

How Should Examinations for 15-19 Year olds in England be Run?

A SCORE response to the Education Select Committee inquiry

7 November 2011

Cc Marianne Cutler (Association for Science Education)
Professor Peter Main (Institute of Physics)
Libby Steele (Royal Society)
Professor Jim Iley (Royal Society of Chemistry)
Rachel Lambert-Forsyth (Society of Biology)

1. SCORE is a partnership of organisations, which aims to improve science education in UK schools and colleges by supporting the development and implementation of effective education policy. The partnership is currently chaired by Professor Graham Hutchings FRS and comprises the Association for Science Education, Institute of Physics, Royal Society, Royal Society of Chemistry and Society of Biology.
2. The current examinations system is not fit for purpose and SCORE welcomes this timely inquiry from the Education Select Committee. The assessments are not testing the specifications; therefore, even students with high grades are not prepared for the next stage in their career or education – despite the fact that the specifications suggest that they should be; and consequently, consumers of qualifications have lost confidence in the examinations system. This has come about because the five main Awarding Organisations (AOs) which cover England, Wales and Northern Ireland¹ are competing for market share on the basis of enabling more candidates to get higher grades rather than on the basis of high quality assessments or high quality curricula specifications. We ask that the Select Committee recommends significant changes that include drivers for quality in the examinations system and bring an end to the ‘race to the bottom’.
3. In summary this SCORE response:
 - Sets out the characteristics for an effective examinations system and analyses these in respect of failings of the current system;
 - Sets out alternative models for examinations systems including some in which competition is not for market share within a qualification;
 - Calls for it to be a formal requirement of the regulator to review assessment material prior to use, to prevent problems in the quality and accuracy in examination papers;
 - Raises serious concerns about the management of conflicts of interest between the awarding functions of an AO and any other activities AOs (and their related companies) undertake.

Characteristics of an effective 15-19 examinations system

4. An effective 15 -19 examinations system should:
 - **Set and maintain standards** – Assessment results should be comparable as far as is reasonably possible from year to year in order to maintain confidence in the system (the results being the public face of the summer examinations). This allows employers and Higher Education Institutions (HEI) to compare fairly two people who took the same qualification years apart. It is fine for grades to improve so long as this is a result of better teaching and learning rather than through schools changing specifications.

¹ AQA, CCEA, Edexcel, OCR and WJEC

- **Produce fair and effective assessment tools** – The assessment tools must effectively measure a learner’s ability in a subject. They must also be designed to differentiate fairly and reliably between the excellent, good and weak candidates.
- **Engender high quality teaching and learning** – There will always be an intrinsic link between assessment and the teaching and learning of a subject. Assessment must therefore test all levels of Bloom’s taxonomy² to encourage high quality teaching and learning.
- **Authentically represent the subject being assessed** – Specifications and assessment should not lose the character and ethos of a subject in the practicalities of setting and marking assessments. Examinations in biology, chemistry and physics should assess the subject-specific capabilities of candidates rather than generic abilities such as being able to recall facts.
- **Encourage and allow for progression** – Learners obtaining any qualification should be equipped with the necessary subject knowledge and skills to progress to the next relevant level in that subject.
- **Embrace subject expertise** – The two previous points emphasise the need for the development of specifications and assessments to be carried out, supported and regulated by people with expertise in that subject. Any AO and its regulator must have subject expertise in-house. In addition, the system should require specialist input from subject communities (including teachers, professional bodies, employers and academics) at all stages of qualification development. Subject experts should be used throughout: setting the criteria, developing the assessments and accrediting specifications.
- **Be transparent** – The roles and responsibilities of any AO, the regulator and the subject communities must be clearly defined and transparent. Furthermore the system should engender supportive relationships between these groups. In addition, **all** AOs with responsibility for administering 15–19 examinations should be obliged to publish or otherwise make available anonymised, subject-based data on examination participation and performance in all national qualifications. This information would allow reliable assessment of how the examinations system in England is performing³.
- **Support innovation** – The examinations system should be responsive to and engaged in educational research to support suitably evidence-informed innovations in assessment, curriculum and pedagogy.
- **Ensure comparability between subjects** – A system must be committed to achieving parity of standards across subjects and specifications. Where this is not possible the system must be transparent and suitably acknowledge a lack of parity.
- **Offer real choice and quality** – Learners (or, more realistically, their institutions) must be offered a genuine educational choice in qualifications to accommodate and

² Bloom's Taxonomy is a hierarchy of learning objectives for education: the lower levels include recall, comprehension, application and the higher levels include analysis, synthesis and evaluation. Learning at the higher levels is dependent on having attained prerequisite knowledge and skills at lower levels, creating a deeper and more holistic form of learning.

