. They may include some of the underpinning knowledge, skill, and understanding that constitutes a solid foundation for developing occupational competence; yet, not enough to enable them to

[23] Note that some of the descriptions of the sub-purposes (column 2) have been adapted slightly, to emphasise that they all relate to types of learning ('expertise') of one sort or another.

[24] A proposal for a revised scheme was discussed at Ofqual's Vocational Advisory Group, 10 July 2015.

function as direct ‘stepping stones’ into an occupation. They may well act as ‘stepping stones’ into vocational readiness qualifications.

Scholastic attainment: A scholastic attainment qualification is predominantly associated with acquiring the knowledge, skill and understanding that is integral to a discipline or domain of inquiry. The defining feature of these qualifications is that they indicate a level of attainment in relation to a curriculum area or similarly defined body of knowledge, skill and understanding.

Functional capacity: A functional capacity qualification is predominantly associated with acquiring the knowledge, skill and understanding that is fundamental to effective functioning in everyday life. The defining feature of these qualifications is that they certify the ability to engage in activities that are central to everyday life and work. No distinction is drawn between qualifications at varying levels of functional capacity, from basic level competence to high level expertise.

Personal achievement: A personal achievement qualification is predominantly associated with acquiring confidence, self-esteem, educational and social integration, and suchlike. The defining feature of these qualifications is that they recognise personal achievement in an area of learning. The content domain (whether vocational or otherwise) is often considered to be of secondary importance. In fact, even the specific set of learning outcomes that is explicitly associated with the qualification may be secondary in importance to the personal achievement that it implicitly facilitates, especially as the sense of achievement will be relative to the learner’s unique starting point. These qualifications may include some of the underpinning knowledge, skill and understanding that constitutes a solid foundation for studying for another kind of qualification, although not necessarily enough to enable them to function as direct ‘stepping stones’ into further study.

Once again, though, concerns over whether these categories were sufficiently MECE led to this scheme not being formally adopted. We concluded that the qualifications we regulate are too divergent to be captured adequately by a high-level scheme of this sort. Although we still lack a formal scheme for classifying syllabus purposes, the 2 schemes just discussed still have some heuristic value.

Contrasting opinions

While both curriculum specialists and subject domain experts will contribute to debates concerning expertise-related purposes, they will typically be making slightly different kinds of contribution. So, while a curriculum specialist might advocate for a particular approach to learning right across the curriculum, a subject expert is more likely to advocate for a certain position on the nature of valuable learning within their own particular domain.

Even at the level of syllabus purposes, there may still be highly contrasting opinions on the nature of valuable learning, depending on the perspective from which each protagonist views the debate. Consider, for example, what might count as valuable learning for a post-16 vocational qualification that supports progression into the workforce. From the perspective of a single local employer – for example, a window installer specialising in wooden frame installation – their interests would be served best if local students were able to study a qualification that focused exclusively on fenestration, especially if it could be customised to wooden frame installation. Conversely, a construction industry body representative might feel that the industry would be better served by a far less specialist construction qualification, which provided a broad foundation of knowledge, skill, and understanding in the construction domain, without bringing students close to occupational competence in any specific area. Finally, a national skills advisor might argue (based on relevant evidence) that most post-16 learners on vocational qualifications of this sort tend not to remain in the sector that they initially transition into. What would be of most value to students like this, the advisor might argue, is a far broader preparation for life and work, which incorporates both general and vocational learning outcomes. Bearing in mind that there will always be differences of opinion concerning the nature of valuable learning for any particular qualification, debate of this sort is inevitable and will need to be resolved, whether communally (by consensus building) or ultimately by diktat (from a policy owner).

When designing a qualification from the expertise perspective, the principal focus for activity is the specification of syllabus outcomes, to construct a proficiency model of one sort or another. This is how the valuable learning is represented for the purpose of teaching, learning, and assessment. Debate from this perspective will tend to revolve around the nature of the expertise at the heart of the qualification, and what syllabus outcomes to include or exclude. Ultimately, though, what really matters is whether the valued learning is actually acquired. This is where the expertise perspective begins to shade into the engagement one.

Engagement-related purposes

For learners to maximise their learning, they will need to exert the required amount of time and effort, as will their teachers. And the strategies that teachers and learners employ (for teaching and learning respectively) will need to be as effective as possible. We call this the engagement perspective because it concerns how learners and teachers actively engage with learning and teaching – both how engaged they are and how they engage with it, that is, how committed or involved they are and exactly how they roll it out or take part in it. Thus, the engagement perspective considers how syllabuses, assessment procedures, and qualification

structure can be designed to maximise the engagement of learners and their teachers. It is helpful to draw a crude distinction between:

- the **quantity** of learning that occurs – which can be related to how much time and effort is committed, which is essentially a motivational issue
- the **quality** of learning that occurs – which can be related to the effectiveness of teaching and learning, which is essentially a pedagogical issue

In other words, there are different routes to better engagement, one motivational and one pedagogical. However, it is not uncommon for pedagogical approaches to be framed with both quality and quantity of learning in mind. Learner-centred education, for instance, is a wide-ranging pedagogical approach that is often premised upon increasing student motivation (Cedefop, 2015).

The sort of person who might champion the engagement perspective is likely to be a specialist in pedagogy, or perhaps a representative for a group of disaffected learners – in other words, a stakeholder with a relevant professional or voluntary interest in this perspective. In championing it, they are advocating on behalf of all those who pass through our education and training systems, and therefore on behalf of society more generally.

We interpret the idea of motivating learners quite broadly, to encompass qualification design decisions that are intended to influence the quantity of learning that occurs in a variety of different ways, for example, by:

- incentivising uptake of a qualification
- securing consistent application of effort while studying for the qualification
- ensuring that students fulfil all necessary qualification requirements (to prevent drop out and failure)
- cultivating attachment to the particular domain of learning (to facilitate both current and future learning within that domain)
- instilling confidence in learning, per se, and commitment to future learning

Developing insights from a review of classroom assessment and learning by Crooks (1988) we note that there are all sorts of ways in which qualification design decisions can help to influence the quality of learning that occurs, for example, by putting in place features or processes that:

- focus attention on the most important aspects of the domain of learning, foregrounding its ‘big ideas’ and mapping optimal progression routes
- explicate, communicate, and reinforce expected learning goals and standards
- influence the choice of, and development of, the most effective teaching and learning strategies

- help teachers to determine whether learners have prerequisite knowledge and skill prior to introducing new material (potentially reactivating or consolidating it)
- provide repeated opportunities for practicing and consolidating new knowledge and skill
- elicit evidence of progress in learning, and gaps in learning, to enable teachers to identify the ongoing learning needs of individual students
- encourage students to monitor their own progress in learning, and to develop skills of self-evaluation and self-enhancement
- reward successful learning, to build confidence in their ability to learn

Of course, teachers might put any or all these pedagogical ideas into practice without having been influenced by qualification design decisions, in exactly the same way that learners might maximise their effort without having been influenced by qualification design characteristics. Yet qualification design has the potential to be used as a tool for helping to improve the quantity and quality of learning that occurs. Indeed, in certain situations, with certain cohorts of learners, this is an extremely important tool.

More fundamentally, we should never forget that qualification design decisions affect the quantity and quality of learning that occurs, either positively or negatively, whether we anticipate those impacts or not. So, it makes sense to attempt to anticipate where negative impacts might arise, and then to decide whether the risk is tolerable or whether it should be mitigated. Equally, it makes sense to consider where it might be possible to engineer positive impacts, and to decide whether active intervention of this sort would be worthwhile.

Teaching to the test

The phenomenon of **teaching to the test** helps to explain why the engagement perspective is not simply incidental to qualification design, but fundamental to it. It is often described in terms of teachers focusing their teaching on elements of a syllabus that are assessed consistently from one session to the next, while neglecting elements of the syllabus that are assessed only occasionally or not at all. This reminds us of the power of high-stakes assessment to drive, not just learning, but teaching too. It also indicates the potential of high-stakes assessment to undermine the effectiveness of teaching and learning in certain situations. When qualifications perform a gate-keeping function in the labour market, they have high stakes for learners, which is why the promise of certification is often assumed to be a positive motivating force in its own right. Qualifications also have high stakes for teachers when their results are aggregated and used for accountability purposes.

In an ideal world, teaching to the test would not be a problem. If qualifications comprehensively and authentically assessed their respective syllabuses, then this should incentivise comprehensive and authentic teaching and learning too. Yet for all sorts of reasons – including straightforwardly pragmatic ones related to cost and burden – qualifications are never assessed entirely comprehensively nor entirely authentically. Instead, sampling (rather than full coverage) is the norm and methods with limited authenticity (such as multiple-choice testing) are common.[25]

Again, when elements of a syllabus are rarely or never assessed, this provides a perverse incentive to focus teaching and learning on the consistently sampled syllabus outcomes at the expense of consistently unsampled ones (Frederiksen, 1984). Inauthentic methods can also be problematic when their inauthenticity makes them vulnerable to gaming strategies that enable candidates whose learning is actually very shallow to produce creditworthy performances (Shepard, 1997).[26]

Interestingly, the power of a qualification to drive teaching and learning also lies behind the use of qualification requirements to drive outcomes that may seem tangential, if not entirely unrelated, to the certified domain of learning. In the Republic of Ireland, for instance, a candidate who answers in Irish on written papers for various Leaving Certificate examinations may be given bonus marks – simply for answering in Irish. This practice was introduced during the 1920s to strengthen the position of the Irish language in the education system and to revive its use in everyday life (Mac Aog´ain, Miller, & Kellaghan, 2010). Relatedly, in 2011, a policy decision was announced in England concerning (the return of) the award of marks for spelling, punctuation and grammar, integrated into the marking criteria of GCSEs in subjects such as geography, history, and religious studies.

The phenomenon of teaching to the test reminds us that the engagement perspective must never be overlooked, as motivational and pedagogical factors are always at work (for better or for worse) whenever a qualification is being delivered.

---

[25] Even when learners are required to demonstrate the acquisition of all specified learning outcomes, the assessment of each outcome will still only sample a limited number of contexts (for demonstrating acquisition) from the set of all possible contexts (Newton, Clarke, et al, 2024).

[26] The idea of 'teaching to the test' can apply at various levels. At the curriculum level, subjects that are formally tested might receive more attention than those that are not. Likewise, at the syllabus level, subject areas or elements that are formally tested might receive more attention that those that are not. In terms of pedagogy, the concern is often the kind of attention that is given to areas or elements that are formally tested, in particular, the possibility that gaming strategies (such as learning model answers by rote) can lead to creditworthy test performances in the absence of robust learning. Each of these levels is relevant to the present discussion.

Motivational purposes

Every single day – in schools, colleges, universities, and workplaces all around the world – learners make conscious and unconscious decisions concerning how much effort they are going to put into the learning that is asked of them. Motivational purposes concern the will to learn, which can be understood in relation to incentives for engaging with education or training.[27] Results from a recent survey of students from state-funded secondary schools in England are relevant to this discussion (DfE, 2025). Students from years 7 to 11 were asked how motivated they are to learn and the results indicated that: 19% felt 'very motivated', 51% felt 'fairly motivated', 20% felt 'not very motivated', 7% felt 'not at all motivated' (and 3% responded 'don't know').

Keep (2009) has provided a fledgeling framework for classifying incentives to learn, distinguishing between internal and external ones.[28] Incentives that are **internal** to the education and training system relate to how attractive (or unattractive) it is for learners to learn, which might be influenced by:

- the intrinsic interest of syllabus outcomes
- approaches to assessment that downplay competition
- not tightly rationing access to desirable progression opportunities
- cultures that nurture potential and celebrate achievement

Incentives that are **external** to the education and training system relate to the perceived value of qualification certificates in the labour market, which might be influenced by:

- occupational returns to holders of certain qualifications (for example, wages, status, job satisfaction)
- labour market requirements for certain qualifications (for example, literacy and numeracy qualifications, licence-to-practice qualifications)

---

[27] This analysis concerns motivation for learning, as distinct from motivation for assessment. Motivation for assessment – more commonly known as test motivation – concerns the effort that is put into assessment tasks and the invalidity that arises from too little effort being exercised (for example, Wise & DeMars, 2005). The analysis is also distinct from concerns related to assessment anxiety, which can also undermine validity, and it is distinct from broader assessment-related impacts on wellbeing, such as an inability to sleep or to focus on everyday tasks (Polesel, et al, 2012).

[28] There is a large literature on motivation within educational settings, which includes different stances on the relative utility of internal and external drivers (for example, Ryan & Deci, 2020; Wang, Henry, & Degol, 2020; Remedios & Sewell, 2024; Hulleman, Totonchi, Barron, Sutter, et al, 2025). The implications of this literature for qualification design have not yet been comprehensively identified or evaluated.

Keep also noted the rise of another category: subsidy-based incentives, including grants or allowances that support learners to study particular qualifications.

The distinction between internal and external reminds us that a qualification that might be extrinsically motivating for a particular learner (owing to its market value) might simultaneously lack intrinsic motivation for them (owing to its particular content), with the latter potentially nullifying the former. Clearly, we might expect learners to be drawn most towards qualifications that they find both intrinsically and extrinsically motivating.

The following sections consider qualification design decisions that might be taken to improve motivation to learn, and thereby the quantity of learning that occurs. We identify features or processes that might, in theory, be motivating for learners, most of which have been implemented in England at some point in time. However, it is essential to recognise that there is no guarantee that these features or processes will actually motivate learners. This is partly because what motivates one learner will not necessarily motivate another. It is also because there is only limited evidence concerning the motivational impacts of different qualification features and processes.

Syllabus outcomes

There are various ways in which syllabus outcomes might be chosen with the explicit intention that studying for the new qualification ought to be an interesting or attractive prospect, thereby motivating uptake, effort, completion, and continuation.

One approach is to specify outcomes that have as much relevance as possible for the target cohort, to help them to appreciate the value of acquiring those outcomes, and to increase their motivation accordingly. Where traditional academic syllabuses have been criticised for seeming far removed from everyday life, making them more obviously relevant has been suggested as a solution. This might involve, for example, specifying syllabus outcomes for a school-based science qualification in terms of everyday scientific issues, phenomena, and controversies, with a specific focus on cultivating everyday scientific literacy (Wilson, Evans, & Old, 2015).

In essentially the same way, it has often been assumed that practical, or vocational, qualifications can be motivating for school and college students, and particularly for learners who have achieved only limited success on the traditional academic route. Over the decades, it has often been assumed that students who are ‘better with their hands’ are likely to find a more practically-oriented qualification more relevant to their personal proclivities and ambitions (Edwards, 1997). From a slightly different perspective, it has often been said that students who decide early that they are going to pursue a career in a particular sector (for example, construction or business) are likely to be more motivated if studying for a related vocational qualification than if continuing with a more general course.

Closely related to the idea of relevance is the idea that qualifications ought to be interesting for students to study. Over the years, numerous qualifications have been designed with the specific intention of motivating via interesting syllabus outcomes. The Certificate of Secondary Education (CSE) provides a radical example of this. According to the Secondary School Examinations Council (1963), effective teacher control of syllabus content and assessment methodology was the foundation of the CSE system, enabling a far more personalised approach to teaching and learning, for both teachers and learners:

> It will be the responsibility of the teachers themselves to ensure that what is examined is what they want to teach; they will not be obliged to teach what someone else has decided to examine.
>
> (SSEC, 1963, page 4)

The report explained that, in subject areas characterised by limited agreement among teachers over the ground to be covered:

> good teaching practice may involve selecting from amongst many possible topics a limited number which the individual teacher knows he can use to capture the interest and imagination of the particular group of pupils with whom he is concerned.
>
> (SSEC, 1963, page 5)

The SSEC acknowledged that prioritising the engagement perspective for these qualifications would mean deprioritising the information perspective, noting the challenge of ensuring comparability of standards under these circumstances:

> it would be better to accept some loss of comparability than to sacrifice the curricular needs of the pupils
>
> (SSEC, 1963, page 6)

Goodson (1995) described another curriculum initiative of the 1960s and 1970s that was developed with the specific intention of engaging learners. European Studies followed in the wake of the movement for comprehensive schooling, as a response to the challenge of teaching low-attaining students who seemed particularly unmotivated to learn a modern foreign language. Its core idea was to engage learners, by focusing as much on culture and people as on language. However, its reception was mixed, and its decision to invoke the interest of low-attaining students as a major criterion for content selection was criticised by some as creating a curriculum for inequality (in the sense of developing qualifications for lower-attaining students that would end up having lower market value). Goodson noted a reticence

from the university-based exam boards to establish bespoke O or A level syllabuses, which meant that European Studies did lack market value for many years.[29]

A different approach to building qualifications for engagement involves the specification of optional syllabus outcomes. Optionality enables students or teachers to tailor a qualification to current interests or future aspirations, under the assumption that the more interesting or relevant a qualification is, the more likely it will be to secure the effort that is necessary to achieve it. Optionality can be engineered in all sorts of ways, for example: choice from a range of optional components within a qualification, choice from a range of optional topics within a component, and so on. An important issue is who gets to choose. Choice can be exercised by teachers choosing an option for all students in a class, or it can be exercised by individual students choosing their own preferred options (Bramley & Crisp, 2019).

When qualifications are designed like this, all the syllabus outcomes are centrally specified (including the optional ones) but the optional elements are assumed to be exchangeable, such that the integrity of the overarching qualification is not undermined. This is common for qualifications that are heavily content-based, and where there is a lot of potentially in-scope material. It is common, for example, with history qualifications, where there might be choice over the periods that are studied and assessed, or with literature qualifications, where there might be choice over the texts that are studied and assessed. In this situation, the prescribed content for each optional element is specified in a manner that ensures comparable demand across the options, to provide a warrant for the assumption of exchangeability.

Essentially the same end can be achieved by making qualifications heavily outcome-based, where the syllabus outcomes are defined in terms of skills that can be applied across a wide range of contexts. In this situation, there may be no choice over the syllabus outcomes for the qualification in question. Yet students can still exercise choice in terms of the contexts in which they acquire and demonstrate those outcomes. England's Extended Project Qualification works in a similar way. All students must plan, research, and conduct a project, and they are all assessed against the same objectives and the same criteria. Within limits, though, the nature of the project is up to each individual student, which means that the assessed evidence might also differ – an essay, or an artefact, or a business plan, for instance.

Over the years, many qualifications have been designed to motivate students by providing an element of choice over syllabus outcomes. Again, choice can be

[29] European Studies was certificated as a CSE subject from early on, although it was years before it was certificated at O level, and even longer before it was certificated at A level. It is no longer a subject at either GCSE or A level. This reminds us just how fundamental the engagement perspective is to qualification design. Simply associating a curriculum area with a recognised qualification type (with an established market value) can be critical to its uptake and longevity.

motivating in terms of increasing the interest or relevance of a qualification. Perhaps just as important though, if not more so, is the potential positive impact on motivation associated with increasing the agency that students have over the course of their own education. The agency of choosing to study a certain kind of qualification can be critical in this respect, with students sometimes being motivated more by the desire to follow a particular route than by the intrinsic interest or market value of the qualification in question, for example:

> For other students the choice of [GNVQ] 'Business Studies' represented more a code of their intention to succeed in life, rather than a specific interest in understanding businesses. Their feelings went deeper than just the promise of a well-paid job at the end of the process (although that is clearly mentioned). The choice of 'Business Studies' was also an affirmation to their families and themselves that they are no longer 'dossers and tossers', but are now adults addressing the issue of constructing a good future for themselves.
>
> (Wahlberg & Gleeson, 2003, page 429)

So, while some of their tutors treated this Intermediate GNVQ as little more than a 'second-class' option, the students in this study saw it as mechanism through which they could re-engage with education, society, and their family (see also Stasz, Hayward, Oh, & Wright, 2004).

Assessment procedure

Quite independently of syllabus outcomes, features and processes built into assessment procedures have the potential to render a qualification more or less energising for learners (and sometimes for their teachers too). One of the features that is often discussed in this respect is the way qualification results are reported.

In particular, there is a longstanding debate over the pros and cons of norm-referencing versus criterion-referencing. Norm-referenced reporting focuses on the candidate's relative achievement within a cohort of learners (for example, they scored in the top 10% of candidates). Criterion-referenced reporting focuses on the candidate's absolute achievement independent of anyone else (for example, they are now fully occupationally competent).[30]

[30] Although it is often assumed that general qualifications are norm-referenced while vocational and occupational qualifications are criterion-referenced, the reality is far more subtle than this. Virtually all qualifications will incorporate a certain amount of both norm-referencing and criterion-referencing, although the balance will differ across qualifications. A levels and GCSEs, for instance, are neither strongly criterion-referenced nor strongly norm-referenced. They are better described as attainment-referenced (see Newton, 2022, to appreciate the nuance of this distinction).

It is often assumed that norm-referenced reporting is highly motivating for high-attainers yet highly demotivating for low-attainers, while criterion-referenced reporting can be motivating for all learners as it indicates that they have achieved something important in absolute terms. In fact, the situation for criterion-referenced reporting is more complicated than this. For example, some qualifications, especially vocational ones, have a single pass-fail threshold, confirming students have achieved all specified learning outcomes. This is a classic example of a criterion-referenced qualification. In practice, however, if the target cohort of learners includes some who are faster learners and some who are slower learners – but they all have the same amount of study time available to them – then a single pass-fail threshold can be problematic. If the standard for each learning outcome is set too high, then the slowest learners will struggle to achieve them all, which will obviously be demotivating for them. Yet if the standard is set too low, then the fastest learners will be insufficiently stretched, which can also be very demotivating.

