

The INGOT qualifications design rationale

Background to the design

There has been an on-going debate about assessment for several decades. The heritage of academic qualifications is in controlled handwritten examinations and norm referencing. More recently, particularly in relation to practical learning, criterion referencing established itself in schools and across vocational education. The rationale was to move to fixed bench marks for learning rather than competition with quotas of grades based on cohort distribution of performance based on the assessment.

Five risks with assessment

1. The purpose of the assessment is lost in other less important details.
2. Is what is assessed what is easy to assess rather than what needs to be assessed?
3. The assessment methods limit learning.
4. Assessors are dishonest.
5. When grading qualifications the declared precision in the grades is not justified.

1. There are three clear purposes for assessment. One is to **ascertain competence** eg to carry out tasks in a work place to an acceptable standard or readiness to progress to the next stage of learning, another is to act as a **filter for progression** when the number of places in the next stage is limited. The third is as a **focus to motivate learning**. If the main function of the qualification is to decide base level competence, matching performance to fixed assessment criteria is good enough. If the requirement is to filter on performance some sort of graded assessment is needed that will differentiate those suitable to fit the scarcer higher level places or to determine competence in a particular type of learning at a higher level.

2. Some subject assessment is much more straightforward than others. **Academic subjects** that are largely about knowledge and understanding can be assessed quite uncontroversially using a **conventional paper based examination**. A pure mathematics exam is fairly certain to be all that is needed to assess competence in pure mathematics. If we want to assess competence in a language we need to assess speaking, listening, reading and writing. While there might be a good argument that a good writer will be a good reader it is far less certain that a good writer will be a good speaker. It might be a lot easier to assess writing but if we want valid assessment of all important aspects of language, assessing speaking is important. Taking account of all aspects of **technical qualifications** provides similar needs in assessing breadth, including representative context.

3. There are a number of ways in which **assessment limits learning**. When focusing on exam technique becomes as important as knowledge it is an indication that there is a serious imbalance. An example was when [Sir Trevor Nunn, Artistic Director of the Royal Shakespeare Company](#) [1], Cambridge graduate and possibly the world's leading authority in Shakespeare was judged to have achieved a Grade B in the Shakespeare questions in A level English exam he was asked to sit as part of a media experiment. The reason he was given was he lacked exam technique. A graded assessment is supposed to check level of competence in a subject, not competence in taking assessments. Other ways in which assessment can limit learning include **over-elaborate bureaucratic procedures** that distract teachers from teaching, excessive certification costs **diverting resources** away from teaching, **discouragement and demotivation**. Targeting ages such that the **highest attainers underachieve** because they only achieve the highest grade possible in the level at the end of a Key Stage when they were capable of a lot more.

4. There is a risk that if assessors are **dishonest** they will return **inaccurate results**. This applies to all forms of assessment although it has traditionally been more associated with **coursework** than

The INGOT qualifications design rationale

-->

with terminal academic testing. The main problem is in **high stakes accountability** providing a reason to be dishonest. Equally there is the incentive to **enforce invalid assessment** merely to reduce the opportunity for dishonesty to take place.

5. For any measurement to be valid it should have an associated [uncertainty calculated](#) [2] for the measurement. If the walls of my room are only parallel to the nearest centimetre there is no point in measuring the room width to the nearest millimeter. There is more to this than simple statistical uncertainty. Sample size for a group might indicate a very different level of precision than for an individual and qualifications are about individuals and their progress just as much as groups and national statistics.

Design principles

Level 1 to 3 qualifications used in the 11-18 sector.

A. The purposes of qualifications in this age range is

1. To provide recognition of milestones in learning that will help motivate higher performance.
2. To provide a focus to ensure that programmes of study are completed.
3. To inform stakeholders of the appropriate progression routes from a particular level of learning.
4. To provide a means of filtering where there is competition for places in the next stage of learning.
5. To provide a means of holding schools and teachers to account.

