



Assessing the Assessment: 1–The New GCSE English

‘Strength’ and ‘rigour’ in the new GCSE English assessment

Peter Thomas explores the changes to assessment at GCSE, and looks at some of their implications for teaching, learning and professional development in English.

The rationale for change

The reformed English GCSEs now being taught for first certification in 2017 represent a major change for English teachers and for students, with both concerned about the implications for results next August. The most obvious aspect of change has been in the mode of assessment, with the ending of phased teacher-assessed classroom work, and its replacement by end-of-course externally-assessed examinations. This change was a direct consequence of two governmental perceptions: firstly, that teacher assessment had become unreliable, and secondly, that external marking would be more accurate and consistent. Both perceptions of the assessment process are worth some scrutiny, particularly in view of the recruitment and training of a vastly increased and inexperienced examiner workforce.

Less obvious, but equally revealing of the thinking behind reform, is the difference in what is to be assessed. In English Language, spelling, punctuation and grammar take on a greater significance with enhanced weighting, and, in English Literature, memory recall of studied set texts becomes a necessary skill, together with use of technical vocabulary, and writing in a literary critical style. The result of these organisational and structural changes is a significant

difference in what students are taught and encouraged to learn. They are already having a major impact on classroom teaching and learning in English Language and English Literature.

In addition to government perceptions of deficiency in assessment and in curriculum content, there was a curiously un-Tory governmental view that a free market competition between awarding bodies had resulted in undesirable consumer choice, with some being seen as ‘easier for a C’ or ‘more demanding for the more able’. Reform was designed to eliminate differences and discrepancies between awarding bodies reflected in variable standards of award.

Assessing assessment

As part of NATE’s role in monitoring and evaluating educational policy and practice, this article examines some of the ways that changes in assessment may impact upon classroom teaching and learning. It recognises evidence that both modes of assessment – centre-assessed and externally-assessed – can be variously successful and flawed, and that a reliably consistent and objective form of assessment is one that reduces the complexity of what candidates are asked to do – such as multiple choice questioning.

However, given the complex mix of knowledge and skills involved in reading in English – an interweaving of the cognitive and the affective, of the interpretative and the analytical – it's clear that any system of assessment has to accommodate elusive and ambiguous elements of human response. Likewise in writing, the clearest assessment emerges from reducing writing to a matter of presentational accuracies, whereas the communicative repertoire in English involves necessary complexities of form, audience and purpose.

Contesting versions of English and its assessment

The changes were triggered by Governmental judgement or perception that GCSE assessment and the GCSE curriculum were both in need of 'strengthening' to correct a tendency to an increasing share of A*-C awards over time. The judgement was couched in terms of a need for greater 'rigour', more emphasis on 'basics' and a check against 'grade inflation'.

Some of the assumptions packaged into this perception of deficiency and need are worthy of scrutiny – for example, the notion that correct SPaG and the ability to write an 'essay' represent the 'basics' of English, and that the ability to identify zeugma and synecdoche represent 'rigour' in subject knowledge, let alone desirable attributes of an educated mind. The reformed English agenda puts a new premium on presentational and knowledge-based virtues for various reasons. Among these it is possible to see the brandishing of cultural tokens and totems and the prioritising of visible and concrete virtues. These may be a valuable part of education, and may also be valuable as headlines and sound-bites featuring terms such as 'crackdown', 'slipping standards' and 'world-class'.

Of course, there will be those who dissent from such views – largely the professionals and academics who have been, every day and throughout a lifetime, engaged with classrooms, learners and their specialist subject. It is unlikely to be a potent dissent. The last time there was potent dissent was in 1993, when John Patten, the Secretary of State for Education was humiliated after declaring 'these tests will take place' by the following day's 92% boycott of the flawed initial KS3 SATs. Since then, every Secretary of State, with the exception of Estelle Morris, the only occupant of that office to have been a teacher, has seen the English teaching community as the enemy rather than the ally.

As a result, for several years there has been no real dialogue between educationalists and politicians – certainly none about fundamental principles and practices in education – because those with power have bypassed those with knowledge, preferring the counsels of the red-top and Murdoch press to those dismissed by a previous education secretary as 'the blob'. It is a sad reflection of the intellectual life of a country that Education, one of its most vital and transformative agencies, has been dominated by glib clichés, wobbly statistics, ill-founded comparisons and the financial reward of those who give the answers favoured by those who ask the loaded questions.