This is why it is important to consider grading from the engagement perspective (to motivate students) as well as from the information perspective (to provide information about student attainment). Indeed, it is part of the reason why apprenticeships have been graded in recent years, to set engaging targets for faster learners (HM Government, 2013).

Incidentally, it is worth noting that qualifications tend not to classify students as having ‘failed’ nowadays, instead designating such candidates as ‘unclassified’, or ‘not-yet-competent’, or suchlike. This decision stems directly from the engagement perspective, as the threat of being labelled a failure is presumed to be demotivating for learners and unnecessarily damaging for candidates who do ultimately fail to achieve the standard (see Reay & Wiliam, 1999, for a related discussion).

Beyond reporting arrangements, there are various ways in which qualification design decisions might help to motivate, or demotivate, learners. For example, motivation can be reduced when assessment arrangements become unnecessarily burdensome – for students, for their teachers, or for both. Assessments that involve portfolio building provide an interesting case in point. Sometimes described as ‘work of the course’ coursework, an assessment of this sort might involve students continually updating an electronic or paper-based portfolio with evidence of their latest and best achievements. On the one hand, portfolio assessments like this can be designed to integrate seamlessly with everyday teaching and learning, helping to minimise the assessment burden. In this instance, the presumed motivational impact derives largely from the removal of the threat of terminal assessment – with its anxiety-inducing preparatory period and its single-shot jeopardy. On the other hand, portfolio assessments sometimes end up not integrating seamlessly with teaching and learning, requiring numerous additional bespoke assessments that consume time that might otherwise be used for teaching and learning. Assessments of this

sort can still be valuable in terms of reinforcing learning, but if they end up feeling unnecessarily duplicative, then this has the potential to be demotivating for students.

Portfolio assessments can also be very burdensome for teachers. This is not just because they have primary responsibility for grading portfolios. It is also because they are required to maintain an evidence trail of their judgements and rationales, for internal and external quality assurance purposes. Again, this can take time away from teaching and learning. But, if it becomes overwhelming, then it can also become extremely demotivating for teachers. Finally, portfolio assessments can be demotivating for learners whose experience of anxiety and jeopardy is actually worse under the pressure of having to demonstrate competence continually throughout the course than under the pressure of a single-shot terminal assessment.

Another way in which qualification design decisions might help to motivate, or demotivate, learners relates to the perceived accessibility of assessment arrangements. Where these arrangements are perceived to play to students' strengths, or to compensate for their weaknesses, this is likely to be more motivating for them.

Methods that incorporate extended written responses, such as essays or project assignments, illustrate this well. Some students find extended writing challenging, and this is pertinent to the extent that being able to provide extended written evidence (of the syllabus outcomes in question) is often technically irrelevant to the construct we are trying to assess. In other words, students who may have achieved those syllabus outcomes – but who struggle to provide satisfactory evidence of this in an extended written response – will be penalised unfairly.[31] Clearly, when this occurs, it will be extremely demotivating. This is not just a matter of students being demotivated by the prospect of having to demonstrate competence in this manner at the end of their course, as they will inevitably have to face repeated assessments of this sort throughout their course, in order to prepare them.

If a certain approach to assessment proves to be particularly challenging for the cohort of learners that a new qualification is intended to target – for reasons that are technically irrelevant to its intended syllabus outcomes – then incorporating that approach would inevitably be demotivating for many of those learners. If it were possible to find a more accessible approach, then this would be likely to increase their motivation for learning, and might also help to improve validity too. Even if it ended up not improving validity – perhaps even leading to a slight reduction in validity, having opened other potential threats – this might still be the best decision to

[31] In assessment terminology, we describe this as leading to 'construct-irrelevant variance' in qualification results, that is, candidates' results varying for reasons that have nothing to do with the construct (syllabus outcomes) that we are trying to assess. Construct-irrelevant variance and construct underrepresentation comprise the 2 principal threats to validity (Newton, 2020c).

make on balance, particularly if the engagement perspective had been prioritised over the information perspective.

Sometimes, particularly with low-attaining learners on vocational courses, centre-based assessors are permitted to determine the most suitable approach for assessing their students. They might also be permitted to adopt an alternative approach when the standard approach fails to reveal evidence of competence. For example, an assessor might supplement a work-based observational assessment with follow-up questioning, where the learner's performance seemed substandard for one reason or another. If the questioning revealed that the learner did actually understand what they were trying to do despite having failed to provide sufficient evidence – particularly if there appeared to have been construct-irrelevant reasons for that failure – then the assessor might judge the student to be competent after all.

During the 1980s and 1990s, it became increasingly common to prioritise the engagement perspective for learners in what was then referred to as the Learning and Skills Sector (LSS), which spanned school sixth forms, further education colleges, workplaces, and adult learning environments. This was largely in response to the raising of the school leaving age in 1972 and rising unemployment throughout the 1970s, which meant that post-16 education and training had to accommodate an expanded cohort of learners, many of whom had experienced only limited success on the traditional academic route. Reporting on an investigation into the impact of different modes of assessment on achievement and progress in the LSS, Torrance, Colley, Garrett, et al (2005) concluded as follows:

> Overall the study found that assessment methods per se do not directly affect learners' choice of award or likelihood of success, but the association of certain awards with methods which employ extensive writing (coursework, exam essays, etc) does. Thus for example, practical tests and/or multiple-choice tests are seen as acceptable – and indeed unavoidable – across most groups of learners in the sector, especially younger trainees, but extensive written work is disliked and largely avoided, except by A-level takers."
>
> (Torrance, et al, 2005, page 1)

Qualification structure

Qualification designers sometimes make high-level design decisions – concerning the overall structure of the qualification – that are explicitly intended to keep learners motivated. These decisions typically have implications for both syllabus and assessment, and often for pedagogy too. Unitisation (modularisation) provides a good example of this, particularly where the units (modules) are designed to be accumulated gradually over time.

The introduction of credit-based awards within the adult learning sector illustrates this approach. Wilson (2010) traced the origins of the Qualifications and Credit Framework (circa mid-2000s to mid-2010s) to the influence of the Open Colleges, Open College Federations, and Open College Networks that began to develop in various parts of the UK during the 1970s. The credit-based systems that these organisations developed were designed to recognise the disparate learning trajectories of adult learners – disparate because of their very different prior attainments and future aspirations – and their fundamental need to reengage with the process of learning itself, potentially quite slowly and discontinuously. The idea of being awarded credit for achieving a certain volume of learning outcomes at a certain level of performance became fundamental to this approach. The most critical feature of this system involved the award of credit itself – which recognised and rewarded each small step in a bespoke learning journey – to motivate learning and to build confidence in the learner.

For similar reasons, unitisation spread throughout the landscape of vocational and technical qualifications from the 1970s onwards, particularly with the introduction of Technician Education Council awards and Business Education Council awards. Prior to their introduction, drop out and failure had plagued the landscape, particularly for the system of classically-designed Ordinary and Higher National Certificates and Diplomas. Unitisation was intended to help address this, by replacing an overarching course with a programme of study built from self-contained units, each of which could be passed in their own right on a unit-by-unit basis (Newton, Clarke, et al, 2024). The potential of unitisation to help improve completion rates – by flexing to meet the needs of individual learners – was viewed very positively (Like, 1986).

Although GCSEs and A levels are currently designed as linear (rather than modular) qualifications, they were mostly modular until around a decade or so ago. The exam boards introduced the idea of modular A level syllabuses during the 1990s, and the Curriculum 2000 reforms meant that all A levels were modular from 2002 onwards. GCSEs also became modular during the 2000s, albeit less systematically. The idea of completing a GCSE module-by-module was not unpopular, especially with its reassurance of being able to resit a poorly completed assessment. Unfortunately, the extent of resitting assessments led to unanticipated levels of disruption on teaching and learning, which was an important part of the rationale for returning to a linear model. An evaluation of this transition failed to find evidence to suggest that either modular or linear GCSEs led to better educational outcomes (Baird, Caro, Elliott, El Masri, et al, 2019).

Finally, students sometimes need the opportunity to experiment with new areas of learning, with the potential to change their progression route if they decide that they have made suboptimal choices. Qualifications can be designed to accommodate this, or to frustrate it. For example, a 2-year vocational qualification might encourage students to choose from a wide range of units during the first year, only requiring

them to specialise in the second year. To encourage experimentation during the first year, it might be decided that grades should only be derived from year 2 units. This can cut both ways, though. If grades from year 1 units do not count, then there may be less incentive to work hard during the first year, especially for year 1 units that a student may not build upon during year 2. Considerations of this sort might even recommend a weighted approach to grading, with (say) 33% derived from the first year and 67% from the second.

Pedagogical purposes

A qualification can be designed to promote, or require, pedagogical techniques, strategies, or approaches that are intended to improve the quality of teaching and learning that occurs while studying for it. Decisions like these are likely to relate to the overall structure of the qualification, with implications for both syllabus outcomes and the assessment procedure.

A good example of this is the idea of mastery certification based upon a conjunctive (hurdle-based) approach to aggregating assessment judgements, which means that learners have to achieve all the specified learning outcomes (jump all the hurdles) in order to pass the qualification. Mastery certification is often associated with occupational qualifications, where the intention is to certify full occupational competence. Yet syllabus outcomes of any sort can be certified in terms of mastery, and when non-occupational outcomes are certified like this, the main reason is to promote mastery teaching and learning (see Bloom, 1968).

Mastery certification provides an extremely strong incentive for mastery teaching and learning, as students who do not master the domain of learning do not pass the qualification, period. Mastery-based qualifications are often also outcome-based, where the specified learning outcomes and associated assessment criteria clarify details concerning the breadth and depth of learning required.[32] Ofqual's recent research into CASLO qualifications, which are both outcome-based and mastery-based, explains this approach to qualification design in considerable detail, and explains how CASLO qualifications are best understood as integrated teaching-learning-assessment systems. Qualifications of this sort tend to be designed to

[32] It is possible to design a qualification to certify mastery without adopting an outcome-based approach, for example, regulated accountancy qualifications in England often certify mastery using very high pass-mark thresholds (applied to classically designed exams). Equally, any hurdle-based qualification, which requires students to pass all units or components for the overall qualification pass, also applies a mastery model of sorts.

incorporate and integrate both formative and summative strategies (albeit keeping them sufficiently distinct to secure validity).[33]

There are many other ways to design a qualification to promote teaching and learning. For instance, by specifying a syllabus in terms of generic skills-based learning outcomes, and by basing the assessment procedure around extended assignments, a project-based or problem-based approach to teaching and learning can be promoted. This has been a feature of certain BTEC qualifications over the years. From a slightly different perspective, the 'Cambridge Approach to Textbooks' notes how Singapore Maths was based upon a model of learning linked to a spiral curriculum approach, with direct implications for both the content of syllabus outcomes and their sequencing (Cambridge Assessment, 2018, pages 13 to 14).

Sometimes qualification features and processes are specified in a certain way to improve teacher understanding of the domain of learning and qualification standards. The decision to use teachers as developers, administrators, and markers in the New Zealand National Education Monitoring Programme provides a good example of this, albeit not specifically in a qualification context (Newton, 2008). Consensus moderation of teacher assessment in Queensland provides another example, this time in a qualification context (Maxwell, 2010).

Sometimes qualifications are designed in ways that make them vulnerable to gaming (Meadows & Black, 2018), for example, when the predictability of an English language qualification effectively allows students to memorise compositions off by heart and reproduce them rote in the exam. Building an optimal amount of unpredictability into an assessment procedure – not too much but not too little – ought therefore to have a positive impact on teaching and learning, by preventing undesirable teaching and learning strategies (Holmes, Khan, Zanini, & Black, 2020).

Wiliam's analysis of principled curriculum design includes various insights into how a qualification might be designed to promote effective teaching and learning (Wiliam, 2013). Particularly important, but often overlooked, is the necessity of designing a qualification to promote progression in learning. This consideration applies when developing an individual qualification, where units are designed to be taken in a particular order, such that outcomes acquired earlier help to scaffold subsequent teaching and learning, which also reinforces and consolidates the earlier learning. Yet the same consideration applies when designing a sequence of qualifications at successively higher levels, to ensure that they are sufficiently vertically integrated.

---

[33] Although the idea of formative assessment seems naturally to fall within this perspective, the related idea of Learning How To Learn (LHTL) is slightly more ambiguous. Where the focus is on how LHTL approaches or techniques can improve learning, in situ, this would seem to fall within the engagement perspective. Conversely, where the focus is on the acquisition of LHTL competencies, as explicit learning outcomes in their own right, this might be said to fall within the expertise perspective.

Wiliam also sounded a word of caution over the need to ensure that syllabus outcomes are selected to focus on the most important elements of a domain:

> Generally, the only way to improve a curriculum is to leave out important material so that the teacher and the students can spend more time on the *more* important material.

(Wiliam, 2013, page 36)

This acknowledges a problematic tendency during qualification reform, which is to add in more new material than the omitted old material left space for.

Contrasting opinions

It is hard to argue with the 2 principal engagement purposes. It is important that students are as motivated as possible, and it is important that their teachers are teaching as effectively as possible. There is, however, plenty of scope for disagreement concerning when, or perhaps whether, it is important to build features and processes into qualifications with the specific intention of motivating learners or directing pedagogical practices.

For instance, when learners are already highly motivated to achieve high grades in their qualifications, the need to actively engage them (through qualification design decisions) would seem to be low. This might, for instance, be true for candidates studying a qualification that provides a licence to practice within a well-paid profession. Conversely, for adults who exited formal education a long time ago, with no formal qualifications and little confidence in their own ability to learn, it might be very important to design qualifications with engagement in mind. The same might be said for many of the lowest-attaining students who remain in full-time education and training.[34] No doubt there will be learners in between these extremes for whom the situation is less clear, such as students who achieve middling grades from school-leaving exams, who continue in full-time education albeit with no particular educational or occupational ambitions.

[34] The collapse of the youth labour market from the 1970s onwards led to a major expansion of post-16 participation. This meant that FE colleges increasingly had to take on a “social inclusion role” (Hodgson & Spours, 2019, page 228), to cater for young people who were able to engage neither with work nor with more formal educational studies. Eraut took this analysis a step further, suggesting that dealing with learners under these circumstances could be understood as a qualification purpose in its own right: “keeping job-seekers gainfully occupied” whilst “delaying entry into the workforce in times of high unemployment” (Eraut, 2001, page 88). Note that this would be an engagement purpose, of sorts, albeit a very different kind of engagement purpose from those described earlier, and effectively a method of social control (Yeomans, 1998).

There have been times when policymakers in England have been positively disposed to designing qualifications with engagement in mind, and there have been times when they have been less inclined to do so. The comparison of policy intentions underlying the GNVQ and its successor, the AVCE (Advanced Vocational Certificate of Education), provides a good case in point (Wellings, Spours, & Ireson, 2010).

A good indication of the need to consider engaging students via qualification design decisions is a high incidence of qualification drop out and failure. This was part of the rationale for the reform of industry and commerce qualifications during the 1970s, which included engagement-related design decisions, such as unitisation and the introduction of outcome-based approaches (Newton, Clarke, et al, 2024). The need to consider engagement-related design decisions might also be influenced by external factors, such as the lack of availability of attractive job opportunities for young adults, reducing the perceived value of becoming qualified (Hargreaves, 1989).

Unfortunately, there is scope for disagreement over the plausibility of claims that particular qualification features and processes will, in fact, improve engagement, without also resulting in unacceptable negative consequences. As noted earlier, there may not always be a great deal of evidence to arbitrate the matter, and some claims may be hotly debated. For example, Records of Achievement were promoted by many education scholars during the 1980s, on the assumption that they would help to engage learners (Hargreaves, 1989). Despite a flurry of interest and political support, which resulted in a variety of schemes, they did not prove to be particularly successful and evidence of their positive impact on motivation was underwhelming.

Finally, it is important to recognise the potential for tension between motivational engagement and pedagogical engagement, in particular, when qualifications are effective in improving the quantity of learning that occurs – by getting learners to the end of their courses – but only at the expense of the quality of that learning. This idea has been explored extensively by Ecclestone in the context of vocational education in England during the 1990s and 2000s (for example, Ecclestone, 2002; 2007; 2010; 2012). She warned of the risk of teachers surrendering their role as vocational educators, lowering their expectations of learners, and – under the guise of mentoring, support, and formative feedback – coaching them to achieve no more than minimal success via superficial teaching and learning strategies.

Ecclestone was part of the team that investigated the impact of modes of assessment in the Learning and Skills Sector. They noted a general trend (that is, for both general and vocational qualifications) towards making assessment criteria far more transparent than in previous decades. In theory, transparency ought to be good for teaching and learning, as it enables teachers and students to improve their understanding of what needs to be achieved, to help them to take a more active role

in closing the gaps in their learning. While acknowledging the benefits of increased transparency, the team concluded that:

> Transparency, however, encourages instrumentalism. The clearer the task of how to achieve a grade or award becomes, and the more detailed the assistance given by tutors, supervisors and assessors, the more likely are candidates to succeed; but succeed at what? Transparency of objectives, coupled with extensive use of coaching and practice to help learners meet them, is in danger of removing the challenge of learning and reducing the quality and validity of outcomes achieved. We have identified a move from what we characterise as assessment *of* learning, through the currently popular idea of assessment *for* learning, to *assessment as learning*, where assessment procedures and practices may come completely to dominate the learning experience, and 'criteria compliance' comes to replace 'learning'. This is the most significant challenge confronting assessment in the LSS: balancing the explicitness of learning objectives and instructional processes against the validity and worthwhileness of learning outcomes.
>
> (Torrance, et al, 2005, page 2)

Ecclestone even warned of the risk of coming to treat education as a kind of therapy, particularly for the most marginalised and disaffected. That is, under the guise of motivating and engaging, educators risk constructing a diminished view of learners, and communicating unduly low expectations of their capacity to achieve (Ecclestone, 2004). Yet the idea that education, and particularly vocational education, has been reduced to therapy in recent years is hotly disputed. Other academics have reasserted the contrary position, that therapeutic strategies are vital when educating disaffected learners with little experience of success and achievement at school (Hyland, 2005). Relatedly, Lucas, Spencer, & Claxton (2012) discussed the serious and widespread challenge of low self-esteem, arguing that its implications for pedagogy (let alone qualification design) have not yet fully been explored.

Information-related purposes

To achieve certain educational, occupational, or societal ends, we use qualification results to help us to make better decisions (than we would be able to make without the information that these results provide). Typically, these decisions concern individual learners. For example, we might use qualification results to decide whether they have mastered enough of an entry-level course to permit them to enrol on a higher-level one. Sometimes, we aggregate qualification results, to support decisions concerning the administration of classes, schools, or regions. For example, we might use class-level results to help us to determine how effectively an individual teacher has taught a class over a period of time.

Decisions that stem from the information perspective typically concern the design of the assessment procedure. They aim to ensure that the qualification will measure the right thing, in the right way, to provide accurate and useful results. This is essentially how we define validity, the technical quality of an assessment procedure (Newton, 2017b).

The sort of person who might champion the information perspective is likely to be an assessment or measurement specialist, as these decisions tend to be highly technical. However, in championing this perspective, these specialists are advocating on behalf of qualification users – those who will use the information that qualification results provide to make real-world decisions – as well as on behalf of the students whose lives will be affected by those decisions.

Assessment purposes

Information-related purposes tend to be known as **assessment purposes**, which concern the **uses** to which assessment results are put. Although it is widely accepted that there are many different kinds of assessment purpose, there are no generally accepted categories for classifying them. For example, when developing proposals for assessing England's new national curriculum, the Task Group on Assessment and Testing (1988) identified 4 high-level categories of purpose:

- formative (assessment for learning)
- summative (assessment of learning)
- evaluative (assessment for accountability)
- diagnostic (assessment for special intervention)

With high-level categories like these, the purposes that cluster within each one are likely to be quite divergent, which can be problematic if the scheme invites global fitness-for-purpose judgements along the lines of: this assessment is fit for summative purposes. After all, there is no reason to assume that an assessment that is fit for one summative purpose will necessarily also be fit for another, let alone for all purposes in the summative category. This is why a similar analysis provided by Newton (2007a; 2010) opted not to use higher-level categories at all, opting instead for the following 22 lower-level ones:

1. to decide whether individual students are making sufficient progress in attainment in relation to expectations or targets; and, potentially, to allocate rewards or sanctions (student monitoring)
2. to identify students' proximal learning needs, to guide subsequent teaching interventions (formative)
3. to judge the social or personal value of students' achievements (social evaluation)

4. to clarify the type and extent of students' learning difficulties, or needs, in light of well-established criteria (diagnosis)
5. to determine whether students meet eligibility criteria for special educational provision (provision eligibility)
6. to identify students who differ significantly from their peers, for further assessment (screening)
7. to segregate students into homogeneous groups – on the basis of aptitudes or attainments – to make the instructional process more straightforward (segregation)
8. to identify the most suitable courses or vocations for students to pursue, given their aptitudes (guidance)
9. to identify the general educational needs of students who transfer to new schools (transfer)
10. to locate students with respect to their position in a specified learning sequence, to identify the level of course which most closely reflects it (placement)
11. to decide whether students are sufficiently qualified for a job, course, or role in life – that is, whether they are equipped to succeed in it – and whether to enrol them or to appoint them to it (qualification)
12. to predict which students – all of whom might in principle be sufficiently qualified – will be the most successful in a job, course, or role in life, and to select between them (selection)
13. to provide legal evidence – the licence – of minimum competence to practice a specialist activity, to warrant stakeholder trust in the practitioner (licensing)
14. to provide evidence – the certificate – of higher competence to practice a specialist activity, or subset thereof, to warrant stakeholder trust in the practitioner (certification)
15. to identify the most desirable school for a child to attend (school choice)
16. to decide whether institutional performance – relating to individual teachers, classes or schools – is rising or falling in relation to expectations or targets; and, potentially, to allocate rewards or sanctions (institution monitoring)
17. to identify institutional needs and, consequently, to allocate resources (resource allocation)
18. to identify institutional failure and, consequently, to justify intervention (organisational intervention)
19. to evaluate the success of educational programmes or initiatives, nationally or locally (programme evaluation)

20. to decide whether system performance – relating to individual regions or the nation – is rising or falling in relation to expectations or targets; and, potentially, to allocate rewards or sanctions (system monitoring)
21. to guide decisions on comparability of examination standards for later assessments on the basis of cohort performance in earlier ones (comparability)
22. to 'quality adjust' education output indicators (national accounting)

Note that 'formative' and 'diagnostic' uses appear as distinct categories within both schemes, with essentially the same meaning. The 'summative' and 'evaluative' categories from TGAT (1988) are slightly more ambiguous, although they are presumably intended to distinguish between decisions that are based upon individual versus aggregated results. In the scheme from Newton (2007a; 2010), purposes 1 to 14 would tend to be based on individual (summative) results – with the exception of purpose 2 which would generally require a more fine-grained analysis – while purposes 15 to 22 would all be based on aggregated (summative) results.