³ Royal Society State of the Nation Report - [Science and mathematics education 14-19](#) (2008)

support different styles of learning. These qualifications must also lead to clearly differentiated career pathways.

- **Promote a cycle of evolution** – The system should operate on an evolutionary cycle where high quality and effective qualifications are continually improved, based on meaningful research and evidence so that each round of specification-development is informed by the successes (and failures) of the previous round. In addition, assessments and specifications should be piloted (with assessments being pre-tested to reduce errors) and the cycle of specifications should be long enough to allow any major changes to be based on evidence.
5. In addition, an effective examinations system must be considered within the context of the broader ecosystem. Awarding Organisation(s) do not exist in isolation. They have interdependent relationships with the regulator, subject communities, learning institutions and the Government and it is the effectiveness of these relationships that will determine the way a model operates in practice (see also paragraph 9).

Current system

6. While SCORE recognises the potential merits of multiple AOs (it spreads risk in the system, reduces the extent to which qualifications are under direct political control, presents diversity in qualifications, and potentially keeps costs down) the current model in which the AOs operate in England is not effective and we strongly believe that it jeopardises the needs of the learner, the consumer (HEIs and employers) and the country, by not assessing the specifications and thereby reducing the demand of what is taught.
7. The current system falls short on almost every point set out in paragraph 4, many of which are interrelated⁴:
- **Standards** – The commercial nature of AOs has led to an erosion of standards. Because it is a priority for AOs to maintain market share in qualifications they will never make a unilateral change to an assessment that makes it more difficult to achieve a high grade (or, put another way, reduce the number of high grades) – as most schools are unlikely to choose an AO that offers fewer high grades. This has led to a continual increase in the number of students getting the high grades. It is reasonable to assume that over time a number of the students who obtain the top grades would not have done so in the past. Individual cases and indeed year groups are difficult to compare, but the impression that standards are either slipping or becoming incomparable between year groups cannot be ignored. Of course it is a good thing for schools to aim to increase the number of their students achieving high grades; but the current system, in which a school's performance is measured mainly by the raw grades of its students, encourages them to connive in a broken market.

⁴ SCORE's comments on the current system are based on member organisations' collective experience of working with Ofqual, QCDA and the Awarding Organisations.

- **Engendering high quality teaching and learning** – The nature of AOs we believe has led to some decisions being made on commercial rather than educational grounds. These decisions have affected both the content of the specifications (chosen to be easily assessable) and the way in which they are assessed (tending to concentrate on the lower levels of Bloom’s taxonomy). The higher levels in the taxonomy (analysis, synthesis and evaluation) are rarely assessed. Attributes like curiosity, enthusiasm, imagination, persistence and teamwork are also relatively un-assessed; and therefore they are less likely to be taught. The effect has been to impoverish the learner experience by including a large number of knowledge-based statements in specifications and straight recall questions in examinations.⁵ Furthermore there is no regulation in place to prevent this from happening, as Ofqual is not required to review assessments prior to use. Rather than acting as an ‘air traffic controller’, preventing problems with examinations papers arising in the first place, Ofqual operates as a ‘crash scene investigator’.
- **Authentically represent the subject** – Multiple AOs producing multiple specifications for the same qualification in the same subject means that the expertise is spread thinly. It calls into question whether there are enough people who have sufficient subject and examining expertise and experience in each subject in each of five main AOs in England, Wales and Northern Ireland. Additionally, having multiple AOs makes it hard for professional bodies and the subject communities to take any role in specification development, as all must be treated equally. This lack of engagement with subject communities results in a lack of confidence from users of the system, including HEIs and employers. Furthermore, Ofqual is responsible for ensuring qualifications authentically represent a subject but, with little in-house subject expertise, it is hard to see on what grounds they can make this judgement (see paragraph below on transparency).
- **Transparency** – The role of the subject communities is not defined in the current system. The Criteria, to which all specifications must adhere for each subject, are set by Ofqual without in-house subject expertise and without formal engagement with the subject communities. Specifications are required to have received official support from their subject community before they are accredited. However, this official support is not defined and could come from any number of organisations whatever their expertise or professional standing.
- **Innovation** – SCORE believes the current model is not supportive of innovation. This is in part due to the competitive commercial nature of some AOs. Sharing best practice and collaborative working are not embraced and there is pressure from institutions (the customers) to minimise change to syllabuses and assessment methods.
- **Comparability** – The Criteria are produced by Ofqual to ensure comparability between specifications within the same subject. However, in reality, evidence shows