Much of what has been offered as justification for these and other changes amounts to familiar political rhetoric and ideologically-driven notions of education, but they are particularly unhelpful when they relate to something as complex as the English curriculum,

its teaching and its assessment. These require an informed grasp of student motivation, learning processes, teaching strategies and reasoned approaches to English and education in their broadest and narrowest contexts. The purpose of this article is not, however, to engage with the myths, misconceptions and ignorance that contribute to most politicians' and press views of education. It is, rather, to examine what has resulted from the proclaimed strengthening and its intention to remove the variability between different awarding bodies.

Teacher expertise and professional development

The effects of these reform changes were intended to be profound, though whether they will be profound in the way they were intended has yet to be seen. It is, in this case as in others, worth setting the situation in a historical, if not pedagogical, perspective. The change in the mode of assessment affects teachers most directly because it ends a history of involving teachers in the assessment process that began with the coursework-based GCSEs linked with the development of the National Curriculum in 1984.

This thirty-year period of enhanced teachers' professional responsibility was supported by an extensive standardisation procedure which involved face-to-face meetings with senior moderators and a bank of exemplar materials serving as benchmarks of applied criteria. The exemplar materials also served as models of varied practice and varied outcomes. Apart from illustrating what work at various levels looked like, it was also possible for awarding bodies to illustrate good and less good task setting, good and less good annotation and, in Literature, good and less good selections of texts for study.

Standardising meetings had other functions productive of developing expertise: they allowed dialogue between teachers and moderators and board staff, and they provided a networking forum where exemplar scripts were argued over, not presented as cascaded instruction, as in, for example, Literacy Strategy meetings, where questioning outside the cascade protocol was disallowed.

The new regime of 100% examined assessment and, in Literature, externally set texts, makes much of the training mentioned above unnecessary. This loss of free, on-going, subject-specific, professional development may or may not be followed by new forms of training, perhaps involving standardising and exemplar work, though the disappearance of local authority advisors has diminished a valuable source of training initiative and expertise. No doubt, the academy chains will recruit their own advisers and schools will find other means of out-sourcing expertise, but the likelihood is that these will be motivated more by target-driven expediency than educational principle.

Teacher response to this new all-exam regime will vary in kind and force. Some teachers will see all this as a welcome lessening of their marking load, particularly with the removal of the controlled assessment treadmill. Others will see it as a sign of de-professionalising teachers and teaching, divorcing a well-established link between assessment and learning, and that disappeared notion of assessment for learning. Those outside the world of English teaching, particularly if influenced by parts of the national press, may consider that the new arrangements put a rightly higher value upon 'the

“Much of what has been offered as justification for these changes amounts to political rhetoric and ideologically-driven notions – particularly unhelpful when they relate to something as complex as the English curriculum, its teaching and its assessment.”

basics' of the English curriculum and the 'tougher' form of testing by timed examinations. Whatever the response, it is likely that students are prepared for the new all-or-nothing mode of assessment by frequent rehearsals of the timed testing. That in itself is a major cultural shift. And additional marking.

A common assessment framework for English

Given the panic of the awarding bodies following the results of 2012, and the ensuing cowed compliance with the Ofqual strictures on GCSE assessment, the new specifications for the three English awarding bodies are more similar than in previous years. This is because the specifications were constructed to a very tight remit. All share the same assessment objectives and meet the required content coverage, so there is less to choose between specifications than before. Given the drive towards a regulated conformity and consistency across the awarding bodies and their specifications, it is worth some scrutiny of the kind of conformity and consistency achieved. It is also worth some scrutiny of the ways in which the awarding bodies have attempted to develop some distinctiveness where possible.

Where distinctiveness is possible it is largely in the construction of mark-schemes. These become a vital part of the whole process, not only in the homes of the thousands of markers recruited – some of them specialist English teachers, some not – but in the classrooms where students are prepared for the specific skill-sets associated with different levels of attainment.

What may be hoped for in a major revision of assessment practice is some national consensus of the core skills of reading and writing expressed in a secure hierarchy of attainment useful to examiners and to teachers. However, it quickly becomes evident that there is a conflict between the needs for conformity and the needs for distinctiveness. Although all awards must conform to the new nine-point numerical scale that replaces the old (nine-point) alphabetical scale, the bands themselves are not consistent across the specifications. In English Language Reading, AQA uses a four-band range, Edexcel uses a five-band range and OCR uses a six-band range.