Figure 7 provides further insight into 6 of these purpose categories, incorporating information from Table 1 of Newton (2007b). The point of this table is to explain why different assessment purposes might have different implications for designing an assessment procedure – and sometimes radically different implications. For instance, information that is used to make formative decisions, related to instructional interventions, will need to be generated frequently and with a fine grain size. This is not the same kind of information that would need to be generated to make student monitoring decisions – confirming whether or not student progress is broadly on track – which would need to be generated far less frequently and with a much higher grain size.

Likewise, the kind of information that would be needed to make selection decisions (based on individual results) is different from the kind of information that would be needed to make decisions related to teacher monitoring (based on aggregated results). For example, selection decisions are typically based upon information relating to the final level of attainment achieved by a student, for instance, their final A level grades. Conversely, how effectively an individual A level teacher has taught their class is likely to require a value-added analysis, which compares the (average) extent of progress that students make within a subject from GCSE to A level.

Even within a single assessment purpose category, different purposes will often necessitate different design decisions. Newton (2008), for instance, compared system monitoring arrangements across a number of countries, noting how their design decisions differed substantially depending on their specific purposes. In New Zealand, the National Education Monitoring Project (NEMP) was intended to be of particular use to teachers and curriculum designers. Consequently, it was designed to report at the level of individual tasks and small clusters of tasks. In the USA, although originally reporting at the level of individual tasks, the long-term trend

National Assessment of Educational Progress (NAEP) was specifically redesigned to be of use to policymakers and stakeholders, subsequently reporting only at the overall subject level.

Even the distinction between high-stakes and low-stakes decision-making may necessitate different design choices, as threats to validity and the quality of information provided by results can increase with rising stakes. In particular, assessments that have high stakes for individuals – such as qualification, selection or licensing – will need to be designed with high levels of security in mind at all stages of the assessment lifecycle: from identity checking at the assessment centre, to the disposal of assessment scripts and portfolios (Millman & Greene, 1989; Schmeiser & Welch, 2006). Such considerations do not arise in the same way when results are used only for low-stakes purposes, where security is unlikely to be a major concern.

Although the 22 categories of assessment purpose are useful for illustrating the wide range of uses to which results might be put, this list is not exhaustive. For instance, it might be decided to redesign a qualification system as part of an international effort to improve the mobility of students across countries, by making qualification results far more transparent for users in terms of the learning that they signify. So, this would represent a new category, a 'mobility' purpose category. Another important omission from the list (in the UK qualifications context) is a category of assessment purpose that falls somewhere between institution monitoring and resource allocation: the use of results to determine whether an institution is eligible for full remuneration of education and training costs, when arrangements dictate that a proportion of this funding will be withheld for students who fail to complete their qualification(s) successfully. We might refer to this as the 'funding eligibility' purpose category, related to the idea of payment on results. Whenever results are used for purposes that lead to payments, rewards, sanctions, interventions, or suchlike – either for institutions or for individual teachers and trainers – qualifications will need to be designed with subversion resistance in mind, especially the threat of qualification standards being manipulated downwards (see Wolf, 2011).

Figure 7. Decisions and design implications

Purpose category	Typical users	Decision to be taken	Example of a design implication
Formative	Students, teachers, parents	Individual (or aggregated) results are used to identify student (or group) learning needs, to direct subsequent teaching and learning.	Frequent, fine-grained analyses of attainment in small subdomains will typically be required to guide interventions for individual students, and reporting will often occur through verbal feedback.
Student monitoring	Students, teachers, parents	A comparison of earlier and later results enables the user to decide whether individual students are making sufficient progress in attainment over time (this is sometimes linked to student performance targets).	Results need to be particularly reliable (across the full range of attainment for the target cohort) given the potential for assessment inaccuracy in both the 'input' and 'outcome' measures.
Placement	Teachers	Individual results are used to place students in teaching groups, or educational programmes, that will be most suitable for them (this might include a decision over which exam tier to enter a student for).	Reasonable reliability is required (since the decisions may have high stakes for students) although – where placement decisions are revisited fairly frequently – reliability may be lower since misplacement can be rectified.

Purpose category	Typical users	Decision to be taken	Example of a design implication
Selection	Admissions tutors or officers, employers	Individual results are used to predict which applicants – all of whom might, in principle, be sufficiently qualified – will be most successful in a job or course of instruction (thereby providing a basis for choosing between them).	Results need to discriminate reliably (across the full range of attainment for the target cohort) especially when different progression pathways mean that applicants competing for certain opportunities (for example, medical degrees) will tend to have similar prior attainment levels.
Institution monitoring	Teachers, school managers, local authorities	Aggregated results are used to decide whether institutional standards are rising or falling over time, for example, the effectiveness of individual teachers (this is sometimes linked to teacher performance targets linked to penalties or rewards).	Where results are used to allocate rewards or penalties to teachers in different departments of a school, there should be comparability of assessment standards across subjects (that is, common evaluation criteria).
Programme evaluation	Evaluators, policymakers	Aggregated results are used to evaluate the success of educational programmes or initiatives, nationally or locally.	Ideally, results should reflect the construct at the heart of the evaluation (which would be challenging if that construct embodied educational aims that were broader than those measured by qualification syllabuses).

Multiple purposes

During the mid-2000s, the issue of how many assessment purposes ought to attach to national curriculum test results became a matter for national debate, including during sessions of the House of Commons Children, Schools and Families Committee (HCCSFC, 2008a). Ken Boston, then head of the Qualifications and Curriculum Authority, referred to the list of 22 purpose categories from Newton (2007a) and suggested that results from key stage tests were being used for 14 of them. Ultimately, the select committee concluded:

> The evidence we have received strongly favours the view that national tests do not serve all of the purposes for which they are, in fact used. The fact that the results of these tests are used for so many purposes, with high-stakes attached to the outcomes, creates tensions in the system leading to undesirable consequences, including distortion of the education experience of many children. […] In short, we consider that the current national testing system is being applied to serve too many purposes.
>
> (HCCSFC, 2008a, p.20)

In a supplement to the select committee report, Ofsted clarified its own, somewhat conflicting, opinion:

> We believe that using national test and exam data for several purposes is a strength. One of the core principles for efficient use of data in government should be 'collect once, use more than once', and this is the case. We believe that the use of test and exam data in inspection is appropriate. Set within the context of the inspection methodology described above, we believe the data are extremely helpful in evaluating schools' effectiveness.
>
> (HCCSFC, 2008b, p.15)

The issue of multiple purposes subsequently became a focus for the Independent Review of Key Stage 2 Testing, Assessment and Accountability (Bew, 2011). Bew recommended specifying the following 3 principal uses:

- holding schools accountable for the attainment and progress made by their pupils and groups of pupils
- informing parents and secondary schools about the performance of individual pupils.
- enabling benchmarking between schools, as well as monitoring performance locally and nationally

It is interesting to note the order of this list, with accountability as the first principal use. This would be extremely unlikely to happen in a qualification context, given the role that qualifications are designed to play in formally certifying learning. Bew further prioritised assessment purposes by distinguishing principal uses from a (somewhat problematically) large list of secondary ones.

The tension between Ofsted and the select committee is important to address. National data collection is extremely costly and burdensome and so it is entirely understandable to want to use the data for as many different purposes as possible. Yet different decisions require different kinds of information, certain sorts of decision require better quality information than others, and certain kinds of inaccuracy are more problematic for some decisions than for others. In other words, results from a particular test or qualification might be entirely fit for one use of results, somewhat less fit for another, while being entirely unfit for others. Fitness for purpose cannot be assumed and needs to be established on a use-by-use basis.

Furthermore, beyond decision-related concerns lie impact-related ones. In 2008, the select committee was particularly concerned about the use of test results for accountability purposes, which raised the stakes of assessment for schools and teachers, which in turn exacerbated the likelihood of a negative backwash impact on the curriculum. For instance, while the national curriculum for English was equally weighted in terms of reading, writing, and speaking and listening, the lack of a speaking and listening test risked teachers focusing unduly on reading and writing at the expense of speaking and listening. Of course, if this occurred, then it would not simply be bad from the expertise perspective (students not being taught to speak and listen), it would also be bad from the information perspective (as inflated results for reading and writing would misrepresent the situation for English overall).[35]

Because qualifications are (by definition) designed to certify intended learning to specified standards, the construct that they are designed to assess will always relate primarily to formal syllabus outcomes of one sort or another. Consequently, the primary use of results from a qualification is likely to relate to an individual-level assessment purpose (for example, placement or selection) while aggregate-level assessment purposes (for example, institution monitoring or organisational intervention) will tend to be considered secondary uses at best.

---

[35] Owing to the omission of speaking and listening (from English) and the omission of the scientific enquiry attainment target (from science), the National Foundation for Educational Research proposed that the key stage 2 tests did not serve the national monitoring function very well. If that had, for instance, been the sole purpose of key stage 2 assessment, then a low-stakes, light-sampling survey would have been a far better approach (see HCCSFC, 2008a, page 21). Note that this is exactly the approach that has been adopted in recent years for the National Reference Tests, which are designed solely to serve a national monitoring function for English and maths at the end of key stage 4. These tests do not even generate results at the student level.

Having said that, it not uncommon for policymakers to want qualifications to serve both primary and secondary uses. Indeed, they will often be used for multiple primary and multiple secondary assessment purposes. For this reason, it is important to prioritise between these purposes, and to orient design decisions toward the highest priority ones. In practice, this is likely to involve planning qualification design based on the highest priority use, then tweaking design decisions accordingly to optimise fitness for any remaining prioritised purposes.

An even more complicated situation arises when a qualification is designed to certify the achievement of intended learning to specified standards, yet, in practice, its results tend not to be interpreted in terms of those syllabus outcomes, but in terms of a more abstract quality, like aptitude for learning a subject, or potential productivity in the workforce. In this situation, the qualification might have high validity in relation to the intended attainment-related interpretation (the design interpretation), but have lower validity in relation to the actual aptitude-related interpretation (the use interpretation). This situation is probably not that uncommon in relation to certain uses of results (Newton, 2007b; Wheelahan, Buchanan, & Yu, 2016). Yet it is not well theorised in terms of its implications for qualification design.

Contrasting opinions

The design of an assessment procedure is a highly technical enterprise. It is supported by an extensive literature, which has been building for around 150 years, and it is championed by a variety of national and international organisations, such as the National Council on Measurement in Education (USA), the Association for Educational Assessment – Europe (AEA–E), the Association of Test Publishers (ATP), and the International Association for Educational Assessment (IAEA).

Although there is still plenty of debate concerning the principles of educational measurement – including fundamental disagreement over how broadly to define its core concept, validity (Newton & Baird, 2016) – there is also a reasonable amount of consensus concerning good and bad practices in designing assessment procedures. As such, we might expect assessment specialists to be able to reach broad agreement concerning design requirements arising from alternative uses of results. For instance, if it were decided that the primary use of results from a particular qualification was licensing – that is, to (dis)allow a person to practice a safety-critical profession, such as electrical installation – then we would expect assessment specialists to agree that this would require a high degree of validity from the associated assessment procedure to ensure consistently accurate results from one assessment to the next.

Having said that, even a group of assessment specialists might not entirely agree over the optimum constellation of design decisions for achieving a high degree of validity. This is because validity is an umbrella concept for characterising overall

measurement quality, which can be subdivided into a variety of dimensions, some of which are sometimes in tension with each other. For example, it is possible to increase score reliability by decreasing construct representation, which often amounts to excluding from the assessment procedure learning outcomes that are particularly hard to assess accurately. This suggests that the same overall level of validity might be achieved through different relative weightings of score (un)reliability versus construct (under)representation.[36]

In terms of contrasting opinions, though, the biggest challenge is to anticipate the full range of assessment purposes for which a qualification will need to be designed, and then to achieve some kind of consensus over their prioritisation. It is important to have assessment specialists guiding this process, as they are best placed to understand the full variety of uses to which assessment results might be put. However, the identification and prioritisation of assessment purposes is essentially a policy decision, which is likely to require input from a wide range of stakeholders. This will include stakeholders from different potential user groups – for example, higher education selectors and inspectors of education – whose views will inevitably diverge from each other in terms of the importance that they attach to alternative uses.

Conflict between perspectives

It is not inevitable that different purpose perspectives – or different purpose categories within a purpose perspective – will recommend different qualification design decisions. Far from it. Sometimes there will be considerable alignment.

Consider, for instance, a qualification that is intended solely to provide a licence to practice a particular occupational role, such as a heating engineer. From the expertise perspective, the body of learning that would need to be acquired while studying for the qualification would comprise the body of learning that enables a budding heating engineer to practice safely and competently. Likewise, from the information perspective, the qualification result would need to certify exactly the same body of learning. In this situation, the 2 perspectives are essentially 2 sides of the same coin, and they are likely to result in essentially the same design recommendations: the syllabus will need to specify the acquisition of full occupational competence, and the assessment procedure will need to certify the acquisition of full occupational competence.

Yet purpose perspectives do frequently come into conflict with each other. For instance, we might assume that students will learn best when given the opportunity

[36] Note that there are no procedures for computing an overall level of validity, so this is a fuzzy concept, which affords no more than rough, holistic judgements.

to study syllabus outcomes that are selected to be as intrinsically interesting as possible (engagement perspective). Yet the demands of life and work sometimes translate into highly valued syllabus outcomes (expertise perspective) that learners can find quite boring. So, should we decide to prioritise how well they learn, or what they learn? This will not necessarily be an easy decision to arbitrate.

Conflict can also arise between the engagement perspective and the information perspective. Optionality provides a good case in point because it enables the personalisation of syllabus outcomes to individual interests or ambitions, which is assumed to be good for motivating students. Unfortunately, the more optionality is built into a qualification – via optional units, optional sections, or optional questions – the less information the qualification certificate can convey concerning exactly what students have learnt. Optionality also introduces a variety of threats to validity, such as the risk that optional elements are differently difficult, which would reduce the overall accuracy of qualification results. Sometimes, however, we will conclude that it is important that students remain engaged, and incorporate an element of optionality accordingly, deciding to tolerate the likelihood of a reduction in overall accuracy.

The expertise and engagement perspectives often clash when considering qualifications for low-attaining post-16 students in full-time education, especially the most fragile who are at risk of dropping out. Instinctively, we would like these students to be studying qualifications that will have serious purchasing power in the job market because they signal to employers that students have acquired a solid foundation of knowledge, skill, and understanding of direct relevance to a particular sector (akin to the vocational readiness syllabus purpose described earlier). This would seem to prioritise both the expertise perspective (the importance, for students, of achieving this readiness) and the information perspective (the importance, for employers, of having this readiness certified). Yet some of these low-attaining students will not be ready to study for a vocational readiness qualification, let alone for an occupational proficiency one. They may still lack some of the basic knowledge, skill, and understanding that provides a necessary foundation for life and work. And they may also have limited, if any, confidence in their ability to learn, given their experiences of education to date. For students like this, the priority might simply be to keep them engaged with learning for long enough to work effectively on improving their basic skills and confidence in learning (akin to the personal achievement syllabus purpose described earlier). In this situation, to prioritise the engagement perspective over the other 2 perspectives would be to argue that these students need to be supported to obtain at least some value from their post-16 learning, even if this learning might have less market value than that achieved by

other post-16 learners, and even if their certification might be somewhat less accurate too.[37]

For reasons like these, it is important to attempt to prioritise between the 3 purpose perspectives. This may not always be possible. For instance, even following sustained debate, it may not be possible to decide between the information and expertise perspectives, and they may end up having to be equally weighted. Alternatively, there might be multiple intended purposes within 2 or 3 of the purpose perspectives, making it just too complicated to prioritise them coherently. Yet it would still have been important to have attempted to prioritise those perspectives, to make sure that there had been a full and frank discussion of what the qualification absolutely needs to do, and what is better considered an optional extra.

In addition to prioritising between the 3 purpose perspectives, it is just as important to prioritise between the potentially relevant categories of purpose within each of them (bearing in mind the full range of 'contrasting opinions' just discussed for each perspective). These alternatives might have radically different implications for qualification design. Consider the information perspective, for instance. Initial purpose analysis discussions might lead to the conclusion that a reformed school-leaving certification system should satisfy 2 quite different assessment purposes (from the list of 22 presented above):

- qualification: to identify students who have achieved sufficient subject-level knowledge, skill, and understanding to warrant enrolling them for higher-level study in the same or related subject areas
- national (system) monitoring: to decide whether the attainment of the nation is rising or falling within specified subject domains (English, maths, science)

Critically, purposes as divergent as these would recommend very different qualification design decisions. In fact, to optimise an assessment procedure for the national monitoring purpose, it would be better simply not to utilise the qualification system (which generates results for all students) for this purpose and to design an independent survey instrument instead (which would generate a result for the nation

[37] It is important to appreciate that there has been some resistance to prioritising the engagement purpose in recent years, partly in response to qualifications of the 1980s and 1990s that did exactly this yet were judged not to have been entirely successful (for example, Hargreaves, 1989; Ecclestone, 2002). A strand of this resistance relates to the debate over skills-based syllabuses, where these have been promoted primarily for engagement reasons – to permit maximum customisation to the bespoke interests of individual students – effectively downplaying the importance of acquiring forms of knowledge that might, in theory, have empowered them (see, for example, Young, 2008; 2013).

but not for individual students).[38] The need to prioritise should be obvious here. If the national monitoring purpose was judged to be more of an optional extra – for which fairly low-quality, merely indicative information would suffice – then it might be possible to satisfy this aspiration on the back of a system designed primarily to achieve the qualification purpose. If, instead, the national monitoring purpose were to be prioritised, then it might be better to explore the possibility of achieving the qualification purpose quite differently (for example, by relying on informal teacher assessment judgements and one-to-one interviews with individual students).

3. Systemic requirements

Systemic requirements stem from the fact that individual qualifications do not operate independently of each other, but as part of a broader qualification system, which plays a fundamental role in organising society by regulating access to job opportunities as well as to education and training opportunities that lead to job opportunities. These high-level systemic considerations also have consequences for the design of individual qualifications, which need to be considered in detail when developing and prioritising a purpose profile for the new qualification.

Although, in a sense, these systemic requirements could be considered purpose perspectives in their own right – albeit higher-level ones – it feels clearer to describe them as independent systemic design requirements, on the assumption that they all embody essentially the same generic societal perspective on how the system (and its qualifications) ought to be designed.[39] This societal perspective is seen most clearly through the eyes of those who are at the centre of the system – those who are studying for qualifications and (particularly for younger learners) their parents too – because these are the individuals whose lives are most immediately and acutely affected by decisions concerning how these systems are designed.

Although all these requirements will have implications for the design of any system, and of all qualifications, their precise implications will differ for different systems and different qualifications – as will the importance attached to each implication – which

[38] This is the approach that we take to measuring trends in attainment over time in English and maths, at the end of key stage 4, via National Reference Tests.

[39] An earlier paper discussed the 'contractual perspective' on qualification design, which emphasised the need for ethical integrity, suggesting that this perspective may sometimes trump both the expertise and the information perspective (Newton, 2023b). This was premised on the idea that a qualification represents a 'social contract' between those who study for a certificate and those who use the certificate to regulate access to job opportunities (or to education and training opportunities that lead to job opportunities). From this perspective, the overriding concern is that this contract is managed fairly. In the present report, similar considerations are discussed as high-level systemic requirements related to ethical integrity.

is why they all need to be considered in detail, up-front, before any development activity takes place. The following requirements illustrate a variety of important considerations – related to structural integrity, cognitive integrity, and ethical integrity, respectively – although this list is not presumed to be exhaustive.

Structural integrity

Structural integrity is about securing the system itself. Because qualification systems regulate access to life-changing opportunities – some of them performing this function for hundreds of thousands of learners each year – it is critical that they are built well and built to last. There are various ways in which a qualification system might need to demonstrate structural integrity, but coherent design and resilient delivery would seem to be particularly important.