The TLM INGOT assessment method targets each of these purposes. The **coursework** element enables the **milestones** to be set **through** each of the broad **national levels** with **Level 1** corresponding to attainment typically achieved by a majority in **Key Stage 3**, **Level 2** typically a majority in **Key Stage 4** and **Level 3** a significant minority in **Key Stage 5**. Too much focus on **age** presents a significant **risk** because attainment is likely to be normally distributed at any particular age. **Bench marks** for all are **scientific nonsense** although **minimum thresholds** make more sense. When nearly all the top quartile achieve the highest grade available in the assessment at age 16, perhaps the **top 10%** could achieve this a year or even **two years earlier**. In such cases it would be far more sensible to start work targeted on the next level otherwise the **brightest individuals** are **significantly under-achieving**. TLM does not force anyone to teach to any particular timings, the main interest is whether or not the learner reaches the standard as quickly as they can and then moves on to the next level. This then **optimises progress in keeping with** [cognitive research evidence for learning](#) [3].

B. Progress through content

The **coursework** element will track **progress through** the programme of study at the base **level**. This measures progress "horizontally" at a **baseline** of knowledge and competence qualitatively specified by the **level descriptor** and the **assessment criteria** and associated guidance. There is **no grading**, either the candidate can meet the criteria and they have completed the full range of they can't. The chances are that if they can get through the material but not meet the level descriptors they are at the level below. The advantage of this approach is that there is **no need** to put a lot of resource into **expensive procedures** that are in themselves not as precise as most of the outcomes suggest. There is a **motivating imperative** to complete the coursework because the candidate can not take the grading exam until they have completed the coursework at the level of the grading. **Bright candidates** have the freedom to excel and **innovate** [4] in their coursework without worrying about the constraints of exam techniques and minor procedural errors costing them a high grade.

C. Progress through grades

The **coursework coupled** to the **grading exam** enables **inclusion** for those that can demonstrate practical baseline competence. The **grading exam** can then act as a **filter** for those with potential for **higher level academic study**. Any candidate that successfully **completes** the coursework is likely to **learn** enough to get enough marks on the grading exam **to pass**. There are a mixture of questions in the grading exam, some **very difficult** for the level. This means we have the **leeway** in the exam to **stretch** the **highest attainers** and at the same time included those that struggle in controlled academic tests. This latter group are the most likely to be demotivated at the prospect of two years work that depends solely on a terminal exam. The rationale of not restricting an exam to a set time allows a more efficient approach to [reasonable adjustments](#) [5]. We have **multiple versions** of the exam and so we can provide **testing on demand** subject to the completion of the coursework. We can also provide practice exams on line exactly matching the real thing to gauge readiness to complete the assessment process with **planned progression** to support the **next level** if **top grades** can be achieved.

The assessment model is easy to tune to make sure it is statistically of **comparable difficulty** to any other qualifications at the same level thus making it suitable in contributing to performance tables.

D. Validity of assessment

Providing **flexibility** in the **ungraded coursework** element makes what we **need to assess** easier to assess. **Flexibility** and removing any requirement to grade the coursework enables assessment to fit **demonstrated competence** in the context of normal work. Teacher/assessors can source their evidence by any valid method that provides **convincing evidence** that the **assessment criteria** have been met. There is no restriction on the organisation or activities beyond **meeting the assessment criteria** as an indicator that the **learning outcome has been achieved** to the characteristic specification of the **broad level descriptor**. The course can be taught in **units, linearly** or in a **combination**, the process is really not for us to determine, as an **awarding organisation** we are quality assuring **outcomes** not the means of achieving those outcomes. This means that **all aspects** of the content of the qualification **can be assessed** without compromise due to either inappropriate assessment methods or by fine grading that is of very dubious precision. There is really very **little quantitative evidence** that would stand up scientifically that eg National Curriculum levels broken in to a, b, c **sub-grades** has any **justification**. If a piece of English is graded **level 5c** by 100 teachers, how likely is it that a different random sample of **100 teachers** would grade the work 5c and what would be the **uncertainty** in the measurement? And that assumes the simple case of one piece of work. Even if such an exercise gave confidence in the assessment of that piece of work it does not mean there will be the same confidence in something different. Even if the **averages** turned out to be 5c for both groups of teachers, there is no certainty that for **individual learners** the variation does not span entire levels, never mind the sublevels. In short, assessing coursework is not that precise.