As the basis for an uneven nine-point award, it may be that the five-point model is better adapted, and a better representation of the national ability profile, and the six-point model may provide examiners with more grounds for fine discrimination, but the AQA model may have the rather different advantage of retaining some opacity when converting raw marks to UMS and establishing borderlines. The way in which attainment is defined by band-descriptors is not only important for training examiners, but it has massive implications for teachers looking for the pedagogy involved, and finding practical ways of translating it into classroom teaching and learning.

The new GCSE mark-schemes

A mark-scheme's embedded pedagogy becomes apparent in its hierarchy of skills defined by mark bands and their attainment descriptors.

English Language: Reading

Those for English Language Reading are listed below on the left, mainly (though not consistently)

expressed as nouns or verbs, for example, '*analysis/analyse*' and mainly indicating cognitive features. Further discrimination emerges from the additional qualifying descriptors on the right, mainly (though not consistently) expressed as adjectives, and indicating operational skills, for example '*perceptive/relevant*'. What immediately becomes clear from comparison is that not all of these key words in the assessment lexicon mean the same thing in all mark-schemes.

“The way attainment is defined by band descriptors has massive implications not only for training examiners but also for classroom teaching and learning.”

| Language (Reading) | | |
|--------------------|--|---|
| AQA | | |
| 4 | analyses | perceptive, detailed |
| 3 | explains | clear, relevant |
| 2 | understands | attempts, selects |
| 1 | awareness | simple |
| Edexcel | | |
| 5 | analyses | discriminating, clarifying |
| 4 | explores | detailed, supported |
| 3 | explains | appropriate, relevant |
| 2 | comments | selects |
| 1 | identifies | limited |
| OCR | | |
| 6 | analysis | skilled, sophisticated, consistent |
| 5 | analysis | perceptive, balanced |
| 4 | explanation | secure, relevant |
| 3 | explanation | understanding, relevant |
| 2 | comment | straightforward |
| 1 | describes | |
| Eduqas | | |
| 5 | analysis, exploration, evaluation, engagement | accurate, convincing, perceptive, effective |
| 4 | analysis, exploration, evaluation, engagement, awareness | accurate, relevant, clear, effective |
| 3 | explanation, evaluation, understanding, awareness | supported, appropriate, some, begins |
| 2 | identification, comment, opinion | supported, straightforward, limited, some |
| 1 | identification, comment, opinion, reference | simple, basic, begins |



“Teachers looking for a distinctive culture of assessment in the various mark-schemes will note that all apart from Edexcel prioritise analysis as the prioritised skill.”

There are various merits and various oddities here. Whereas all agree on the highest level of skill being ‘analysis’, there are interesting differences in the sequence of rungs in the attainment ladder. The gap between AQA’s ‘explains’ and ‘analyses’ and OCR’s similar stepping, seems a large cognitive leap, much better filled by Edexcel’s ‘explores’, which indicates some independent and investigative reading. On the other hand, AQA’s bottom two rungs establish a progression from ‘awareness’ to ‘understanding’, which is well suited to the cognitive progression leading up to ‘explain’. In this respect, OCR’s failure to distinguish between its two levels of ‘analysis’ and ‘explanation’ seems to make a 6-point scale unnecessary, though the additional descriptors help to discriminate within the named skills.

OCR also has a very odd use of ‘consistent’ as a top band descriptor, above ‘perceptive’, which AQA, quite properly, identifies as a top band feature. ‘Perceptive’ surely merits top-band status if it is understood to entail independence of notice and discovery, rather than application of a learned formula, which could be a route to OCR’s top-band ‘skilled’. What seems difficult to see as a progression is the OCR stepping of ‘describe’ and ‘comment’, though this may reflect a belief that attitude or opinion may be regarded as higher skills than recording what is observed. That’s a distinction worth some discussion, particularly if ‘comment’ is unsupported by evidence of having read enough to ‘describe’ it. Edexcel’s ‘identifies’ may be a clearer indication of what candidates can do at this level.

Where the three sets of descriptors agree, it is in the nature of *explanation*, where all agree that ‘relevant’ is a helpful qualifier. Otherwise, the mark-schemes seem to direct examiners to assessment bands based on different models of progression in reading. Whether these implied models of progression allow teachers to choose a specification for preferred pedagogy is possible but doubtful. Much will depend on published exemplars justifying the application of the mark-schemes.