Coherence

England's qualification systems have evolved over hundreds of years, and hundreds of awarding organisations have contributed to this process, especially in relation to vocational and technical qualifications. Although these systems have responded organically to changing employment and societal demands, they have not always embodied **coherence**. Indeed, during the second half of the 20th century, England's vocational and technical landscape was often described as a "jungle" (De Ville, 1986, page 7), with too many overlapping organisations and too many overlapping qualifications.

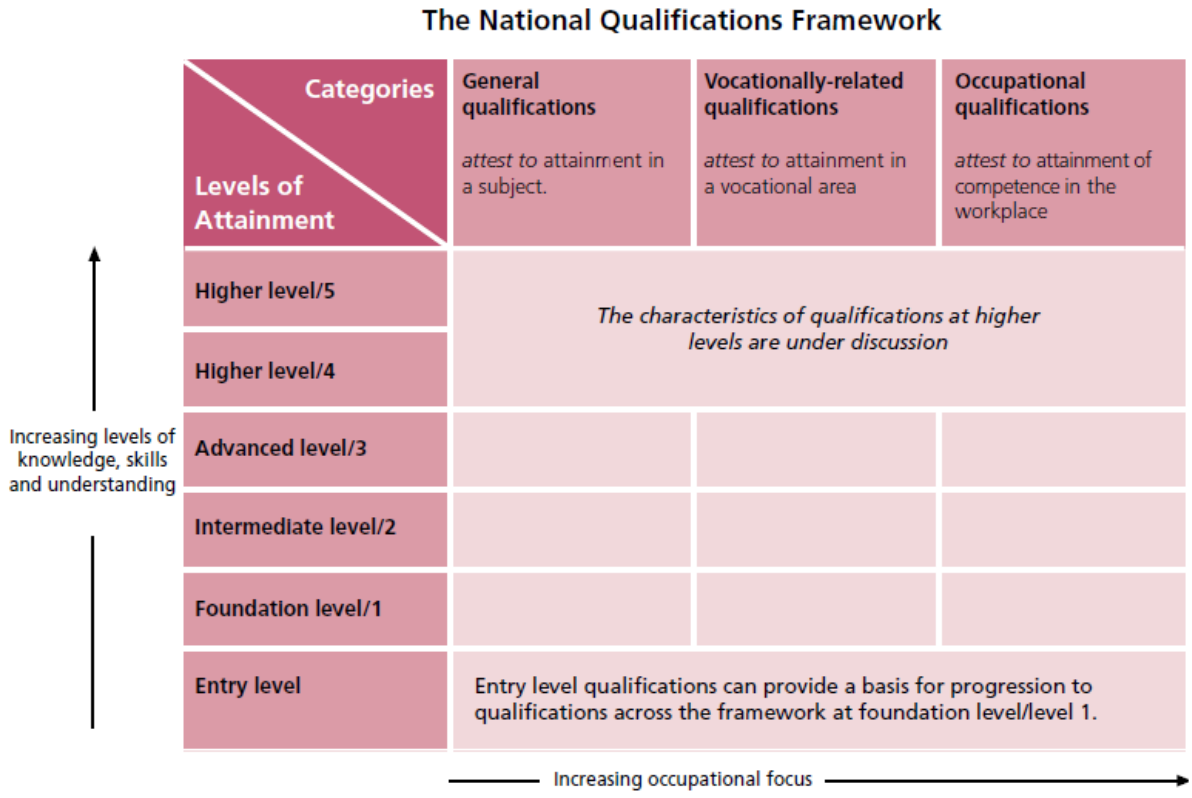
Although there has been far less central planning in England than in many countries, government has intervened at critical junctures to secure coherence, for example, when it co-ordinated the creation of a national system of examining at 16 and over (School Certificate) and 18 and over (Higher School Certificate) during the 1910s. It also intervened during the 1970s to establish 2 new bodies – the Technician Education Council and the Business Education Council – to rationalise the provision of critical qualifications for industry and commerce.

Ideally, a coherent system should provide for the needs of all relevant learners at all relevant stages of their learning. In other words, it should be a comprehensive system. But it should also be parsimonious – in the sense of not being unnecessarily complicated – to avoid providing so much choice that the system becomes incomprehensible and no longer cost-effective. In recent decades, qualification frameworks have been introduced to help engineer coherence and reduce complexity.

Introduced in 2000, as part of the Curriculum 2000 reform effort, the National Qualifications Framework (NQF) provides a good example of this. It had a vertical dimension, to indicate a trajectory of progression in learning through successively

higher qualification levels. And it had a horizontal dimension, to indicate alternative (albeit complementary) learning routes, characterised by general qualifications, vocationally-related qualifications, and occupational qualifications (see Figure 9, reproduced from QCA, 2000, page 5). To be part of this system, qualifications had to be accredited to the framework, as one of the 3 possible qualification types at a particular level. This involved having to satisfy type-specific accreditation criteria, which were intended to promote transparency, quality, and rigour (QCA, 2000).

Figure 9. The National Qualifications Framework



The idea of a qualification type embodies the principle of coherence and parsimony. Grouping qualifications within a type indicates that they serve essentially the same (or a similar) set of purposes, which suggests they are likely to need to be designed in the same (or a similar) way. Ofqual currently specifies these common purposes and design requirements within Qualification Level Conditions for GCSEs, A levels, T Levels, and so on. Within each qualification type, individual qualifications sometimes incorporate different design features, either in ways that are not covered by the common requirements or when specific requirements have been disapplied (within Subject Level Conditions).

Although system coherence is important, there are serious risks associated with over-rationalising a qualification system, especially via heavily prescriptive frameworks that incorporate very detailed accreditation criteria. The most obvious problem arises when qualifications that worked extremely well under a previous framework can no longer be accredited to the new one owing to incompatible design characteristics. Sometimes these misfitting qualifications are reformed, to enable them to satisfy the new accreditation criteria, but then end up no longer fulfilling the purposes they previously served. As noted earlier, the QCF, which superseded the NQF during the late-2000s, was heavily criticised for its very detailed qualification design rules that were not appropriate for all the qualifications it needed to accommodate (Newton, Clarke, et al, 2024).

Because qualifications are related to each other through networks of progression opportunities – and because it would never be possible to reform all these interacting qualifications simultaneously – any attempt to reform a qualification or qualification system is likely to present a threat to coherence (even when it is intended, in the long term, to secure coherence). This is because the reformed qualifications will still be connected, through progression opportunities of one sort or another, to adjacent unreformed qualifications. This helps to explain why reforms are often very ambitious in scope, for example, the Curriculum 2000 reforms, where level 3 general and vocational qualifications were reformed simultaneously to secure their closer integration.

When qualifications are reformed horizontally – for example, focusing on level 3 learners studying full-time in schools and colleges – this will create a new threat to coherence for qualifications that are vertically related to them (in this instance, qualifications at level 1/2 and level 4 and above). Whether, or when, qualification reform is better undertaken vertically rather than horizontally is an important question to consider, especially when it is critical that the system is optimised for progression.

Resilience

The importance of systemic **resilience** has assumed far greater significance in recent years, in the wake of the pandemic and the increasing prevalence of cyber attacks (HM Government, 2025). The importance of qualification system resilience became apparent during the pandemic, when preparations for the summer exam series ground to a halt during the spring of 2020. It was simply not viable for GCSE and A level students not to progress to their next education or training destinations, so alternative arrangements had to be put in place to regulate the process. In the end, we relied upon Centre Assessed Grades rather than exam results.

Resilience concerns the capacity of a qualification system to continue to fulfil its core functions in the context of a large-scale external crisis. Examples of crises that might have very serious consequences for a qualification system include:

- a pandemic that makes it impossible to run an exam series
- a cyber attack that makes it impossible for a major exam board to deliver their exams or release results
- a major exam board failing financially and having to be shut down
- industrial action by postal delivery workers that makes it impossible to transport question papers and scripts to and from schools and colleges
- industrial action by teachers that impacts negatively on learning and makes it impossible to conduct non-exam assessments
- a major war that disrupts teaching, learning, and certification in all sorts of challenging ways
- a national student protest that results in the majority of students failing to comply with certification arrangements

Although it might not be possible to anticipate when an emergency like this is likely to arise, it may still be important to design the system in ways that help to mitigate risks associated with its occurrence. For instance, in the wake of the pandemic, Ofqual and the DfE decided that guidance should be provided to schools and colleges concerning how to gather evidence of student performance that could potentially be used to support the awarding of grades if a situation arose that prevented normal assessment arrangements from operating.

Unfortunately, different crises present different threats, which means they will require different mitigation strategies. For instance, a strategy that might work in the context of a national pandemic (that did not affect an exam board's ability to gather data and award results) might not work in the context of a cyber attack (that did). Likewise, a strategy that relied on using teacher assessment judgements might not work in the context of a national teacher strike. Furthermore, any active mitigation strategy would need to be considered in relation to the likelihood of the crisis occurring, in terms of the additional costs and burdens that the strategy would incur.

Cognitive integrity

The idea of **cognitive integrity** assumes a qualification system can only fulfil its core functions if it is capable of securing both understanding and trust. In other words, to secure their ongoing willingness and ability to participate actively and optimally, it is important that those who participate in the system both understand how it is supposed to work and trust that it actually is going to work.

This recognises that consent cannot simply be presumed, because it is grounded in respect for the system, which needs to be earned. Students, for instance, can withdraw their consent by choosing not to study certain qualifications, or not to follow

particular qualification pathways. Admission officers or employers can withdraw their consent by choosing not to use results from particular qualifications for recruiting or selecting purposes. Parents can withdraw their consent via elective home schooling or, more worryingly, by allowing their children to absent themselves from school for extended periods. The public can withdraw its consent by expressing uproar over qualifications or qualification systems that seem not to be defensible for one reason or another.

Transparency

Qualifications and qualification systems need to be **transparent** to secure understanding. The more coherent the system is – in terms of comprehensiveness and parsimony – the more likely it is to be understood. But transparency is an important consideration in its own right. It is quite possible, for instance, to imagine a system that was entirely coherent but that lacked transparency and was not well understood by many of its core participants, including learners and those who use results, let alone members of the public.

Governments, regulators, awarding organisations, and others too can take steps to promote system transparency independently of prior qualification design decisions. Efforts to bolster understanding and prevent misconceptions are especially important during the early years of a new qualification or qualification system. Yet design decisions are sometimes taken specifically with transparency in mind. For instance, transparency is a primary consideration when decisions are made concerning qualification grading scales. Grades are tools for communicating information about learners' achievements. So, it is critical that they convey the information they need to convey as clearly as possible, and it is equally important they do not accidentally appear to convey inappropriate (surplus) meaning.

When the GCSE grading scale was revised during the late-2010s – to incorporate additional grades that would stretch and challenge the highest-attaining students – the need for transparency was paramount. It was decided to adopt a numerical scale, to minimise the potential for confusion with the previous alphabetic one. It was decided to grade from 9 (top) to 1 (bottom) because it made sense to associate higher numbers with higher-value grades. And it was decided to establish an empirical link between the old and new grading scales, particularly at grade C, given the historic significance of the threshold between grade C and grade D. So, for each GCSE subject qualification, the bottom of the grade 4 scale for the new qualification was linked back to the bottom of the grade C grade scale for the old one. These design decisions were publicised via an extensive communications campaign.

It is fair to say that threats to the transparency of the GCSE grading scale still exist. For example, since 2015, qualifications in England, Wales, and Northern Ireland have been regulated independently, and NI decided not to adopt a radically new

grading scale for its own GCSEs. Instead, it retained the A* to G grading scale, albeit introducing a new C* grade to parallel England's grade 5. This presents a transparency challenge because it is possible to study GCSEs from both England and NI (in NI). The situation in Wales is different again, as Wales retained the A* to G scale without introducing a C* grade. Wales and NI also adopted different stances on the A* threshold. When the new grading scale was introduced in England, the standard of the NI grade A* threshold was realigned to the standard of England's grade 9 threshold, while the standard of the Welsh grade A* threshold remained unchanged (at a standard lower than the grade 9 threshold).

Another challenge relates to widespread misinterpretation of the GCSE grade 4 threshold, particularly in the wake of a distinction that was drawn by the Secretary of State, Justine Greening, when it was introduced:

> Rather than reporting on the "good pass", we will instead distinguish between a grade 4 as a "standard pass" and a grade 5 as a "strong pass" and report on both.
>
> Under the new system, a grade 4 and above will be equivalent to a C and above. This is – and will remain – the level that pupils must achieve in order not to be required to continue studying English and maths post 16. Therefore, a GCSE pass at new grade 4 will continue to have real currency for individual pupils as they progress to further study and employment. Where employers, FE providers and universities currently accept a grade C we would expect them to continue recognising a grade 4.
>
> (Greening, 2017, page 1)

A subsequent report from the Association of School and College Leaders (ASCL) reflected on the apparent implications of this narrative as follows:

> The demoralising impact of the language of these descriptors should not be underestimated. The descriptor of a grade C as a good pass in the previous system at least held on to the notion that lower grades constituted a pass. If a grade 4 is now a standard pass and a grade 5 a strong pass, what does that mean for grades 1, 2, and 3? We have avoided the language of failure in this report as far as possible for fear of reinforcing the inevitable conclusion of these changes.
>
> (ASCL, 2019, page 8)

With repeated reference to GCSE as a "system which 'fails' a third of students" (page 30) and "fails so many" (page 5), perhaps more could have been done to avoid the language of failure. But, ASCL was making a point about the likely misinterpretation of the grade 4 threshold, and this misinterpretation seems not to be uncommon (see, for example, Rethinking Assessment, 2023). This is unfortunate as

the whole point of the GCSE system was to provide a positive indication of success for all students (Kingdon & Stobart, 1988), including those whose ultimate level of attainment would be among the lowest in each cohort.

Plausibility

Narratives surrounding the design, development, and delivery of qualifications and qualification systems need to be **plausible** to secure trust. That is, they need to hold water when awarding organisations are held to account.

Historically, the custodians of England's qualification systems have had limited success in constructing plausible narratives, and in earlier decades they seem to have made limited effort along these lines. A century ago, these custodians tended to be internationally renowned universities (including the universities of London, Oxford, and Cambridge) or bodies with longstanding royal charters (such as City & Guilds of London Institute or the Royal Society for the Encouragement of Arts, Manufactures and Commerce). Prestigious organisations like these may have felt little need to account for their practices back then. But times have changed.

In her 2002 Reith Lecture on Trust and Transparency, Onora O'Neill suggested that: "openness and transparency are set to replace traditions of secrecy and deference, at least in public life" (page 2). Yet she added that openness and transparency were not enough. Trust only emerges once we have been able to test the evidence that is made available to us, to establish its veracity and reliability. For those who are responsible for qualification systems, this emphasises the importance of making the necessary evidence available, so that members of the public can judge for themselves the plausibility of qualification claims and practices, and therefore whether to place or refuse trust (O'Neill, 2013).

Historically, qualification practices and claims were not very amenable to active checking of this sort. As a secretary to the Joint Matriculation Board once put it:

> There has previously not been any reluctance on the part of the Board to answer fair enquiries reasonably made by responsible folk.
>
> But on the whole it has preferred to get on with the job and not to proffer comment or information where it has not been asked for.
>
> (Petch, 1963, page 3)

This observation was made in the context of allegations that qualifications were insufficiently comparable, which begged the question of why the boards had not published any evidence to the contrary, which:

> would have helped to prevent the spread of 'horror-stories' about such things as lack of equivalence which is an inevitable concomitant of the present cloak of secrecy
>
> (Wiseman, 1961, page 154)

During the 1970s, the boards invested heavily in research of this sort, and did begin to publish relevant evidence, including the first of a series of reports that summarised outcomes from an ongoing programme of comparability studies (Bardell, Forrest, & Shoesmith, 1978).[40] Yet publications of this sort failed to silence critics, particularly when some of the claims that they made seemed to lack plausibility.

Claims concerning the maintenance of exam standards over time have always caused problems for the boards, particularly when trends in pass rates have been inconsistent with presumptions concerning trends in educational standards.[41] During the 1970s, for instance, public debate was dominated by concerns over falling educational standards. Yet O level and A level pass rates remained remarkably consistent over time, even as the proportion of the national cohort sitting these exams increased. This led to allegations that the boards were 'norm-referencing' exam standards, thereby concealing the true drop in educational standards.[42] The boards denied this flatly, explaining that grades were always referenced to students' actual attainments, but the myth of norm-referencing proved extremely hard to bust, and the concomitant threat to trust has persisted.

Ironically, exactly the opposite situation threatened trust in the system during the early-2010s. A level pass rates – technically, the cumulative percentages of students who achieved grade E or better in each subject area – had been rising consistently for decades, yet there was growing scepticism over whether this could be attributed to changes in educational standards over time. The question of whether exam standards were, in fact, gradually falling over time has never been answered definitively. Indeed, some exam board experts argued it was impossible to answer

---

[40] Its Foreword read as follows: "In presenting this booklet to the public […] we in the GCE boards have found ourselves in a dilemma. If we merely state that comparability exercises are regularly conducted and do not show our hand, we appear to have something to hide. If we try to explain them, their complexities and limitations invite misunderstanding and misrepresentation. On balance, the preferable alternative seemed to be to 'publish and be damned'. We have, and probably shall be." (Bardell, Forrest & Shoesmith, 1978, Foreword)

[41] The term 'pass rates' will be used here and subsequently as shorthand for the cumulative percentages of students who achieve particular grades (or better) from one year to the next.

[42] This allegation suggested that the boards simply awarded grades to the same proportions of students each year, regardless of whether their actual attainments might have been better or worse from one year to the next, or from one board to the next.

the question definitively, due to a lack of agreement over what it might mean for exam standards to rise or fall over a period that spanned decades (Newton, 1997).

By the early-2010s, however, even the experts had given up arguing the toss, as annual reports of pass rates continuing to rise – for the 26th year in a row, for the 27th year in a row, for the 28th year in a row, and so on – began to sound increasingly implausible. This led Ofqual to implement a much stricter policy on the maintenance of exam standards over time, with far greater priority given to statistical indicators (which were now far more robust than in previous decades) and less weight given to the judgement of senior examiners (which was suspected of ratcheting exam standards downwards – albeit imperceptibly and unintentionally – through a tendency to give students the benefit of the doubt when choosing between adjacent grade boundary marks when they were unsure which one to opt for).[43]

The strategy of giving far greater weight to statistical indicators has been described as the 'comparable outcomes' approach, which has been associated with a new wave of allegations that exam standards are, in effect, norm-referenced. Thus, ASCL proposed that:

> Comparable outcomes mean that one third must 'fail' in order that two thirds 'pass'
>
> (ASCL, 2019, page 6)

Technically, there is an important distinction to be drawn between applying the comparable outcomes principle (which happens during the first couple of years of a reformed qualification) and maintaining standards by giving far greater weight to statistical indicators (which happens once a reformed qualification has bedded in). What this means is that today's grades are referenced to students' actual attainments, just as they always have been. It is true that pass rates do tend to remain fairly stable over time. But we now have convincing evidence from the National Reference Tests that educational standards in English and maths seem to be fairly stable too, which suggests pass rate stability (in general) should not be that surprising.

To conclude this section, it is not true that Ofqual's grading strategy is tantamount to norm-referencing, let alone that it condemns a third of GCSE students to failure. Yet stakeholders do need to be persuaded of this, to secure trust. Ultimately, the plausibility of the narratives that underpin our qualification systems cannot be presumed – and sometimes their case will need to be argued strongly. When a narrative continues to sound implausible, despite repeated attempts to argue its

[43] See Newton (2022) for discussion of the persistent myth of norm-referencing. See Newton (2020a) for information on Ofqual's approach to maintaining examination standards over time from the early-2010s onwards.

case, it may even become necessary to change the system, and to adopt a new narrative.

Ethical integrity

Ethical integrity relates to the perceived fairness of a qualification or qualification system, which is critically important for securing public confidence and consent. Although no one would doubt the importance of fairness in this context, the idea of fairness is tricky to pin down – let alone to secure – because it means different things in different contexts, and because different stakeholders are liable to express different intuitions concerning what counts as fair (and unfair) within any particular context.

The idea of fairness in assessment has become far more prominent in recent decades, particularly within the educational assessment literature (see, for example, Jonson & Geisinger, 2022). Scholars have developed the idea in different directions. Nisbet & Shaw, for instance, identified 6 different senses of fair and unfair when applied to assessment:

> The four main senses of 'fair' are as follows:
>
> 1. A *formal* sense, denoting accuracy or appropriate application of a rule or design.
>
> 2. An *implied contractual* sense, in which something is fair if it meets the *legitimate expectations* of those involved.
>
> 3. A *relational* sense – treating (relevantly) like cases alike. This is reflected in a requirement of comparability between marks/grades awarded to candidates, and also in the concept of impartiality, which implies disregarding irrelevant considerations, and can be applied to judgements about one person/instance, with no others immediately concerned.
>
> 4. A *retributive* sense, in which an outcome is regarded as fair if it is an appropriate reward (or penalty) for what has gone before.
>
> We also distinguished two senses which were used mainly in discussions of *un*fairness:
>
> 5. *Consequential* – in this sense, an assessment might be judged as unfair if its outcomes might be used as a basis for unfair actions in the future.
>
> 6. *Retrospective* – in this sense, an assessment might be judged as unfair if its outcomes were the consequence of unfair/socially unjust actions in the past.
>
> (Nisbet & Shaw, 2020, pages 146 to 147)

A different kind of framework for thinking about fairness was proposed by Colquitt & Rodell (2015), for use in organisational contexts that involved selection, performance, evaluation, training, compensation, and so on. They distinguished between 4 types of 'justice rule' (which are illustrated in Figure 10, adapted from Colquitt & Rodell, 2015, page 189). Although targeting slightly different contexts, the contrast between Colquitt & Rodell and Nisbet & Shaw is useful, and the approaches certainly overlap. For instance, the retributive sense of fairness from Nisbet & Shaw overlaps with the distributive rule described by Colquitt & Rodell. Both seem to be grounded in the idea of just deserts. Yet the approaches also depart from each other in significant ways. For example, the informational rule described by Colquitt & Rodell does not seem to have a direct counterpart in the Nisbet & Shaw framework.