The **narrower** we make the assessment the more likely a **statistical** exercise will demonstrate **reliability**. If the assessment is limited to spelling words, a good level of agreement can be expected. If the assessment is about **writing style** or speaking effectively to an audience the level of **agreement will fall**. For a valid assessment, sampling **all aspects** of the subject matter is important, not just those that can be measured with precision. We need to prevent **less precise** but important aspects from being **ignored** especially in high stakes systems that will result in pressure to do exactly that and thus **compromise learning**.

E. The risk in trusting teacher assessors

There is no doubt that **high stakes** performance data puts pressure on teachers to be **dishonest**. This then adds risk to the system. We can **mitigate risk** presented by teacher dishonesty by **adding risk** that assessment is not valid or reduces motivation. **Substituting** one risk for another might be **justified** if one can be shown **objectively** to have a **bigger effect** than the other.

The prevalent view is that **coursework** presents the **biggest risk** of teacher **dishonesty**. To treat

```
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)})(window,document,'script','//www.google-analytics.com/analytics.js','ga');ga('create','UA-46896377-2','auto');ga('send','pageview');
```

The INGOT qualifications design rationale

-->

this risk objectively requires going back to purpose. At Level 2 at age 16 no-one now gets directly into employment, the **qualifications** are primarily about **motivation, informing progression routes and accountability**. Coursework has an important role to play but it has been discredited by coursework only approaches that resulted in pupils achieving the **highest possible grades** when they could not even **pass** in other subjects assessed by controlled end testing. This meant that some pupils believed they would cope with **Level 3** academic courses with **terminal examinations** and that was clearly not going to be the case. If it is a practical subject there is no **validity** in an assessment that is **void of practical** context. The question is, will providing extended controlled practical testing **reduce risk**? The **evidence** seems to be **against** this. If a controlled practical exam is taking place over several weeks it only takes one learner or accomplice to post solutions to that assessment on the internet and the entire security is compromised. We might be mitigating the threat of teacher dishonesty but we are massively increasing an even bigger potential risk. The approach of **controlled practical** examinations has many of the characteristics that detract from learning and can **weaken validity** eg by considering contrived rather than real contexts and by putting numbers to marking schemes that provide a **misleading sense of precision**. In addition the **cost** is significant in teacher administration and the opportunity cost in teaching time.

The first **strategy** in maintaining coursework but **mitigating** the associated **risk** is to make it very unlikely that the coursework on its own will **route** a learner to the **wrong place** for progression. Learners' interests must come first. That then makes the effects of the risk of teacher dishonesty much less important. By providing a competence based coursework element as a **qualifier** for taking the exam, we make it **important** but **not critical** in the assessment overall for **high attainers** that are likely to go on to **academic routes**. We keep the design purpose of **motivation** and tracking progress through the basic core content for academically **weaker learners** and maintain their engagement and learning leaving open options as some individuals will change their levels of competence at unpredictable rates with age. If a school wants to recognise particularly good coursework there is nothing to stop them doing this outside the qualification and the learners themselves have the flexibility to gain recognition eg by using their coursework in other contexts eg by the number of YouTube plays.