Eduqas provides a wider range within its hierarchy of skills, with an attempt to show where banded skill-sets may show overlap. This may be a useful and realistic guide to ‘best-fitting’ where responses are, as they often are, uneven. The bottom band is very well defined with ‘simple’, ‘basic’ and ‘begins’, though ‘some’ is rather unhelpfully used in the band above.

English Literature

A similar scrutiny of the various Literature mark-schemes results in similar observations.

| Literature | | |
|------------|-------------------------------------|--|
| AQA | | |
| 6 | analysis & exploration | critical, convincing |
| 5 | consideration | thoughtful, developed |
| 4 | understanding | clear, appropriate |
| 3 | explanation | structured, supported |
| 2 | comments | relevant, supported |
| 1 | awareness | simple |
| Edexcel | | |
| 5 | interpretation | assured, critical, discerning |
| 4 | engagement | developed, effective |
| 3 | personal response | appropriate, relevant |
| 2 | narrative | inconsistent, insecure |
| 1 | response | simple |
| OCR | | |
| 6 | analysis | informed, pertinent, perceptive, sensitive |
| 5 | examination | developed, thoughtful, convincing |
| 4 | personal response | detailed, relevant, competent |
| 3 | personal response | reasonable, relevant |
| 2 | response | straightforward, supported |
| 1 | comments | basic, limited |
| Eduqas | | |
| 5 | analysis, appreciation, originality | assured, sensitive, evaluative, perceptive |
| 4 | analysis, understanding | thoughtful, apt, secure |
| 3 | comment, understanding, engagement | relevant, straightforward, appropriate |
| 2 | comment, understanding, reference | simple, limited, some |
| 1 | comment, understanding, reference | basic, simple, limited, some |

Here, AQA seems not to have pursued the notion of consistency across reading in English Language and reading in English Literature, for ‘understanding’ and ‘explanation’ are reversed in the hierarchy. The useful ‘exploration’ in Literature is linked at the top with ‘analysis’, both indicative of very assessable active reading, but the rung below, featuring the more nebulous ‘consideration’, seems open to many kinds of interpretation, and therefore confusion. Such ambiguity is not the best basis for assured examiner judgements – especially when a vastly expanded and largely inexperienced examiner workforce has to be trained to ensure consistency and accuracy.

Other differences immediately appear. Edexcel appears to favour ‘interpretation’ over everything else, and its other rungs also emphasise the nature of personal rather than academic merits in ‘engagement’ and ‘personal response’. Whether this will result in a

“Such ambiguity is not the best basis for assured examiner judgements when a vastly expanded examiner workforce has to be trained to ensure consistency and accuracy.”

greater reward for impressionistic and personal work rather than forensic scrutiny is an interesting point. It could be a selling point for Edexcel among teachers trying to make of English something more than the arid linking of textual features with terminology, supported by whatever acronym takes over from AFOREST and the like. However, the liberal steer apparently privileging the reader over the text in Edexcel’s ‘interpret, engage, personal response’ top descriptors is not wholly echoed in the additional descriptors, where all bands apart from Band 2 establish what the candidate can do, whereas the Band 2 descriptors ‘insecure, inconsistent’ define what the candidate *cannot* do.

Again, the subsidiary descriptors are what may provide teachers and students with the best keywords to read and write by. AQA has a secure distinction between ‘supported’ comments and ‘developed’ ones, though the insertion of the bland ‘clear and appropriate’ between them is a less plausible cognitive stepping.

The OCR additional descriptors offer a more plausible progression, but the main descriptors are less plausible. The difference between band 2 ‘response’ and band 3 and band 4 ‘personal response’ does not seem well-calculated to lead examiners to clear discrimination, nor does the band 1/band 2 difference between ‘comments’ and ‘response’. It is left to the additional ‘supported’ to provide a clear discrimination here.

Teachers looking for a distinctive culture of assessment in the various mark-schemes will note that all apart from Edexcel prioritise analysis as the prioritised skill. Looking at the hierarchy of qualifiers reveals some potentially interesting and helpful discriminators. Whereas AQAs and Edexcel’s top band

‘critical, convincing’ and ‘critical, discerning’ seem to privilege a discursive and cognitive model of response, OCR and Eduqas seem to privilege a more responsive and affective response with the inclusion of ‘sensitive’ as a mark of top attainment.