The fact that these 2 approaches are different, albeit complementary, illustrates that there are no generally agreed frameworks that can be used to represent a full range of issues or activities related to fairness concerns. Furthermore, the fact that they identify so many distinct senses of fairness in assessment highlights the challenge of designing a qualification or qualification system with ethical integrity in mind.

Figure 10. Justice rules from Colquitt & Rodell (2015)

Type	Name	Description
Procedural	Process control	Procedures provide opportunities for voice
	Decision control	Procedures provide influence over outcomes
	Consistency	Procedures are consistent across persons and time
	Bias suppression	Procedures are neutral and unbiased
	Accuracy	Procedures are based on accurate information
	Correctability	Procedures offer opportunities for appeals of outcomes
	Representativeness	Procedures take into account concerns of subgroups
	Ethicality	Procedures uphold standards of morality
Distributive	Equity	Outcomes are allocated according to contributions
	Equality	Outcomes are allocated equally
	Need	Outcomes are allocated according to need
Interpersonal	Respect	Enactment of procedures are sincere and polite
	Propriety	Enactment of procedures refrain from improper remarks
Informational	Truthfulness	Explanations about procedures are honest
	Justification	Explanations about procedures are thorough

Neither of these 2 approaches indicates that it is also helpful to distinguish between dimensions of fairness that relate specifically to measurement (and the information perspective on qualification purposes) and dimensions of fairness that seem to be judged according to somewhat different criteria. Note how each of the following 3 dimensions is relevant to qualification design, albeit in different ways:

- the **technical** dimension – focused on **measurement** fairness
- the **interpersonal** dimension – focused on **treatment** fairness
- the **supervisory** dimension – focused on **custodial** fairness

Figure 11 explains what each of these 3 dimensions is intended to capture, also identifying their differing bottom lines. Note that the ultimate criterion of measurement fairness is accuracy, which is also the ultimate criterion in relation to the information perspective, as discussed earlier.[44] So, these 2 concepts are clearly in the same ballpark. The following section on the systemic requirement for robustness explains how these concepts differ. The subsequent section on justice explores the concept of custodial fairness in more detail.

[44] Earlier, we noted that the information perspective bottom line was to ensure that the information that qualification results provide is of the right kind and of sufficient accuracy (to support effective decision-making).

Figure 11. The 3 dimensions of fairness, their characteristics, and associated activities and issues

	Technical Dimension	Interpersonal Dimension	Supervisory Dimension
Goal	Measurement fairness	Treatment fairness	Custodial fairness
Ultimate criterion	Accuracy: Qualification result users should be provided with accurate information about learner competence.	Considerateness: Learners should be treated with consideration during all stages of the qualification lifecycle.	Defensibility: The design, development, and delivery of the qualification should be consistent with an overarching duty of care, and publicly defensible.
Related criteria	Validity, freedom from bias, reliability, comparability.	Dignity, respect, politeness, sensitivity, impartiality.	Acceptability, credibility, tolerance, humanity.
Fairness aspiration	All learners will be measured accurately via the assessment procedure, and all groups of learners will be measured equally accurately.	All learners will be treated with consideration during all stages of the qualification lifecycle, and all groups of learners will be treated with equal consideration. (Consideration will also be extended to anyone who participates in activities associated with running the qualification.)	The potential for harm arising from the design, development, or delivery of the qualification will be considered thoroughly – with respect to learners from a wide range of backgrounds in a wide range of circumstances – and systems or practices that are inconsistent with an overarching duty of care, and that would be hard to defend publicly, are avoided or amended.

	Technical Dimension	Interpersonal Dimension	Supervisory Dimension
Examples of activities intended to improve fairness within each dimension	Adopting a Universal Design approach when designing qualifications. Bias training and materials. Scoring exam scripts blind to candidate details. Providing Reasonable Adjustments and Special Considerations.	Processing qualification materials through Sensitivity Reviews Engaging effectively with learners, and other key participants (learner voice) Escalating safeguarding concerns over potential for harm	Applying the comparable outcomes principle when grade awarding during periods of reform. Providing compensation. Publishing information on likely misuses of results.
Examples of classic fairness cases or concerns	Bias against ethnic minority students attributable to stereotyping in teacher assessment judgements. Bias against lower-socioeconomic status learners attributable to the use of less familiar contexts in constructed response exam tasks.	Questions involving contexts that play to group stereotypes. Lack of diversity among authors of set texts.	Lack of Opportunity to Learn. Qualification flags, which flag whether adjustments or accommodations have been provided. Using different cut-scores (grade boundaries) for different groups, given evidence of substantial differential achievement.

Robustness

The idea of **robustness** is intended to capture how resistant a qualification or qualification system is to well-established causes of measurement unfairness, whether intentional or unintentional. The issue looms largest whenever there seems to be an association between inaccurate qualification results and a questionable qualification practice, for instance, where there seems to be:

- unintentional judgemental bias (exhibited by examiners) in favour of neatly handwritten exam scripts, which leads to higher-than-deserved exam grades
- unintentional judgemental bias (exhibited by teachers) against learners with specific needs or disabilities, which leads to lower-than-deserved teacher assessment grades
- intentional judgemental bias (exhibited by managers) against disliked apprentices, which leads to lower-than-deserved workplace assessment results
- insufficiently directive mark schemes, which result in different judgements across examiners and (consequently) to harsh or lenient grading
- inadequately moderated coursework judgements, which result in different standards across schools and (consequently) to harsh or lenient grading
- insufficiently robust administration arrangements, which fail to prevent students from cheating in their exams or coursework
- insufficiently robust centre controls, which fail to prevent teachers or training providers from intentionally inflating students' teacher assessment grades

Situations like this all refer to result inaccuracy of one sort or another, and for that reason they all fall within the purview of the information perspective. Yet when they give the appearance of being manifestly unfair – and especially when it seems that this manifest unfairness could straightforwardly have been prevented – they assume a deeper significance, which is social and not merely technical. Consequently, they present a far greater threat to public confidence than more covert or abstract causes of result inaccuracy tend to.

For instance, each of the cases noted above might raise an explicit question concerning why the practice-in-question was ever sanctioned in the first place. Why permit teachers to award grades for non-exam assessments, if they are liable to judgemental biases, whether intentional or unintentional? Why permit non-exam assessments at all, given the potential for students to use artificial intelligence to complete coursework tasks for them? Why have any laxity in exam administration arrangements, if there is a possibility that even just a few students will cheat?

The robustness requirement invites us to consider result inaccuracy, as manifest unfairness, through the eyes of those who are at the centre of the system – those who are studying for qualifications and (particularly for younger learners) their parents too – because these are the ones whose lives are most immediately and acutely affected by qualification design decisions. Through the eyes of students and their parents, causes of measurement inaccuracy differ in terms of the threat that they pose to confidence and consent.[45] In particular, the less resistant a qualification is to bias, malpractice, cheating, or fraud, the less confidence it is likely to secure. On this note, it is important to remember that our qualification systems arose from the need to regulate the distribution of education, training, and employment opportunities on a fairer basis than was commonplace during the early-1800s – when payment, patronage, and nepotism were the norm. So, it should come as no surprise when media reports lambast qualifications that are insufficiently resistant to these obviously inappropriate practices (Newton, 2023b).

Threats to the ethical integrity of a qualification system have the potential to create extremely perplexing qualification design dilemmas. For example, situations sometimes arise in which the existence of widespread cheating or fraud becomes well established, and it becomes necessary to redesign the qualification system to make it more resistant to subversion of this sort. While necessary to shore up public confidence in the qualification, this might well mean departing from an approach that would previously have been judged to have been optimal (in the absence of evidence of widespread cheating or fraud). Consequently, this might make it harder for learners in certain circumstances to demonstrate their knowledge, skill, and understanding under the more robust approach than under the previous one, for instance, if a system were redesigned to incorporate less teacher-based assessment and more exam-based assessment (and there were no longer any opportunities for providing bespoke assessment arrangements designed to optimise the accuracy of assessment for learners with particular challenges).

Sometimes, however, situations arise that are even more perplexing, for instance, when the existence of cheating or fraud is well established, but its prevalence is unknown and could potentially be relatively low. In situations like this, the need to shore up public confidence in the qualification might still loom large, as qualification cheating or fraud can easily become a headline story within national media outlets. Yet whether the threat to public confidence might be sufficiently high to justify departing from an approach that might well have been optimal in the absence of evidence of cheating or fraud would be an extremely challenging call to make, where the actual prevalence of cheating or fraud remained unknown.

---

[45] Conversely, a measurement expert might see inaccuracy as inaccuracy, period, with no reason to tolerate one cause of error more or less than any other.

Justice

The idea of **justice** is intended to capture issues of fairness in assessment that extend beyond the technical dimension (measurement fairness) to the interpersonal dimension (treatment fairness) and especially to the supervisory dimension (custodial fairness). The requirement for justice will be illustrated in relation to custodial fairness, with reference to 'applying the comparable outcomes principle when grade awarding during periods of reform' and 'providing compensation' (both from Figure 11).

In these 2 examples, qualification design decisions that would appear to follow from a purely technical analysis would also appear to threaten the ethical integrity of the system. In circumstances like this, it becomes extremely challenging to determine an appropriate course of action. In very exceptional circumstances, like those discussed below, we may decide to prioritise fairness for students over the precise alignment of attainment standards with previous years.

Comparable outcomes

The GCSE and A level exam boards have formally recognised the **comparable outcomes principle** since the late-1990s. It was applied during the Curriculum 2000 reforms, and it was subsequently applied during the Gove reforms of the mid-2010s, when GCSEs and A levels were revised to:

- adopt more rigorous subject content matter
- utilise more written exams and less teacher assessment
- switch from modular (unitised) assessment to linear (terminal) assessment

When qualification reforms are as radical as this, we would expect that it will take some time for teachers to get up-to-speed with teaching the new (unfamiliar) content. In other words, in the first year of the new qualification, we would expect teachers to teach the new syllabus somewhat less effectively than they had come to teach the previous one (after many years of experience). If so, then we would expect the first cohort of students, on average, to achieve a lower level of attainment (on the new syllabus) than the previous cohort of students had achieved (on the old one). In fact, we have good empirical evidence to suggest that this is the case, and that it takes a few years for the effect to wash out (Cuff, et al, 2018; Newton, 2020a; Newton, 2020b).

Assuming that this the case, by applying (essentially) the same exam standard to the new qualification as we applied to the old one – that is, if we rewarded similar performances with similar grades – we would expect to see pass rates fall for the

first cohort of learners.[46] All other things being equal, we would not expect pass rates to catch up until teachers had become as effective at teaching the new syllabus as they used to be in the final year of teaching the old one.

This raises an ethical dilemma. Why should students be penalised for their teachers' lack of experience in teaching the new syllabus? The situation is particularly acute for A level students who might even find themselves competing for university places against students who achieved their A level results in the previous year. Over the past few decades, we have agreed that this situation is not acceptable, and Ofqual now requires exam boards to apply the comparable outcomes principle during grade awarding for the first few years of a new qualification.

The name given to this principle derives from a report that explained how it ought to be applied:

> candidates taking the new examinations should receive, as a group, comparable grades [that is, comparable outcomes] to those which they would have received had they followed the old courses
>
> (Cresswell, 2003, page 14)

In other words, although their exam performances might not be comparable during the first few years, their grade outcomes ought to be (on average). This is not the same as awarding an identical distribution of grades from one year to the next, which might be described as norm-referencing. If, for instance, the 2 cohorts were quite different in terms of their prior attainments – for instance, if the reformed syllabus attracted many more high-achieving learners – then it would not be defensible to award exactly the same distribution of grades, pre-reform versus post-reform. Instead, the exam boards use statistical prediction matrices to adjust for differences in prior attainments from one cohort to the next. In other words, if the pre-reform and post-reform cohorts were:

- equivalent in terms of their prior attainments, then grade boundaries would be adjusted to ensure that the overall grade distributions were also equivalent, but if they were
- different in terms of their prior attainments, then grade boundaries would be adjusted to reflect the anticipated impact of those differences on their current attainments (and therefore their grades)

[46] The use of 'essentially' in parenthesis acknowledges that it is not at all straightforward to define what it means to apply the same standard across qualifications that assess (somewhat) different content in (somewhat) different ways, let alone to link these standards empirically.

This process is repeated for a few years, until we can be confident that teachers are up-to-speed, and we then revert to the traditional approach to maintaining exam standards, which we refer to as 'attainment-referencing' (Newton, 2020a).

Providing compensation

The pandemic threw up all sorts of ethical dilemmas but, for qualifications, perhaps the most challenging decision related to the process for awarding grades in the summer of 2021, more than a year after COVID-19 had first struck. The process for awarding grades in the summer of 2020 was challenging enough. After exams had been cancelled, the intention was to provide results for GCSEs, A levels, and many other qualifications too, using statistically moderated Centre Assessment Grades (CAGs). Unfortunately, concern over anomalous result patterns led to the moderation process being judged to have been unfair, and students ultimately ended up being awarded their unmoderated CAGs.

Yet the situation in 2020 was relatively straightforward in comparison with that in 2021. In the first year, students had more-or-less completed their courses before the pandemic led to students having to remain at home. So, it was not unreasonable to ask teachers to derive their CAGs by predicting the grade that students would most likely have achieved, in the summer of 2020, based on their experiences of having taught and assessed them until the end of the previous term. In 2021, the situation was completely different, after the pandemic had seriously compromised teaching and learning across the country for more than a year.

The most problematic aspect of this situation was that different students had been affected in different ways. Some students were not actually that badly affected. Although the vast majority of students had to study at home (rather than at school) for some of the pandemic, some students had very accommodating home-learning arrangements. And some of these students learnt better under these circumstances than they would have done had the pandemic never happened. Other students had home-learning arrangements that were entirely unsatisfactory. And some of these students spent more time learning at home than others. Yet other students suffered family bereavements, and so on. When we apply the comparable outcomes principle, it is often not unreasonable to assume that all students will have been affected to a similar extent during the first year of the new qualification. Conversely, learning was highly compromised for some students in summer 2021, significantly compromised for many, while for others it was positively affected. Under these circumstances, applying the comparable outcomes principle, as a blanket adjustment, would simply not have been appropriate.

It was not at all obvious how to proceed when exams were abandoned for the second year of the pandemic. It would have been possible to ask teachers to assess students based on their actual attainments, by summer 2021, but their grades would

then have reflected the differential impact of the pandemic in full, which would have been unfair to the most badly affected students. Alternatively, teachers could have been asked to predict the grades that students would have achieved, had the pandemic never happened. But, then on what basis would they have done so, as actual attainments in 2021 would have been irrelevant to a judgement of this sort? Also, in a sense, it would have made a mockery of the learning that had actually been achieved during the pandemic, under such awful conditions. And what about those students whose achievements were better than might otherwise have been the case – would they have ended up with grades that were lower than the standard of their actual attainments? That would have been uncomfortable, to say the least.

In the end, a compromise was reached in which teachers were asked to provide some **compensation** for the impact of the pandemic, but only at the class level, and only in terms of how much of the curriculum it had been possible to teach that class. In other words, teacher assessed grades were supposed to represent students' actual attainments, by summer 2021, but only on those elements of the curriculum that they had actually been taught. This amounted to partial compensation for the (differential) impact of the pandemic, but certainly not full compensation.

A couple of years later, a situation arose for which parallels were drawn with the challenge of teaching and learning during the pandemic. This was for schools and colleges that had been badly affected by the long-term consequences of premises having been built using reinforced autoclaved aerated concrete (RAAC). The seriousness of these consequences meant that students in some schools were learning in hallways, and some schools had even had to close for extended periods, just as when the pandemic struck and online learning became the new norm.

One of the most badly affected schools was St Leonard's Catholic School in Durham, and academics from Durham University calculated its summer 2024 exam grades could fairly be increased by 10%, by way of compensation. After considering the matter extensively, the DfE and Ofqual agreed that compensation would not be appropriate in this situation. It was certainly regrettable that students had had to learn under unsatisfactory conditions, but the idea of compensation is so antithetical to the principle of educational certification in England that it could not be extended beyond the very limited amount that was granted during the hugely abnormal context of the pandemic. After all, students' achievements are affected by all sorts of regrettable circumstances every single year – when a parent dies, or when a student spends an extended period of time in hospital, or suchlike – but none of these circumstances is taken into account when grades are awarded.[47] In the end,

[47] It is not unreasonable to propose that, if all the factors that result in suboptimal learning from pre-school to the end of secondary were to be compensated for, then every single student would end up with a profile of perfect grades in their GCSE and A level exams. Note that awarding bodies do

students at St Leonard's ended up performing better in their 2024 exams than they had done in 2023, even without any compensation.

Other allegations of harm

A wide variety of concerns might be classified as issues of justice in assessment. This would include any allegation of **harm** associated with the operation of a qualification system. For instance, YoungMinds (2025) has recently published a report, which argues that:

- exam stress is driving school absence
- exam stress is causing mental health problems
- exams have a greater negative impact on mental health than social media

As noted in Figure 11, issues such as these need to be considered thoroughly, and systems or practices that are inconsistent with an overarching duty of care, and would be hard to defend publicly, are avoided or amended. This does not mean any substantiated allegation of harm will necessarily lead to the overhaul of a qualification system. For any particular allegation, the issue that would need to be investigated is the likelihood of judging the harm to be tolerable in relation to the individual, organisational, economic, and societal benefits associated with operating the system, and in relation to the potential for preventing or mitigating the harm through alternative qualification design decisions.[48]

4. Constraints

The purposes and requirements that drive qualification design are moderated by pragmatic constraints related to factors as varied as logistics, politics, economics, and the law. It is well-known that assessment design decisions are guided by considerations of this sort, especially those related to cost-effectiveness, which is part of the justification for using assessment formats that lack a certain amount of authenticity (such as multiple-choice testing). Yet the same considerations need to be borne in mind when deciding upon qualification structure and when specifying syllabus outcomes too. The following sections illustrate considerations of this sort, without dwelling excessively on the most obvious ones. This remind us that qualification design is not a matter of maximising fitness-for-purpose, but of

operate a special consideration process, which accommodates situations in which student learning has not been compromised, but the ability to demonstrate that learning in an assessment has been.

[48] For instance, the YoungMinds report suggested that, while GCSEs and A levels should continue to incorporate exams, it would be possible to reduce their pressure by reformatting them, for example, as modular exams, open-book exams, or on-demand online exams.

optimising it in relation to a variety of concerns. It is important to understand that qualification design is always a matter of trade-off and compromise.

Financial costs

The most obvious pragmatic constraint is the availability of funds to pay for a qualification – to cover its design, development, delivery, and review – especially when it is ultimately funded by the public purse. This foregrounds issues of cost-effectiveness and value-for-money. Although high-quality multiple-choice tests are not easy to produce, they can be extremely cost-effective in terms of covering a large amount of syllabus content under very controlled conditions, within a relatively small amount of time, and with potential for automated scoring of student responses. What these tests lack is a certain amount of authenticity – in the sense that life rarely presents challenges in a forced-choice format – and they can also tend towards valuing breadth of knowledge over depth of understanding. Yet under certain circumstances, a cost-benefit analysis might come down heavily in favour of incorporating an assessment format like this.

The more authentic an assessment procedure needs to be, the higher its costs are likely to be: partly because authentic assessments tend to involve extended tasks (which can be costly to develop, administer and mark) and partly because there is likely to be a requirement for learners to attempt multiple tasks (to ensure adequate sampling and generalisability of results). The need for medical students to be assessed comprehensively and authentically has been recognised increasingly over the past few decades, which has led to debate over how best to ensure cost-effectiveness under these circumstances (see, for example, van der Vleuten & Heeneman, 2016).

Presenting a more general case for authentic assessment, Wiggins argued some time ago that:

> The costs are deceptive: while the scoring of judgment-based tasks seems expensive when compared to multiple-choice tests (about $2 per student vs. 1 cent) the gains to teacher professional development, local assessing, and student learning are many.
>
> (Wiggins, 1990, page 1)

This reminds us that the costs of alternative design decisions need to be weighed against benefits that go beyond validity, to include impacts on teaching and learning. Having said that, there may also be opportunity costs to consider, particularly when authentic assessments consume additional time that might otherwise be spent more effectively on learning activities.

Although cost considerations are prominent when designing an assessment procedure, they may also need to be considered when specifying syllabus outcomes. For example, outcomes related to complex practical activities – whether in general, vocational, or technical contexts – might be extremely desirable from an educational perspective, but might simply not be viable from a financial one. This situation might arise if a college were required to provide very expensive specialist equipment for just a small number of students each year. This might ultimately recommend the exclusion of those desirable learning outcomes from the syllabus in question.

Finally, qualifications also need to be designed with a view to the viability of their delivery within schools and colleges. The more niche the specified syllabus outcomes, the lower the demand there is likely to be for the qualification, and the less financially viable it may be to deliver the qualification. Over the past few decades, the viability of many vocational and technical qualifications has been enhanced by modular qualification structures, which enable a certain amount of co-teaching across qualifications and year groups.