The second **strategy** is to make it clear to the teacher/assessor that they are **accountable for standards**. All assessors must **sign a declaration** before they can use the on-line recording system and that is the only means of recording the assessments made. All centres must have a **Principal Assessor** who also signs an agreement to take **responsibility for standards** across the centre. While there is no certainty that no assessor will ever act dishonestly, it seems rather less risky than **relying on** all the **students** to act honestly by not sharing the solutions to a **controlled test**. Since the TLM coursework assessment is **delegated** to local assessors they are well-placed to ensure the tasks they set for assessment and more importantly the **learner outcomes** are **fairly indicative** of the learner's **competence** in relation to the learning outcomes that the criteria underpin. If the evidence is on the evidence management system any of it can be sampled at any time and the teacher assessors don't know which will be.

The third **strategy** is to use the grading exam as an indicator of **coursework fidelity**. It is unlikely that large numbers of learners will fail to gain sufficient marks in the grading exam if they have completed the coursework to the appropriate level. If large numbers from a particular centre do not get sufficient marks for the lowest exam grade we can **investigate** to see why. This causes absolutely no additional work for the centre because they would be doing the grading exam in any case. Another spin off from moderation on demand is that we can provide early feedback reducing the risk of any coursework surprises after it becomes too late to do anything about it.

If we are genuinely worried about teachers and learners being dishonest, why are we not more concerned about **certificate forgery**? It is far easier to forge a certificate than it is to systematically cheat in coursework. It is also easier to cheat in an exam from the point of view of not having to do as much work.

F. Grade inflation

Grade inflation has gradually crept up over the years. Whether or not this is a good or bad thing

depends to an extent on point of view. It is partly a symptom of the lack of rigour in determining measurement uncertainties outlined in paragraph 5 above. If you can't quantify the uncertainty how do you know whether or not consistency over time is being maintained? There is a very poor tradition of measurement practice when qualifications are compared to science and engineering. On the one hand we pretend that qualitative descriptions are cast iron benchmarks and on the other that they have no place when norm referenced assessment of large populations can give much more precise quantitative results. The TLM assessment model addresses this problem by using a competence based criteria matching element for practical assessment that we don't pretend can be graded with multi-fine structured precision and we use a grading exam that can be norm referenced to other qualifications to provide as tight or as loose correlation between grades as is desired. Even between qualifications of the same type, eg mathematics and English GCSE there are significant attainment differences for large population cohorts of the same learners. It is difficult to understand why it is acceptable for eg GCSE English to be easier than GCSE mathematics in terms of the percentage success rates when it would be easy to normalise them. If the results are different we are saying either that one subject is fundamentally more difficult than the other or the teaching in one is better than the other. Difficulty is arbitrarily determined by setting grades across a normal distribution. That distribution can be distorted by over-focus on particular grades but in the end all of that is arbitrarily rather than rationally determined. With all this in mind the TLM design makes it easy to tweak either the grade boundaries or the questions to change the difficulty of the examination. We expect to make the examinations somewhat more demanding in the near future because to start with they are an unknown quantity and teachers will inevitably prepare learners better once they are more familiar with the methods.

G. Summary

The TLM strategy is to require coursework to be completed to a minimum level in keeping with the qualification level so that important practical aspects of assessment count and are covered. We want candidates to demonstrate practical competence in realistic rather than artificially contrived contexts. We do not want to cause a large assessment overhead on schools. The strategies we have adopted balance risk against cost and validity and unlike many other vocational qualifications that tend to be more expensive than their academic counterparts, we can provide qualifications at lower cost not only in the monetary terms of the qualifications but also in the cost in teacher time spent on administration, moderation of grading and similar activities is almost certainly far greater. The methods scale well and they are easy to adjust in order to make assessment more or less demanding.

Source URL: https://theingots.org/community/qual_design

Links

- [1] http://en.wikipedia.org/wiki/Royal_Shakespeare_Company
- [2] <https://www.dit.ie/media/physics/documents/GPG36.pdf>
- [3] <https://www.youtube.com/watch?v=gtmMMR7SJKw&feature=youtu.be>
- [4] <https://www.youtube.com/watch?v=DvqxjZr79eE>
- [5] <https://theingots.org/community/node/5494>

The INGOT qualifications design rationale

-->