What all this amounts to is, firstly, a considerable degree of difference between the awarding bodies in what they value in a hierarchy of attainment in reading. There seems little agreement on the fundamental subject-specific pedagogy here, but this may be something of appeal to teachers. If it is thought that Edexcel favours one kind of response, and AQA another, then there will be immediate implications for what happens in classrooms. However, it does not seem to be what was intended in the government aim to import ‘strength’, ‘rigour’ and raised ‘standards’, or to remove differences between the various awarding bodies.

English Language: Writing

Consideration of the assessment of writing reveals a similar pattern of band variety with 4, 5 and 6 (using the bands for AO5).

| Writing | |
|---------|---|
| AQA | |
| 4 | convincing, crafted, developed, complex, varied |
| 3 | clear, engaging, connected |
| 2 | successful, controlled, relevant, paragraphed |
| 1 | simple, limited |
| Edexcel | |
| 5 | subtle, sophisticated, sustained |
| 4 | effective, cohesive, deliberate |
| 3 | appropriate, developed, connected, clear |
| 2 | aware, straightforward |
| 1 | basic, limited |
| OCR | |
| 6 | sophisticated, ambitious, skilful |
| 5 | confident, sustained, controlled |
| 4 | clear, chosen, well-managed |
| 3 | sustained, appropriate, clear |
| 2 | mostly appropriate, attempts |
| 1 | some, limited |
| Eduqas | |
| 5 | sophisticated, sustained, confident, convincing |
| 4 | consistent, secure, well-judged |
| 3 | clear, appropriate, coherent |
| 2 | some, clear, limited |
| 1 | basic, some, simple |



Again, differences emerge in the evaluative language of the band descriptors, with Edexcel, OCR and Eduqas favouring *'sophisticated'* at the top, and AQA avoiding, as elsewhere, a term it has previously trusted, but glossing it with a compound formula of *'convincing, crafted, developed, complex, varied'*. This compounding of skills provides a helpful variety of access routes to the top band, which may help teachers to develop a repertoire best suited to individual students, but the AQA hierarchy of writing attainment seems lacking in clear stepping when, below this full description of top-band performance, bands 2 and 3 seem very similar – even exchangeable. It is difficult to see how *'clear, engaging, connected'* define attainment superior to that of *'successful, controlled, relevant, paragraphed'*. That band discrimination would seem to need some very careful exemplification if it is to be translated into reliable marking, or be helpful in teaching and learning skills progression in the classroom.

“Teachers will not want to wait much longer for further exemplification to pave the way for acceptance of a realistic first award in 2017.”

Teachers who care about motivating and valuing the less academic – those guilty of failing to meet government performance baselines – need guidance from mark-schemes. It is not nerdish to want consistency, for example, in the hierarchy of *'some'* and *'limited'*. It would be helpful if all agreed that there has to be *'some'* before it can be *'limited'*, so *'limited'* should be above *'some'* in any hierarchy

Finally

Generally speaking, the evidence over the years is that teachers' (and examiners') assessment of writing is more confident, consistent and secure than the assessment of reading. This would suggest that reading mark-schemes are a priority case for establishing a reliable model of progression in the new regime. The links between reading skills and writing skills would also suggest some need for a harmonising of mark-schemes for both. Some of the inconsistencies noted above would suggest that these are yet to be achieved.

As these mark-schemes are unlikely to be altered for some time, we must look with interest for ways in which the awarding bodies support their work with exemplars and commentaries which flesh out some of the implications of the mark schemes and eliminate some of the ambiguities and uncertainties. Teachers will not want to wait until 2017 for this, and will want early exemplification which paves the way for acceptance of a realistic first award in 2017.

That first award will be eagerly and anxiously awaited. The anxiety may be excessive, however. No government is likely to risk the anger and disappointment of teachers and students, or vote-wielding parents, by delivering a massive drop in attainment or *'standards'* as the price of its much-vaunted *'strength'* and *'rigour'*. It is likely that statistical data will result in similar award distributions whatever the judgemental differences or changes in the reformed GCSE.

August 2017 will be most interesting as a measure of how much has changed, at what cost, and to what advantage. Then we will all know the true length of the strength and the true vigour of the rigour.

Peter Thomas

is Vice Chair of NATE. He has been a Principal Examiner and Principal Moderator and was Lecturer in Education at the Institute for Learning at the University of Hull.