Human resources

Beyond the financial costs associated with a new qualification, it is essential to consider the availability and expertise of those who would be responsible for teaching, training, and assessing it, and the implications that this might hold for qualification design. This challenge might sometimes be reducible to a financial cost: if, for example, the viability of a particular approach to marking scripts for a new website development qualification required an exam board to establish, train, and employ a new community of expert teacher-examiners. Yet if teachers of the new qualification preferred to spend their spare time on more interesting or lucrative activities, then the challenge might prove to be more than just financial.

Qualifications that serve established disciplines (such as geography or mathematics) or established occupations or professions (such as nursing or engineering) benefit from extended communities of practice and bodies of received wisdom that have been accumulated over generations. This typically includes professional organisations that assume a remit for influencing education and training policies and practices (such as the Geographical Association or the Royal College of Nursing). In this situation, it might be safe to assume that appropriately qualified recruits should have sufficient domain knowledge and pedagogical content knowledge to be able to teach or train effectively, and a sufficient grasp of domain standards to be able to assess effectively.[49] Under circumstances like this, much can be left to the

[49] Of course, this assumes that there is an adequate supply of appropriately qualified recruits. The crisis in recruiting teachers, particularly for STEM subjects, has been identified as a major barrier to radical qualification reform in England (Freedman, 2022).

professional expertise of these actors, when it comes to teaching, training, and assessment. Exam boards used to rely heavily on this presumption of expertise, for example, when publishing only minimally specified syllabus content lists.

When expertise of this sort cannot be presumed, qualifications will need to be designed accordingly. This might be necessary, for instance, when qualifications are being designed to serve new and unusual jobs, like a personal shopper or a body painter, where there is unlikely to be an established community of practice of any sort. In a situation like this, there would need to be a lot of foundational work to agree the knowledge, skills, competencies, dispositions, and attitudes that were important to these roles, and the standards at which they needed to be acquired. More to the point, this information would need to be communicated very explicitly and clearly to potential teachers, trainers, and assessors, to ensure that they all ended up singing from the same hymn sheet. Extensive training would be required, extensive standardisation, and extensive quality assurance too.

In other situations, it might be possible to assume extensive domain knowledge and a thorough grasp of domain standards, but not pedagogical content knowledge or assessment expertise. This might be the case, for instance, when responsibility for assessment has been devolved to supervisors of apprentices based in their own workplaces. In this situation, qualification training, guidance, and support materials would need to focus heavily on the principles and practices of assessment, and the assessment procedure would need to be designed accordingly, to support consistently accurate judgements across different workplace settings:

> Employers are not experts in pedagogy, qualification design or assessment, unless from a training background, but they can contribute in other ways. Therefore, when asked to engage the focus should be upon what they can uniquely contribute (“authenticity”), what is straightforward, meaningful, related to sector needs and within their competence. Any guidance offered should be clear, not constantly changing, easily followed and unequivocally set out the expectations in terms of the scope of the tasks and time involved.
>
> (Huddleston, 2020, page 25)

Finally, qualification designers cannot simply assume that employers will be willing and able to support colleges in their locality with aspects of qualification delivery, whether that might involve providing work experience opportunities or assuming substantial responsibility for training and assessment. Securing employer engagement and involvement has been a persistent challenge for the technical and vocational education and training sector for many decades (Phoenix, 2025).

Physical resources

More than 2 decades ago, government announced that schools would be given e-learning credits, worth £100 million a year for 3 years, to purchase innovative and professional digital resources (DfES, 2003). Plans were also announced to modernise the exam system, which had been described by the head of the Qualifications and Curriculum Authority, Ken Boston, as a virtually unsustainable Victorian cottage industry:

> And “developmental work” will be carried out to increase the use of information and communication technology in the examinations system, including some pilots of e-marking. Mr Boston has said he is keen to introduce an emailable form of marking, to stop the annual reliance on the Royal Mail to transport the 7 million exam papers each year.

(Curtis, 2003, online)

So, why do most students in England still complete their exams on paper, as they always have done, which then gets transported to processing centres in vans? In fact, the exam boards have made a great deal of progress over the years, which has included scanning script images at processing centres, then sending those images electronically to markers working on-screen at home. Yet the wholesale transition to on-screen assessment – including on-screen delivery – has not been as rapid as many predicted. An investigation by Ofqual a few years ago identified multiple reasons, including:

> a lack of enough devices of a consistent specification for whole cohorts to sit assessments at the same time
>
> insufficient or unreliable internet and local network capabilities
>
> insufficient staff with the expertise or capacity to support the adoption and ongoing use of new online or on-screen assessments
>
> a lack of physical spaces with the electrical and network facilities suitable for large cohorts to take assessments on devices concurrently
>
> insufficient ability of the variable infrastructure to manage security risks

(Ofqual, 2020, page 4)

These issues illustrate the many different ways in which physical constraints can substantially inhibit qualification design aspirations.

Political commitments

Governments sometimes come into power with strong opinions concerning what counts as a good education and, consequently, what good qualifications ought to look like. For many school- and college-based qualifications in England, government has assumed responsibility for specifying core subject content, and this might be chosen to reflect a particular ideological stance, depending on the preference of those in charge. These considerations, however, would be classified under the expertise perspective on qualification purposes, since they relate to design drivers, not design constraints.

Existing political commitments can, however, act as design constraints when they conflict with recommendations that seem to follow from an emerging profile of prioritised purposes. Increased use of teacher assessment provides a good example of this, because it potentially conflicts with a longstanding political commitment to tackle the very real challenge of excessive teacher workload.

Teacher workload has been an issue of concern in England for many decades, and concerns have increased in prominence over time. In recent years, government has commissioned various task forces to address the challenge, and it has developed a website dedicated to improving workload and wellbeing for school staff, which houses a workload reduction toolkit. This includes resources that were developed in the wake of the report of the Independent Teacher Workload Review Group (2016), which focused on eliminating unnecessary workload around marking.

In this context, any proposal to increase the role that teachers play within the qualification system is likely to receive a mixed reaction. On the one hand, it may be welcomed by those who see assessment, including high-stakes summative assessment, as part of the professional responsibility of a teacher, and important for continuing professional development too. Yet on the other hand, if it were to lead to an increase in teacher workload, then it would become an issue of serious concern for teacher associations and teacher unions. These concerns might ultimately end up frustrating the proposal to incorporate teacher assessment.

Laws and regulations

It almost goes without saying that qualifications and qualification systems must be designed to comply with the law. But it is worth illustrating the point by noting that some qualifications have age requirements for exactly this reason. For instance, the VTCT Skills Level 2 Certificate in Wine Service (RQF) has an age requirement of 16 because this is the minimum age at which it is permissible to serve alcohol in a restaurant.

It is also worth mentioning that 'qualifications bodies' have legal duties under the 2010 Equality Act (beyond those that apply to all employers). These regulations state that they must not discriminate against, harass, or victimise people in relation to arrangements for conferring or withdrawing qualifications, and that they have a duty to provide candidates with reasonable adjustments.

Market forces

Many of the qualifications that Ofqual regulates are designed to regulate access:

- to initial job opportunities, as young people transition from school, college, or university into employment, or
- to better-paid jobs, as workers seek more attractive opportunities, either with their current employer or with a new one

So, the nature of the job market will clearly have implications for qualification designers. Considerations of this sort will be relevant to discussion of the expertise, engagement, and information perspectives on qualification purposes (and so, in theory, could be left implicit within them). But it is worth highlighting the potential for market forces to impose hard constraints on qualification design, particularly as this phenomenon is easily overlooked.

For example, it is tempting to think that providing a new suite of qualifications with the potential to up-skill low-skilled workers would incentivise them to achieve these qualifications and progress into better-paid jobs. This might recommend prioritising the expertise perspective on qualification purposes, such that the new qualifications were designed to support the acquisition of specific skillsets that should enable these low-skilled workers to progress. Yet factors such as the attenuation of internal labour markets, the increasing use of agency workers, the cascade down of graduate labour, and the flattening of hierarchies might all conspire to limit progression opportunities for low-skilled workers (Keep & James, 2012). And, if the opportunities were not actually available to them, then simply providing a new suite of qualifications would be unlikely to motivate them to engage in work-related learning, even if these qualifications were optimally designed from the expertise perspective. Under these circumstances, Keep & James suggested that it might be better to shift the focus of policymaking from up-skilling individual workers to trying to create better jobs and more productive and innovative organisations.

Keep & James also identified the challenge of weak labour market regulation in England, a phenomenon that Wheelahan & Moodie (2017) have associated with Anglophone countries more generally. Wheelahan & Moodie argued that qualifications for regulated occupations, especially qualifications that provide a licence to practice, are used by employers to signal that individuals have the specific skillset that is required for the intended occupation. So, for these qualifications, there

is strong alignment between the body of learning that is acquired while studying for the qualification (expertise perspective) and the information that employers infer from results (information perspective). In contrast, qualifications for unregulated occupations tend to be used to indicate a more abstract 'potential' for productivity, and to screen individuals accordingly. This suggests a certain amount of misalignment between the expertise that is acquired (and certified) and the information that is inferred: qualifications are designed to certify achieved learning (in relation to specified syllabus outcomes) and can they provide no more than a rough, heuristic indication of more abstract constructs, like potential productivity.

The implication is not so much that employers should refrain from interpreting qualification results as proxy indicators of potential productivity and stop using them for screening purposes. It would be very hard to shift established hiring practices. The implication relates more to the body of learning that qualifications for unregulated occupations ought to facilitate. If employers are not especially interested in the exact nature of the body of learning certified by qualification results, then this suggests that syllabus outcomes ought to be weighted towards what would be of most value to learners and to society (rather than to employers, per se). When combined with the observation from Wheelahan & Moodie (2017) that individuals who do not intend to enter a regulated job market frequently progress into jobs that are unrelated (or weakly related) to their achieved qualifications, this might suggest that it would be better for their syllabus outcomes to be broadly defined, to open up a wider range of progression opportunities, rather than being narrowly defined in terms of specific occupations (Keep, 2012).

One final point to mention in this section is the operation of a qualifications market in England – the fact that awarding organisations may compete for market share by offering comparable qualifications – which is a relatively uncommon situation, internationally speaking. This, too, can place constraints on qualification design. For example, if an awarding organisation wishes to make available a new qualification that is intended to compete with a similar, preexisting one (made available by another awarding organisation) then it will need to ensure that standards set for the new qualification are consistent with standards set for the preexisting one. This is important for ensuring that the new qualification does not undercut the preexisting one in terms of its standards, potentially making it more attractive to customers.[50]

Another hard constraint in this context relates to the viability of developing new qualifications for very small, niche, cohorts of learners, even when there might be a strong case for the new qualification from a societal perspective. For instance, many of the modern foreign languages that are currently examined at GCSE (like Persian,

[50] This is sometimes described as a 'race to the bottom' scenario. Ofqual's Condition H3.1 formalises the constraint.

Japanese, or Gujarati) have very small cohort sizes, which raises the question of their financial viability for an awarding organisation. Historically, many small-entry general qualifications have been operated by individual exam boards at a loss, effectively being cross-subsidised by large-entry qualifications. No doubt, plans for other qualifications will have failed to get beyond the planning stage, owing to the lack of a viable business case.

Purpose analysis

Qualifications and qualification systems need to be fit for the purposes that they are intended to serve, which is ultimately determined by how well they have been designed. Unfortunately, there are all sorts of purposes for which they might be designed, and they always need to be designed to serve multiple purposes simultaneously. Furthermore, because these purposes sometimes recommend conflicting design decisions, it is essential to establish their relative importance for when design compromises must be made.

Purpose analysis involves anticipating and interrogating design implications that follow from alternative purpose-related aspirations, including their likely consequences. This embodies the principle ‘think first, then do’ from the book ‘How Big Things Get Done’, which explains that projects do not go wrong, they start wrong, so we must put “enormous care and effort into planning to ensure that delivery is smooth and swift” (Flyvbjerg & Gardner, 2023, page 18). When potential problems become apparent during the planning stage, they can be averted. When they only become apparent during the delivery stage, they can be catastrophic. Qualifications are too important to fail for any reason, but certainly not for lack of adequate planning.

Rational

It seems fashionable for policy theory to emphasise that policymaking cannot be understood as though it were a straightforwardly rational process, with clearly ordered stages of decision making (Cairney, 2020). Yet this does not diminish the need for detailed consideration of better and worse approaches to policymaking. Consequently, this report has provided an unashamedly rational account of principles that underpin high quality decision making in relation to the design of qualifications and qualification systems. It may not always be obvious how best to apply the principles and frameworks we have developed (see below) but that is not to deny their importance in guiding effective policy and practice.

Importance

The importance of purpose analysis is brought home by the prevalence of failed qualifications in England, particularly within the landscape of regulated vocational and technical ones (Newton, Clarke, et al, 2024). Examples include the General National Vocational Qualification (GNVQ), as well as its successor, the Diploma qualification.

In her analysis of GNVQ policymaking, Ecclestone observed that these radically new qualifications were expected to:

> - motivate learners who would otherwise not stay on in post-16 education or who were disaffected in Key Stage 4 by responding to, and rewarding, learners' expressed interests and notions of relevance;
> - expand routes into higher education whilst also making sure that expansion did not lead to over-subscription for limited places;
> - prepare for progression to work and job-related NVQs;
> - keep students labelled by defenders of A-levels and GCSEs as 'less-able' from 'undermining' standards in these qualifications;
> - convince learners, teachers, admissions tutors that GNVQs had parity of esteem with A-levels and GCSEs;
> - ameliorate poor levels of achievement in numeracy and literacy;
> - unify disparate and confusing post-16 qualification pathways;
> - satisfy demands from different constituencies to include 'essential' content and skills.
>
> (Ecclestone, 2002, page 81)

She characterised these as "huge, if not impossible" demands (page 81). A decade later, Isaacs raised similar concerns over the Diploma qualification, which was associated with a very similar range of purposes. She focused particularly on ambiguity related to the target cohort:

> Because the issue of who the diploma was for was never satisfactorily resolved, it could not lead to easily understood progression pathways. Was it for the less able? For those who were at risk of disengaging with the education system? For those who wanted to go into employment directly after secondary education? For high-flyers applying to selective universities? It could not serve all of these groups equally well, although policymakers blindly contended that it would.
>
> (Isaacs, 2013, page 284)

Qualification designers need to be aspirational thinkers, but not wishful thinkers. Purpose analysis involves identifying, and then testing, purpose-related aspirations, to ensure that they are realistic – and revising them accordingly if not.

Integration

> Examinations are used for many purposes, and many of the difficulties and evils connected with their use arise from the attempt of those who frame the Examinations, either to effect too many things by one instrument; as, for instance, to pick out the most proficient, to reject the dunces, and to give an impetus to a certain kind of study all by the same series of papers; or from their not having a clear view of what they are aiming at.
>
> (Latham, 1877, pages 2 to 3)

The idea that qualifications serve multiple purposes is not new, and nor are allegations that policymakers have paid insufficient attention to them. Hence, this quotation from 150 years ago. So, why do we still fall foul of this criticism?

Part of the problem is that purpose analysis has not been theorised adequately. Historically, curriculum purposes have tended to be theorised by education scholars, while assessment purposes have tended to be theorised by measurement scholars. Similarly, education practitioners have traditionally assumed responsibility for teasing out the implications of curriculum purposes for qualification design, while measurement practitioners have traditionally assumed responsibility for teasing out the implications of assessment purposes for qualification design. Unfortunately, insufficient attention has been paid to the integration of these perspectives, in theory as well as in practice.

This report has attempted to explain the necessity and centrality of purpose analysis as a mechanism for both addressing and resolving potential conflict through an **integrative analysis** of alternative perspectives on qualification purposes and of the purposes and sub-purposes subsumed within each. To build solid foundations for qualifications and qualification systems, we need to pay more than lip service to the idea of fitness for purposes. This is why purpose analysis needs to be understood and practiced as an activity in its own right, which precedes all other design-related activities. The objective in writing this report was therefore to help systematise the considerations that need to be taken into account whenever a new qualification or qualification system is being designed, thereby promoting an integrated account of purposeful design.

But why is an integrative analysis so important? Is there not a natural order to qualification design, which means that curriculum considerations come first and assessment considerations come second, both in theory and in practice? In other words, is it not a “cardinal principle” that assessment “should follow the curriculum and not determine it” (Spens, 1938, p.80)? And, if so, then why not assume a simple division of labour in which curriculum practitioners begin by making curriculum decisions – including the specification of syllabus outcomes – after which they hand

over to assessment practitioners who are then responsible for designing assessment arrangements that simply mirror those curriculum decisions?

Unfortunately, we know from over a century of running national qualification systems that this is not how it works. In fact, that is exactly the point that the Spens report was making:

> In the actual working of the examination this principle has been reversed, and there can be little doubt that in many Secondary Schools the Certificate Examination is now the dominant factor in determining the curriculum for the majority of the pupils below the age of 16.
>
> (Spens, 1938, page 80)

Decisions made when designing an assessment procedure always end up departing to some extent, in various ways, from curriculum aspirations. There is no such thing as a perfect assessment procedure, which assesses its target proficiency entirely comprehensively and entirely authentically. Compromises must be made for all sorts of very good reasons, and they will often have consequences for teaching and learning. These compromises need to be part of conversations that happen between curriculum and assessment practitioners (and others too) right at the outset of the qualification design process. These conversations might, for instance, lead to a significantly different specification of syllabus outcomes, even when this might seem less than ideal from a curriculum perspective.

The point is a more general one, which is that qualification design needs to be undertaken from multiple perspectives simultaneously. Hence, the need to commence qualification design with an extended analysis of purposes, to engage those multiple perspectives in conversation with each other. Approaches to designing curriculum, pedagogy, and assessment are not related to each other hierarchically, as though one naturally trumps the others in terms of importance and therefore has the final say. Effective qualification design treats recommendations from all 3 perspectives seriously, carefully weighing them against each other.

Stages

As explained earlier, purpose analysis is the first stage within a 3-stage framework for qualification design:

1. purpose analysis
2. syllabus outcomes and qualification structure
3. assessment procedure

Purpose analysis lays the groundwork for subsequent decision-making, with a particular focus on ensuring the resultant profile of prioritised purposes is realistic, that is, coherent and viable. It involves:

- **eliciting** purpose-related aspirations for the new qualification
- **anticipating** their likely design implications (in terms of syllabus outcomes, qualification structure, and assessment procedure) and their likely consequences
- **testing** the logical coherence and practical viability of these design implications (in terms of their mutual compatibility, suitability for learners from the target cohort and their circumstances, compatibility with broader systemic requirements, and their practical viability in relation to wider constraints)
- **debating** the pros and cons of alternative approaches to resolving conflicts that have been identified
- **agreeing** a profile of prioritised purposes for the new qualification that appears, in principle, to be achievable and defensible

Purpose analysis therefore invites us to consider the plausibility of our purpose-related aspirations – explicitly in terms of their likely design implications – before progressing to the stages in which formal and final design decisions are made. It will not be possible to anticipate and interrogate each and every potential design decision, but the more thorough the analysis that occurs at this stage, the smoother the progress that will be made during subsequent stages, and the lower the risk of having to return to the drawing board a long way into the design process. Note that, once the development process begins in earnest, it may no longer be feasible to return to the drawing board, so discovering a fatal flaw at this stage will transform qualification development into an extended, but inevitable, train crash.

Rather than a principled hierarchy, there is a pragmatic order to the 3 stages: purposes first, then syllabus and structure, and finally assessment. However, as comprehensive as purpose analysis aims to be, complications inevitably arise during subsequent stages, as design implications are drawn out in far greater depth and detail. When this happens, it may be necessary to revisit debates undertaken during the purpose analysis stage, and it may even be necessary to revise the profile of prioritised purposes itself.

Idealised

We described this as an ‘idealised’ framework for various reasons. One relates to the historical lack of an overarching integrative theoretical framework for qualification design, which has left qualification designers somewhat rudderless over the years. We hope that the 3-stage framework might help to systematise qualification design, going forward. But, for now, the framework is simply a tool for thinking about

effective qualification design, and we have limited evidence of how best to put it into practice.

Another reason relates to the potential size and challenge of the task of qualification and qualification system design. Conceptually, it is helpful to break it down into its component elements, consideration by consideration, as attempted in the previous main section. Yet purpose analysis is unlikely to be structured quite so systematically in practice.

Perhaps the hardest practical challenge concerns how best to co-ordinate contributions from a multitude of potential participants (see below). But there is also a more abstract challenge, which concerns the principle that qualification design decisions should never be arbitrary, nor unthinkingly default to longstanding traditions of custom and practice, but must always be tailored consciously to a bespoke, prioritised profile of intended purposes. While this is certainly the ideal, it also seems likely that debating every single, conceivable, design decision from first principles would be unrealistic. Consequently, some design decisions are likely to receive more attention than others. Furthermore, certain lower-level design decisions are likely to be taken implicitly when higher-level design decisions are taken, and decisions concerning one aspect of a qualification might logically entail decisions related to other aspects. So, while the principle of reflective design is a critical one, the processes by which decisions actually get taken may not always be as linear or co-ordinated as we might anticipate.

Participants

Because qualification design needs to address multiple perspectives on qualification purposes simultaneously, it matters a lot who contributes to this process and how their contributions are co-ordinated. There are many questions to answer concerning the participants of purpose analysis:

- who ought to be responsible for it?
- who ought to coordinate and manage it?
- who ought to contribute to it?

Of course, this begs the question of exactly what the ‘it’ is, beyond a process that involves eliciting, anticipating, testing, debating, and agreeing. In other words, is this a discrete (albeit extended) activity that involves a core group of planners, who may draw upon individuals with different kinds of expertise at different points in time? Or is it an assembly of activities, each with its own core group, perhaps with overlapping memberships, or perhaps coming together at different stages? Or is it something completely different, and perhaps far more nebulous? It is not at all obvious how best to implement purpose analysis, given the multitude of voices that need somehow to

be represented and orchestrated. It is, however, fairly easy to identify the sorts of voices that would benefit from being heard, which include:

- those with ultimate responsibility for decisions related to different aspects of qualification design (for example, government, regulators, awarding bodies)
- specialists with detailed conceptual and technical understanding (for example, qualification design experts, curriculum and domain experts, assessment experts, pedagogy experts, logistics experts, legal experts)
- practitioners who will be responsible for qualification delivery (for example, teachers, trainers, assessors, relevant associations)
- practitioners who will assume responsibility for certificate use (for example, employers, higher education, relevant associations)
- members of the public (especially students and their parents)

None of these voices is any more nor less important than any other to the success of purpose analysis, as they all represent different pieces of the qualification design jigsaw puzzle, providing different insights into the likely success or failure of the emerging profile. But how, exactly, their perspectives ought to contribute to the activity, or activities, of purpose analysis is harder to say.[51]

Where it is not feasible for certain voices to contribute directly to ongoing debates, it will be important to ensure that their perspectives are not entirely overlooked. This reminds us that the activity, or activities, or purpose analysis need to be co-ordinated. Ideally, they will be orchestrated by qualification design professionals, with technical expertise in purpose analysis.

Anticipatory

In reports 4 and 9 of our recent research into outcome-based qualification design, we argued that qualification reforms are best understood as education and training reforms that are initiated through changes to certification requirements (Newton, Clarke, et al, 2024; Newton, Curcin, et al, 2024a). If so, then there needs to be a clear line of sight from the point at which qualification design is initiated through to each stage of its implementation. We referred to this as **anticipatory qualification design** because it involves anticipating all the arrangements that need to be put in place: for the qualification to succeed (anticipating likely requirements for effective rollout and ensuring that they are fulfilled), and for the qualification not to fail

[51] One of the ways through which Ofqual integrates insights from multiple stakeholders is through formal consultation exercises. Proposals for new regulations become the subject of formal consultations, and proposals are amended accordingly before final decisions are made. This includes (changes to) the purposes served by different types of qualification.

(anticipating likely threats to effective rollout and mitigating them). This is fundamental to all 3 stages of the idealised design framework, including purpose analysis, which involves anticipating likely design implications and their likely consequences, and testing them for logical coherence and practical viability.

Part of the importance of reflecting on likely design implications relates to an observation that those advocating for qualification reform often come armed, not simply with purpose-related aspirations, but also with implicit or explicit assumptions about how those aspirations can best be achieved, through particular design decisions. A good example of this relates to the aspiration for increasing the breadth of post-16 study programmes, which has been voiced repeatedly over the decades (in general, vocational, and technical contexts). Perhaps the most commonly voiced aspiration is for students to continue studying English and maths, alongside other subjects that they choose to specialise in. This is often packaged as a proposal for a diploma-based award structure, which rewards students for successfully completing the full programme, not just their specialist subject-based qualifications. The implicit assumption underpinning a proposal of this sort is that the diploma structure itself will motivate students to study the full programme to its completion. Yet it is exactly this kind of assumption that needs to be exposed, and tested, before it is implicitly or explicitly committed to.

In this example, the implicit theory of action would seem to assume that students will commit to studying subjects beyond their specialist ones because they know that they will not be awarded the diploma (at all) if they fail to achieve any one of its components (for example, if they fail to achieve a level 2 or 3 award in English). This logic would hold water if the overarching diploma certificate came to possess significant market value in its own right. However, if employers or higher education selectors saw no value in the overarching diploma certificate – and therefore did not ask for evidence of its award, continuing to rely solely on the separately certificated subject-based awards for hiring and admission purposes – then the incentive to complete any non-specialist components would be missing. And students would continue to focus their attention on the specialist components, at the expense of the broader non-specialist ones. If, instead, the specialist subject-based qualifications were no longer certificated – that is, if only the overarching diploma were to be certificated – then this might have the potential to incentivise students to complete all the diploma requirements. Yet if employers and HE admissions teams found that they now lacked the kind of information that they had come to rely upon, then the new qualification system would most likely fail for different reasons, related to public confidence.

Examples like this explain why qualification design needs to be proactively forward-thinking, even at the earliest stage of purpose analysis. In this example, detailed analyses of the sort just outlined might result in a decision to down-weight the aspiration for broader post-16 study programmes. Or it might result in a decision to

pause the reform until it had been possible to persuade all key stakeholders – especially qualification result users – of the importance of broader post-16 study programmes, to ensure that they all signed-up to the role that they would need to play in the new order before any design decisions were committed to.

The deeper point is that purpose analysis needs to focus on how the new qualification or qualification system is actually going to work in the real world, and not on how policymakers would like it to work in an ideal world. Designers need to be aspirational thinkers, but not wishful thinkers. In particular, policymakers cannot determine how learners will engage with the qualifications that they make available, and nor can they determine how their results will be used. Our systems operate by consent, which means that the trade-offs that need to be made during qualification design need to be understood as genuine compromises.

Prioritisation

When designing a new qualification, the ideal would be for:

- all students to be very highly motivated to take, study, and complete it, and for all teachers to be enabled to adopt the very best strategies for teaching it
- all students to acquire learning outcomes with very high intrinsic value for their future life and work and very high market value for their future employers
- all awarded certificates to possess a very high level of accuracy, and therefore to for all certificate users to be enabled to make the very best decisions

In practice, it is never possible to achieve all these aspirations to the extent that we would ideally like them to be achieved. Trade-offs and compromises are the norm during qualification design, which is also why purpose analysis needs to be serious about prioritising aspirations.

This is a hard message to sell, because it is hard to stomach the idea of setting out to design a qualification that will be less than perfect, especially given how important qualifications are to individuals, organisations, the economy, and society. Yet by not confronting this reality up-front – and persisting with a have-cake-and-eat-it-too mentality – the challenge is simply kicked into the long grass. Trade-offs and compromises will still have to be made, but they will end up being forced rather than planned, with potentially disastrous consequences. When decisions are forced in this manner, they are likely to be made inconsistently, and with a lack of transparency, neither of which are good for instilling public confidence.

Figure 12. Progression qualification purpose profiles (level 2 versus entry level)

Level 2 Technical Progression Qualification	Entry Level Progression Qualification
Purpose A: Demonstrating that Learners have the breadth of knowledge, skills, and understanding that prepares them to take a level 3 technical qualification that is not delivered as part of a T Level programme.	Purpose 1: The qualification will support an engaging course of learning that motivates and (re-) engages Learners, who are considering progressing to level 1 study.
Purpose B: Demonstrating Learners' ability to apply their knowledge, skills and understanding in relevant practical contexts.	Purpose 2: The qualification will provide Learners with a breadth of knowledge, skills and understanding that prepares them for relevant level 1 qualifications.
Purpose C: Supporting an engaging programme of learning that enables and motivates Learners who are preparing for progression to level 3 technical study.	Purpose 3: The qualification will demonstrate Learners' ability to apply knowledge, skills and understanding in relevant practical contexts.
Purpose D: Providing evidence of attainment that can be used in combination with other information to inform decisions about a Learner's readiness to progress onto higher levels of study in the relevant occupational pathway.	Purpose 4: The qualification will provide evidence of attainment that could be used in combination with other information to inform decisions about a Learner's readiness to progress on to level 1 pre-technical study.
Purpose E: Providing evidence to differentiate between Learners' attainment in relation to the knowledge, skills and understanding assessed as part of the qualification.	

Note that expertise-related purposes are presented in red, engagement-related purposes are presented in green, and information-related purposes are presented in blue.

The product of purpose analysis is an achievable and defensible profile of prioritised purposes, which provides a single point of reference for all subsequent qualification design and development activity. The profile is likely to be presented at a high level of abstraction – as a very detailed and nuanced profile would probably be very hard to apply – but it should still have immediate practical value in guiding and arbitrating qualification design decisions.

At Ofqual, we are still developing our approach to prioritising qualification purposes for new qualifications – typically, for new qualification types, for which we write Qualification Level Conditions. Figure 12 contrasts 2 prioritised purpose profiles that were prepared to support decisions on Regulating Level 2 and Below Progression Qualifications, in response to the DfE Review of Post-16 Qualifications at Level 2 and Below in England. They are taken from Ofqual's Qualification Level Conditions for Progression Qualifications, which included purpose profiles for

- level 2 T Level Foundation Qualifications (supporting progression to T Levels)
- level 2 Technical Progression Qualifications (supporting progression to level 3 Technical Qualifications that are not delivered as part of a T Level programme)
- level 2 Academic Progression Qualifications (supporting progression to level 3 Alternative Academic Qualifications)
- level 1 Progression Qualifications (supporting progression to higher level technical qualifications)
- entry level Progression Qualifications (supporting progression to level 1 pre-technical qualifications)

It was anticipated that these qualifications would be taken primarily by young people as part of a one-year, post-16, study programme, alongside other components including English and maths. These students would not have achieved the necessary level of attainment, pre-16, to progress directly to level 3 study, or (alternatively) would be presumed to benefit from extra development or support before taking a level 3 qualification. Students enrolled for the lowest-level qualifications would be among those with the lowest levels of attainment, pre-16, a significant subset of whom would be students with special educational needs and disabilities (SEND).

There are many noteworthy points from Figure 12. First, neither profile presents a comprehensive list of all plausible and defensible purposes. The purposes are written at a very high level, specifically to foreground alternative purpose perspectives: expertise (red), engagement (green), and information (blue).

Second, where it seemed important to draw distinctions within a purpose perspective, separate statements were drafted. For instance, Purpose A states that the kind of expertise that students should acquire (at level 2) is that which would

adequately prepare them for a corresponding level 3 qualification. Purpose B supplements this by stating that the target proficiency for the qualification should emphasise practical application.

Third, the purpose perspectives are prioritised. At level 2, the expertise perspective is prioritised above the engagement perspective, which is prioritised above the information perspective. It is significant that the information perspective comes last in this list. For many qualifications, especially academic qualifications at level 3, the information perspective would come first, where high-stakes decisions are made based on these results and we need to be especially confident of the accuracy and consistency of those results. For instance, when we published purposes for level 3 Alternative Academic Qualifications (AAQs), the highest-priority purpose corresponded to the information perspective:

> (a) Providing Users of the qualification with reliable evidence to differentiate between Learners' attainment in relation to the knowledge, skills and understanding assessed as part of the qualification and, in particular, supporting decisions regarding the selection of Learners for further study (General Purpose A).
>
> (Level 3 Alternative Academic Qualifications and Technical Occupation Qualifications Qualification Level Conditions)

De-prioritising the information perspective for progression qualifications does not mean that result accuracy is unimportant.[52] Far from it. It simply means that, if we had to take a hit on any of the 3 perspectives, we would be more prepared to take a hit on the information perspective (result accuracy) than on the other 2. Prioritisation therefore involves being explicit about what we are least prepared to compromise on, when qualification design decisions are made. Prioritising the engagement perspective over the information perspective sends a message to qualification designers that the target cohort cannot be assumed to be motivated, and that it would be legitimate to make design decisions with engagement in mind (even if this meant taking a bit of a hit on result accuracy, or a bit of a hit on its market value).

Fourth, the highest-priority purpose for Entry Level Progression Qualifications reflects the engagement perspective. It is uncommon to weight the engagement perspective above the others, and this tends to happen only for cohorts with many disengaged or fragile learners, for whom completing the qualification is likely to be an accomplishment in its own right, even if the qualification certificate ends up

[52] It is hard to imagine zero-weighting any of the perspectives. The precise content of a qualification will always matter to some extent (expertise), and it would be perverse to award certificates with complete ambivalence toward result accuracy (information). Likewise, if there were nothing motivating about the award of a certificate (engagement), then we would presumably have to admit that the qualification was worthless.

having limited extrinsic market value. In this situation, the aspiration is more about building (or re-building) a learner than about building a specific body of learning.

One final observation concerns a purpose that was not formally recognised for these lower-level post-16 qualifications, but that frequently does need to be recognised, which is the use of qualification results for holding schools to account. Again, this was included in the list that was recently specified for AAQs:

> (d) Providing a basis for schools and colleges to be held accountable for the performance of their Learners (General Purpose D).
>
> (Level 3 Alternative Academic Qualifications and Technical Occupation Qualifications Qualification Level Conditions)

Although this is an information-related purpose, it focuses on information generated from aggregated results (to appraise teaching), not results from individual learners (to measure learning). This was the fourth-ranked purpose in the AAQ list and, with a secondary focus on teaching rather than a primary focus on learning, there are good reasons for locating purposes of this sort toward the bottom of any priority list.[53]

The approach illustrated in Figure 12 is the one that Ofqual has been using, and developing, since the late-2010s. It represents an advance on historical approaches in terms of distinguishing explicitly between alternative purpose perspectives, and in terms of explicitly prioritising them. We would not claim it is necessarily the best possible approach, however, and we will continue to hone our practices.

Indeed, we would have to admit our approach is not yet ideal, in the sense that critical design decisions sometimes get finalised before Ofqual's purpose analysis formally begins. To some extent this is attributable to the historical separation of responsibilities for qualification design in England, which was mentioned earlier. This means expertise-related decisions (related to syllabus outcomes) are sometimes finalised prior to, and therefore independently of, a thorough consideration of implications arising from the information and engagement perspectives.

Compromise

Over the past century or so, the state has assumed much greater control over critical design decisions related to pre-university qualifications – both general and vocational ones – particularly those funded from the public purse. Nowadays, the Department

[53] Having said that, when results are used for an accountability purpose, and teachers play an active role in the assessment process, this can create a perverse incentive to engage in malpractice, which presents a risk to the ethical integrity of the system (one of the principal systemic requirements discussed earlier). So, considerations of this sort might influence qualification design decisions howsoever the accountability purpose had been prioritised.

for Education specifies core content for many school- and college-based qualifications, while Ofqual specifies rules governing many of the remaining high-level design decisions. If so, then does purpose analysis really need to emphasise trade-off and compromise, or does the central specification of qualifications policy nowadays obviate this need?

There are 2 responses to this question. First, even if decisions of this sort were made by a single policymaker, they would still need to grapple with challenges that required trade-offs and compromises. Even a single policymaker will have multiple aspirations for the new qualification or qualification system, related to a variety of different purposes and purpose perspectives, and the implications of these aspirations would still need to be explicated and interrogated in order to determine their coherence and viability. This would uncover tensions, perhaps sometimes straightforward contradictions, that would then need to be resolved. Ultimately, there is no such thing as a perfect qualification: trade-offs and compromises are inevitable when designing any qualification or qualification system.

Second, even within a highly autocratic system it would still need to be recognised that qualifications are community-mandated assets that operate by consent. As noted earlier:

- students can withdraw their consent by choosing not to study certain qualifications, or by choosing not to follow particular qualification pathways, or simply by failing to put any effort into their studies
- admission officers and employers can withdraw their consent by choosing not to use results from particular qualifications for recruiting or selecting purposes
- parents can withdraw their consent via elective home schooling or by allowing their children to absent themselves from school for extended periods
- the public can withdraw its consent by expressing uproar over qualifications or qualification systems that seem not to be defensible for one reason or another

In situations like these, design decisions might have been made with a specific profile of purposes in mind, but if they ended up exasperating participants or stakeholders, then these purposes could easily end up being frustrated or undermined. This links back to the idea of systemic requirements for the success of a qualification system, which include public understanding, public trust, public confidence, and so on. Even in the most autocratic of qualification policy contexts, policymakers would still need to bring participants and stakeholders on board with the qualification reform agenda. And that inevitably requires a certain amount of trade-off and compromise on behalf of all interested parties.

Scope

Before concluding, it is worth mentioning one final point concerning the scope or ambition of purpose analysis, as outlined above. The guiding principle is that aspirations for a new qualification or qualification system need to be both logically coherent and practically viable. This viability has been discussed in relation to how the world actually is, not how a policymaker might like it to be, hence the idea that qualification designers need to be aspirational but not wishful thinkers. The risk in framing purpose analysis like this is that it might seem to rule out the possibility of more radical reform, which envisages a new kind of qualification for a new kind of education system for a changed world. In other words, this framing of qualification design and fitness for purposes might seem inherently conservative (with a small c).

The idea of accountability as a backdrop to education in England is helpful for exploring this possibility. Schools and colleges in England are currently held to account for the effectiveness of their provision based on aggregated qualification results. Using exam results to judge the effectiveness of teachers dates to the earliest of exams, but the current approach (which involves publishing school and college performance tables) has taken root over the past 3 or 4 decades. So, accountability systems of this sort are deeply entrenched, and they are an important part of the conversation when qualifications are being designed.

For instance, accountability is part of the conversation when considering the information perspective, given that results are used for assessment purposes like institution monitoring, resource allocation, and organisational intervention. But it is also part of the conversation when gauging the likely ethical integrity of the system, particularly its robustness, and especially in relation to the use of alternative assessment formats. The decision concerning whether to incorporate teacher assessment judgements provides an important case in point. Results-based accountability systems create perverse incentives for teachers to inflate students' grades – even if just by a little, as though giving them the benefit of the doubt – so qualifications may need to be designed to be resistant to grade inflation, or straightforward subversion, of this sort. Concerns like this are sometimes grave enough to rule out using teacher assessment judgements entirely.

But what if we envisaged a radical reform of the education system, such that the new system was no longer judged in terms of performance tables built from aggregated qualification results? Assuming that radical qualification reform might then also be on the agenda, what might be the implications for purpose analysis in this situation?

Even if we were attempting to design a radically new kind of qualification for a radically new kind of education system, purpose analysis (as outlined above) would still be the place to start. This would include asking important questions concerning the incoming accountability system and its mechanisms, like:

- how would the new approach to accountability operate?
- what would be the (anticipatable) consequences of its operation?
- how soon could we expect it to be operating like that (with those consequences)?
- how certain could we be of our answers to these questions?

Assuming that we could be confident in our predictions, they could straightforwardly be incorporated within the purpose analysis process. Having said that, with its insistence that designers should be aspirational but not wishful thinkers, there is still a sense in which (small c) conservatism is built into the logic of purpose analysis, linked to the idea that our qualification systems are too important to fail through suboptimal planning. In other words, if radical reform were to be on the agenda, then so too would be painstaking planning, as the stakes associated with inadequate purpose analysis would be extremely high.

Conclusion

Although qualifications play a very important role within education and training systems all over the world, qualification design practices remain under-theorised. A major threat arises from the lack of integration between different perspectives on qualification design, including a tendency for curriculum, pedagogy, and assessment planning to be theorised and practiced in silos, with few co-ordinating structures. This report proposes a broad, integrated definition of 'qualification' as follows:

> A qualification is a procedure for certifying attainment in an education or training context, which includes specifications for syllabus outcomes, qualification structure, and an assessment procedure. These specifications have direct intended implications for what students ought to learn, aspects of teaching and learning processes, and how to interpret qualification results.

The implication of this broad definition is that there are multiple perspectives on the purposes that underpin qualification design, including these 3 critical ones:

- expertise perspective – concerning the body of learning that needs to be acquired for the award of a certificate
- engagement perspective – concerning the engagement that needs to be exercised (by learners and by teachers) to achieve the required standard
- information perspective – concerning the information that needs to be provided by a qualification certificate

The broad definition reminds us that a qualification is not just an assessment. If it were just an assessment, then we would expect measurement professionals to dominate the design process, applying principles that follow from the information perspective. A qualification is not just an assessment, however, and other principles must come into play when designing a qualification. Sometimes recommendations from the expertise or engagement perspective trump recommendations from the information perspective. Other considerations also come into play, related to:

- the composition of the target cohort for the new qualification, and the anticipated circumstances of their teaching, learning, and assessment
- broader systemic requirements related to the structural, cognitive, and ethical integrity of the qualification and the system within which it is located
- pragmatic constraints related to costs and resource availability, and constraints related to wider political, legal, and commercial concerns

We have introduced the term 'purpose analysis' to describe the process of reflecting on these drivers and constraints and their implications for qualification design, which is undertaken before finalising any design decisions. Purpose analysis involves:

- eliciting purpose-related aspirations for the new qualification
- anticipating their likely design implications (in terms of syllabus outcomes, qualification structure, and assessment procedure) and their likely consequences
- testing the logical coherence and practical viability of these design implications (in terms of their mutual compatibility, suitability for learners from the target cohort and their circumstances, compatibility with broader systemic requirements, and their practical viability in relation to wider constraints)
- debating the pros and cons of alternative approaches to resolving conflicts that have been identified
- agreeing a profile of prioritised purposes for the new qualification that appears, in principle, to be achievable and defensible

Purpose analysis invites us to consider the plausibility of purpose-related aspirations before progressing to stages in which formal and final design decisions are taken. This is central to qualification design because it lays the groundwork for all subsequent decision-making. Purpose analysis focuses specifically on ensuring the resultant profile of prioritised purposes is realistic, that is, coherent and viable. It is therefore the key to ensuring that qualifications and qualification systems are fit for the purposes we wish them to serve.

Finally, it is important to remember there is no such thing as a perfect qualification or qualification system: trade-offs and compromises are inevitable in this business. The decision-making that purpose analysis facilitates is therefore both technically complex and politically sensitive. Needless to say, it is not a process that should be rushed.

References

Adamson, B. & Morris, P. (2007). Comparing curricula. In M. Bray, B. Adamson, & M. Mason (Eds.). *Comparative Education Research Approaches and Methods. CERC Studies in Comparative Education 19* (pp.263–282). Hong Kong: University of Hong Kong with Springer.

ASCL (Association of School and College Leaders) (2019). *The Forgotten Third: Final report of the Commission of Enquiry*. Leicester: Association of School and College Leaders.

Baird, J., Caro, D., Elliott, V., El Masri, Y., Ingram, J., Isaacs, T., Pinot de Moira, A., Randhawa, A., Stobart, G., Meadows, M., Morin, C., & Taylor, R. (2019). *Exam Reform: Impact of linear and modular exams at GCSE* (OUCEA/19/2. Ofqual/19/6506/1). Coventry: Office of Qualifications and Exams Regulation.

Bardell, G.S., Forrest, G.M. & Shoesmith, D.J. (1978). *Comparability in GCE: A review of the board's studies, 1964–1977*. Manchester: Joint Matriculation Board on behalf of the GCE Examining Boards.

Bew, P. (2011). *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final report*. London: Department for Education.

Biesta, G.J.J (2010). *Good Education in an Age of Measurement: Ethics, politics, democracy*. Oxford: Routledge.

Bloom, B.S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1–12.

Bramley, T. & Crisp, V. (2019). Spoilt for choice? Issues around the use and comparability of optional exam questions. *Assessment in Education: Principles, Policy & Practice*, 26(1), 75–90.

Cairney, P. (2020). *Understanding Public Policy: Theories and issues* (second edition). India: Bloomsbury Academic India.

Cambridge Assessment (2018). *The Cambridge Approach to Textbooks*. Cambridge: Cambridge Assessment.

Cedefop (2009). *The Shift to Learning Outcomes: Policies and practices in Europe*. Luxembourg: Office for Official Publications of the European Communities.

Cedefop (2010). *Changing Qualifications: A review of qualifications policies and practices. Cedefop Reference series, 84*. Luxembourg: Publications Office of the European Union.

Cedefop (2015). *Vocational Pedagogies and Benefits for Learners: Practices and challenges in Europe*. Luxembourg: Publications Office of the European Union.

Cedefop (2025). *The Influence of Learning Outcomes-Based Curricula on Teaching Practices*. Luxembourg: Publications Office of the European Union.

Colquitt, J.A. & Rodell, J.B. (2015). Measuring justice and fairness. In R.S. Cropanzano & M.L. Ambrose (Eds.). *The Oxford Handbook of Justice in the Workplace* (pp.187–202). New York, NY: Oxford University Press.

Cresswell, M.J. (2003). *Heaps, Prototypes and Ethics: The consequences of using judgements of student performance to set exam standards in a time of change*. London: University of London Institute of Education.

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.

Cuff, B.M.P., Zanini, N., & Black, B. (2018). *An Exploration of Grade Inflation in 'Older Style' Level 3 BTEC Nationals: 2006 to 2016 (Ofqual/18/6459/2)*. Coventry: Office of Qualifications and Examinations Regulation.

Curtis, P. (2003). £100m to modernise 'cottage industry' exam system. The Guardian online. 17 November.

De Ville (1986). *Review of Vocational Qualifications in England and Wales*. London: Her Majesty's Stationery Office.

DfE (Department for Education) (2025). *Pupil Experiences in School: Academic year 2024 to 2025. Findings from Department for Education's omnibus surveys with pupils and parents.* London: Department for Education.

DfES (Department for Education and Skills) (2003). *A New Specialist System: Transforming secondary education*. Nottinghamshire: Department for Education and Skills.

Ecclestone, K. (2002). *Learning Autonomy in Post–16 Education: The politics and practice of formative assessment*. Oxford: RoutledgeFalmer.

Ecclestone, K. (2004). Learning in a comfort zone: Cultural and social capital inside an outcome-based assessment regime. *Assessment in Education: Principles, Policy & Practice*, 11(1), 29–47.

Ecclestone, K. (2007). Commitment, compliance and comfort zones: The effects of formative assessment on vocational education students' learning careers. *Assessment in Education: Principles, Policy & Practice*, 14(3), 315–333.

Ecclestone, K. (2012). Emotional well-being in education policy and practice: The need for interdisciplinary perspectives and a sociological imagination. *Research Papers in Education*, 27(4), 383–387.

Ecclestone, K., with Davies, J., Derrick, J. & Gawn, J. (2010). *Transformative Formative Assessment in Lifelong Learning*. Berkshire: Open University Press.

Edwards, T. (1997). Educating leaders and training followers. In T. Edwards, C. Taylor Fitz-Gibbon, F. Hardman, R. Haywood, & N. Meagher (Eds.). *Separate But Equal? A levels and GNVQs* (pp.8–28). London: Routledge.

Eraut, M. (2001). The role and use of vocational qualifications. *National Institute Economic Review*, 178, 88–98.

EU (European Union) (2012). Council Recommendation of 20 December 2012 on the validation of non-formal and informal learning. *Official Journal of the European Union*, C 398/01.

FAB (Federation of Awarding Bodies) (2017). *The Qualification Manager's Handbook*. London: Federation of Awarding Bodies.

Flyvbjerg, B. & Gardner, D. (2023). *How Big Things Get Done: The surprising factors behind every successful project, from home restoration to space exploration*. London: Macmillan.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39(3), 193–202.

Freedman, S. (2022). *The Exam Question: Changing the model of assessment reform*. London: Institute for Government.

Goodson, I.F. (1995). *The Making of Curriculum: Collected essays (2nd edition)*. London: The Falmer Press.

Gray, R., & Lockhart-Smith, A. (2015). *Personal and social development provision: Content, format and impact*. Leicester, UK: National Institute of Adult Continuing Education (England and Wales).

Greening, J. (2017). Letter to Neil Carmichael, Chair of the Education Select Committee. 28 March.

Gronlund, N.E. & Brookhart, S.M. (2009). *Gronlund's Writing Instructional Objectives (8th edition)*. Upper Saddle River, New Jersey: Pearson Education.

Hargreaves, A. (1989). *Curriculum and Assessment Reform*. Buckinghamshire: Open University Press.

HCCSFC (House of Commons Children, Schools and Families Committee) (2008a). *Testing and Assessment: Third Report of Session 2007–08. Volume I. HC 169-I*. London: The Stationery Office Limited.

HCCSFC (House of Commons Children, Schools and Families Committee) (2008b). *Testing and Assessment: Government and Ofsted Responses to the Committee's Third Report of Session 2007–08. HC 1003*. London: The Stationery Office Limited.

HM Government (2013). *The Future of Apprenticeships in England: Implementation plan*. London: Department for Business, Innovation and Skills.

HM Government (2025). *The UK Government Resilience Action Plan: The UK's strategic approach to resilience*. London: Cabinet Office.

Hodgson, A. & Spours, K. (2019). Further Education in England: At the crossroads between a national, competitive sector and a locally collaborative system? *Journal of Education and Work*, 32(3), 224–237.

Holmes, S., Khan, A., Zanini, N. & Black, B. (2020). *Predicting Predictability: Investigating question paper predictability and the factors that influence this through a question prediction exercise (Ofqual/20/6714/1)*. Coventry: Office of Qualifications and Examinations Regulation.

Huddleston, P. (2020). *A Short History of Employer Engagement: Once more round the buoy or set fair for a better voyage?* London: Education and Employers.

Hulleman, C.S., Totonchi, D.A., Barron, K.E., Sutter, C.C., Hulleman, T., Tibbetts, Y., & Thoman, D. (2025). From student to system change: The evolution of the utility-value intervention. *Advances in Motivation Science*, 12, 165–227.

Hyland, T. (2005). Learning and Therapy: Oppositional or complementary processes? *Adults Learning*, 16(5), 16–17.

Independent Teacher Workload Review Group (2016). *Eliminating Unnecessary Workload Around Marking*. London: Department for Education.

Isaacs, T. (2013). The Diploma qualification in England: An avoidable failure? *Journal of Vocational Education & Training*, 65(2), 277–290.

Jonson, J.L. & Geisinger, K.F. (2022) (Eds.). *Fairness in Educational and Psychological Testing: Examining theoretical, research, practice, and policy implications of the 2014 Standards*. Washington, DC: American Educational Research Association.

Keep, E. & James, S. (2012). A Bermuda triangle of policy? 'Bad jobs', skills policy and incentives to learn at the bottom end of the labour market. *Journal of Education Policy*, 27(2), 211–230.

Keep, E. (2009). *Internal and External Incentives to Engage in Education and Training – a Framework for Analysing the Forces Acting on Individuals? Monograph No. 12*.Cardiff: ESRC Centre on Skills, Knowledge and Organisational Performance, Cardiff University.

Keep, E. (2012). *Reforming Vocational Qualifications: Some Problems Ahead? Issues Paper 28*. Cardiff: ESRC Centre on Skills, Knowledge and Organisational Performance, Cardiff University.

Kingdon, M. & Stobart, G. (1988). *GCSE Examined*. Sussex: The Falmer Press.

Latham, H. (1877). *On the Action of Examinations Considered as a Means of Selection*. Cambridge: Deighton, Bell and Co.

Like, C.B. (1986). The BTEC HND in physics – five years on. *Physics Bulletin*, 37(5), 196–197.

Lucas, B., Spencer, E., & Claxton, G. (2012). *How to Teach Vocational Education: A theory of vocational pedagogy*. London: City & Guilds Centre for Skills Development.

Mac Aog´ain, E., Millar, D., & Kellaghan, T (2010). The bonus for Irish in the leaving certificate examination. *Irish Journal of Education*, 38, 25–42.

Mager, R.F. (1997). *Preparing Instructional Objectives: A critical tool in the development of effective instruction (3rd edition)*. Atlanta, GA: Center for Effective Performance.

Maxwell, G.S. (2010). Moderation of Student Work by Teachers. In P. Peterson, E. Baker, & B. McGaw (Eds.). *International Encyclopedia of Education (3rd edition)* (pp.457–463). Oxford: Elsevier.

Meadows, M. & Black, B. (2018). Teachers' experience of and attitudes toward activities to maximise qualification results in England. *Oxford Review of Education*, 44(5), 563–80.

Millman, J. & Greene, J. (1989). The specification and development of tests of achievement *and ability. In R.L. Linn (Ed.). Educational Measurement (3rd edition)* (pp.335–366). New York: American Council on Education/Macmillan.

Newton, P.E. & Baird, J. (2016). Editorial: The great validity debate. *Assessment in Education: Principles, Policy & Practice*, 23(2), 173–177.

Newton, P.E. & Lockyer, C. (2022). *How 'CASLO' Qualifications Work (Ofqual/22/6895)*. Coventry: Office of Qualifications and Examinations Regulation.

Newton, P.E. (1997). Examining standards over time. *Research Papers in Education*, 12(3), 227–248.

Newton, P.E. (2007a). *Evaluating Assessment Systems. Paper 1. Submission to the Education and Skills Select Committee Inquiry Into Testing and Assessment.* London: Qualifications and Curriculum Authority.

Newton, P.E. (2007b). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170.

Newton, P.E. (2008). *Monitoring National Attainment Standards: A Collection of Working Papers (Ofqual/08/3916)*. London: Office of the Qualifications and Examinations Regulator.

Newton, P.E. (2010). Educational Assessment Concepts and Issues: The multiple purposes of assessment. In P. Peterson, E. Baker, & B. McGaw (Eds.).

International Encyclopedia of Education (3rd edition) (pp.392–396). Oxford: Elsevier.

Newton, P.E. (2017a). There is more to educational measurement than measuring: The importance of embracing purpose pluralism. *Educational Measurement: Issues and Practice*, 36(2), 5–15.

Newton, P.E. (2017b). *An Approach to Understanding Validation Arguments (Ofqual/17/6293)*. Coventry: Office of Qualifications and Examinations Regulation.

Newton, P.E. (2020a). M*aintaining Standards: During normal times and when qualifications are reformed (Ofqual/20/6710/2)*. Coventry: Office of Qualifications and Examinations Regulation.

Newton, P.E. (2020b). *What is the Sawtooth Effect? The nature and management of impacts from syllabus, assessment, and curriculum transitions in England (Ofqual/20/6710/1)*. Coventry: Office of Qualifications and Examinations Regulation.

Newton, P.E. (2020c). Deficiency, contamination, and the signal processing metaphor. *Educational Measurement: Issues and Practice*, 39(1), 82–91.

Newton, P.E. (2022). Demythologising A level Exam Standards. *Research Papers in Education*, 37(6), 875–906.

Newton, P.E. (2023a). Taking purpose into account when designing, using and evaluating educational assessments. In R.J. Tierney, F. Rizvi, & K. Ercikan (Eds.). *International Encyclopedia of Education (4th edition)* (pp.136–146). Elsevier.

Newton, P.E. (2023b). Certificating learning on the basis of teacher assessment. In C. Harrison, C. Leung, & D. Pepper (Eds). *Educational Assessment: The influence of Paul Black on research, pedagogy and practice* (pp.221–240). London: Bloomsbury Academic.

Newton, P.E., Clarke, L., Curcin, M. & Brylka, A. (2024). *Origins and Evolution of the CASLO Approach in England: The importance of outcomes and mastery when designing vocational and technical qualifications (Ofqual/24/7156)*. Coventry: Office of Qualifications and Examinations Regulation.

Newton, P.E., Curcin, M., Clarke, L. & Brylka, A. (2024a). *Understanding Qualification Design: Insights from the 2020 to 2024 CASLO qualification research programme (Ofqual/24/7176)*. Coventry: Office of Qualifications and Examinations Regulation.

Newton, P.E., Curcin, M., Clarke, L. & Brylka, A. (2024b). *The CASLO Approach: A design template for many vocational and technical qualifications in England (Ofqual/24/7159)*. Coventry: Office of Qualifications and Examinations Regulation.

Nisbet, I. & Shaw, S.D. (2020). *Is Assessment Fair?* London: SAGE Publications Ltd.

O'Neill, O. (2013). Intelligent accountability in education. *Oxford Review of Education*, 39(1), 4–16.

Ofqual (2009). *Identifying purposes for qualifications in the Qualifications and Credit Framework: Testing and trialling (Ofqual/09/3985)*. London: Office of the Qualifications and Examinations Regulator.

Ofqual (2014). *A Consultation on Withdrawing the Regulatory Arrangements for the Qualifications and Credit Framework (Ofqual/14/5490)*. Coventry: Office of Qualifications and Examinations Regulation.

Ofqual (2020). *Online and on-screen assessment in high stakes, sessional qualifications: A review of the barriers to greater adoption and how these might be overcome (Ofqual/20/6723/1)*. Coventry: Office of Qualifications and Examinations Regulation.

Perie, M. & Huff, K. (2016). Determining content and cognitive demand for achievement tests. In S. Lane, M.R. Raymond & T.M. Haladyna (Eds.). *Handbook of Test Development (2nd edition)* (pp.144–164). New York: Routledge.

Petch, J.A. (1963). *The Joint Matriculation Board: What it is and what it does. Occasional Publication 16*. Manchester: Joint Matriculation Board.

Phoenix, D. (2025). *Increasing Employer Support for the Tertiary Skills System in England*. Oxford: Higher Education Policy Institute.

Polesel, J., Dulfer, N., & Turnbull, M. (2012). *The Experience of Education: The impacts of high stakes testing on school students and their families. Literature Review*. Sydney: The Whitlam Institute, University of Western Sydney.

QCA (Qualifications and Curriculum Authority) (2000). *Arrangements for the Statutory Regulation of External Qualifications in England, Wales and Northern Ireland (QCA/00/589)*. London: Qualifications and Curriculum Authority.

Raymond, M.R. (2016). Job analysis, practice analysis and the content of credentialing examinations. In S. Lane, M.R. Raymond & T.M. Haladyna (Eds.). *Handbook of Test Development (2nd edition)* (pp.144–164). New York: Routledge.

Reay, D. & Wiliam, D. (1999). I'll be a nothing: Structure, agency and the construction of identity through assessment. *British Educational Research Journal*, 25(3), 343–354.

Remedios, R. & Sewell, P. (2024). Employability and motivation: Which motivational theories are most appropriate? *Higher Education, Skills and Work-based Learning*, 14(4), 908–919.

Rethinking Assessment (2023). Creating an Assessment System that Values the Breadth of Strengths of Every Child. Rethinking Assessment.

Ryan, R.M. & Deci, E.L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, Article ID 101860.

Schiro, M.S. (2008). *Curriculum Theory: Conflicting visions and enduring concerns*. Thousand Oaks, CA: Sage Publications Inc.

Schmeiser, C.B. & Welch, C.J. (2006). Test development. In R.L. Brennan (Ed.). *Educational Measurement (4th edition)* (pp.307–353). Westport, CT: American Council on Education/Praeger Publishers.

Shepard, L.A. (1979). Purposes of assessment. *Studies in Educational Evaluation*, 5(1), 13–26.

Shepard, L.A. (1997). *Measuring Achievement: What does it mean to test for robust understanding?* Princeton, NJ: Educational Testing Service.

Spens, W. (1938). *Secondary Education with Special Reference to Grammar Schools and Technical High Schools.* London: HM Stationery Office.

SSEC (Secondary School Examinations Council) (1939). *The Higher School Certificate Examination: Being the report of the panel of investigators appointed by the Secondary School Examinations Council to enquire into the eight approved Higher School Certificate examinations held in the summer of 1937*. London: His Majesty's Stationery Office.

SSEC (Secondary School Examinations Council). (1963). *The Certificate of Secondary Education: Some suggestions for teachers and examiners.* London: Her Majesty's Stationery Office.

Stasz, C., Hayward, G., Oh, S., & Wright, S. (2004). *Outcomes and Processes in Vocational Learning: A review of the literature*. London: The Learning and Skills Research Centre.

TGAT (Task Group on Assessment and Testing) (1988). *Report of the Task Group on Assessment and Testing*. London: Department of Education and Science and the Welsh Office.

Torrance, H., Colley, H., Garratt, D. Jarvis, J., Piper, H., Ecclestone, K., & James, D. (2005). *The Impact of Different Modes of Assessment on Achievement and Progress in the Learning and Skills Sector.* London: Learning and Skills Development Agency.

van der Vleuten, C.P.M. & Heeneman, S. (2016). On the issue of costs in programmatic assessment. *Perspectives on Medical Education*, 5(5), 303–307.

Wahlberg, M. & Gleeson, D. (2003). 'Doing the business': Paradox and irony in vocational education – GNVQ business studies as a case in point. *Journal of Vocational Education and Training*, 55(4), 423–446.

Wang, M., Henry, D.A., & Degol, J.L. (2020). A development-in-sociocultural-context perspective on the multiple pathways to youth's engagement in learning. *Advances in Motivation Science*, 7, 113–159.

Wellings, D., Spours, K. & Ireson, J. (2010). Advanced GNVQs, AVCEs and Level 3 Diplomas in England: A motivational analysis. *Research in Post-Compulsory Education*, 15(4), 387–403.

Wheelahan, L. & Moodie, G. (2017) Vocational education qualifications' roles in pathways to work in liberal market economies. *Journal of Vocational Education & Training*, 69(1), 10–27.

Wheelahan, L. & Moodie, G. (2025). Revisiting credentialism – why qualifications matter: a theoretical exploration. *British Journal of Sociology of Education*, 46(6), 874–892.

Wheelahan, L., Buchanan, J., & Yu, S. (2016). *Linking Qualifications and the Labour Market Through Capabilities and Vocational Streams*. Adelaide: National Centre for Vocational Education Research.

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research & Evaluation*, 2(2), online access.

Wiliam, D. (2013). *Principled Curriculum Design (Redesigning Schooling 3)*. London: SSAT (The Schools Network) Ltd.

Williams, R. (1958). The long revolution. In R. Hooper (Ed.) (1971). *The Curriculum: Context, design & development* (pp.42–48). Edinburgh: Oliver & Boyd with The Open University.

Wilson, F., Evans, S. & Old, S. (2015). Context-led science courses: A review. *Research Matters: A Cambridge Assessment publication*, 19, 7–13.

Wilson, P. (2010). *Big Idea, Small Steps: The making of credit-based qualifications.* Leicester: National Institute of Adult Continuing Education.

Winch, C. (2023). Learning outcomes: The long goodbye: Vocational qualifications in the 21st century. *European Educational Research Journal*, 22(1), 20–38.

Wise, S.L. & DeMars, C.E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1–17.

Wiseman, S. (1961). The Efficiency of examinations. In S. Wiseman (Ed.). *Examinations and English Education* (pp.133–164). Manchester: Manchester University Press.

Wolf, A. (2011). *Review of Vocational Education – The Wolf Report*. London: Department for Education.

Yeomans, D. (1998). Constructing Vocational Education: From TVEI to GNVQ. *Journal of Education and Work*, 11(2), 127–149.

Young, M. (2008). *Bringing Knowledge Back In: From social constructivism to social realism in the sociology of education*. Oxford: Routledge.

Young, M. (2013). Overcoming the crisis in curriculum theory: A knowledge-based approach. *Journal of Curriculum Studies*, 45(2), 101–118.

YoungMinds (2025). *Missing the Mark*. London: YoungMinds.

Published by:

ofqual

2nd floor, 1 Friargate
Station Square
Coventry
CV1 2GN

0300 303 3344

public.enquiries@ofqual.gov.uk

www.gov.uk/ofqual

July 2026

Ofqual/26/7348/2