⁵ For example, evidence from SCORE commissioned research into the assessment of ‘How Science Works’ at Key Stage 4 found that many of the assessment items were low-level recall and few gave the opportunity for students to demonstrate higher level understanding [Andrew Hunt (2010) *Ideas and evidence in science: Lessons from assessment*].

substantial differences in how the Criteria are interpreted, particularly in terms of assessment. In recent SCORE reviews of GCSE⁶ and A-level⁷ examinations papers the type, the quantity and the difficulty of the mathematics assessed varied considerably across the five main AOs in England, Wales and Northern Ireland

- **Choice and quality** – The current model allows for choice. However, this is usually made on the basis of price (and sometimes the cost of a suite of qualifications) and the likelihood of students achieving high grades.
- **Evolution** – The current lifetime of specifications does not allow Ofqual or AOs to use evidence about the impact of the previous specifications when developing a new specification. The time constraints also adversely affect the piloting of new qualifications and the consultation process.

Alternative models to the current system and their potential

8. There are numerous models (and layers within models) for structuring AOs and the way that qualifications are provided. In appendix 1 we have attempted to summarise this in a diagram, highlighting the interactions between the different models and the ways in which qualifications are produced. It is important to note there is not a direct mapping between the structure and the offering to schools (it is possible, for example, to have multiple AOs with just one of them or all of them producing a given qualification). In Appendix 2 we have used this diagram to describe the potential advantages and disadvantages of the models.
9. In summary, based on the analysis set out in Appendix 2, SCORE sees very few advantages of providing the same qualification for a given subject, in competition, by multiple AOs. Although there are a number of risks, we would favour a model in which competition is not for market share within a qualification. While we have highlighted the potential risks and gains for these models there are a number of external factors that will affect how the examinations system operates in practice and SCORE strongly recommends the following factors are included within the remit of this inquiry:
 - a) **Status of an organisation** – An organisation’s status (e.g. charity, not-for-profit, commercial) will affect how they respond to the various incentives that any assessment system promotes.
 - b) **Definition and role of a subject community** – How these communities are defined and their subsequent role in qualification development will impact on the level of confidence in the system. There are likely to be different definitions of subject communities across the different subjects and across the different qualifications. For example, the strength and representation of professional bodies/learned societies vary across different subjects and it is not as simple as appointing the main subject association (in many cases there will not be one).
 - c) **Role of the regulator** – Whatever system is in place there is a need for some form of external regulation or scrutiny – via a national board, or ultimately

⁶ SCORE Report – [GCSE Science 2008 Examinations](#) (2009)

⁷ Preliminary findings from SCORE commissioned research into the assessment of mathematics in science A-levels. The final report is due for publication in Spring 2012.

Parliament itself. Who the regulator is will affect the system differently. For example a governmental regulator is likely to exert more direct political control, whereas subject community regulators will have the expertise to recognise authentic subject qualifications but may not be able to offer subject comparability and an independent body may not have the relevant expertise but be able to offer comparability. There is also the question of how the regulator is itself regulated and also the power the regulator is able to exert over the system.

- d) **Geographical remit** – The current model of qualification development allows specifications developed by AOs in Northern Ireland and Wales to be used in English learning institutions (and vice versa the specifications developed by AOs in England are available to Northern Irish and Welsh learning institutions). Will (and can) this still be a requirement with increasing divergence between the nations?

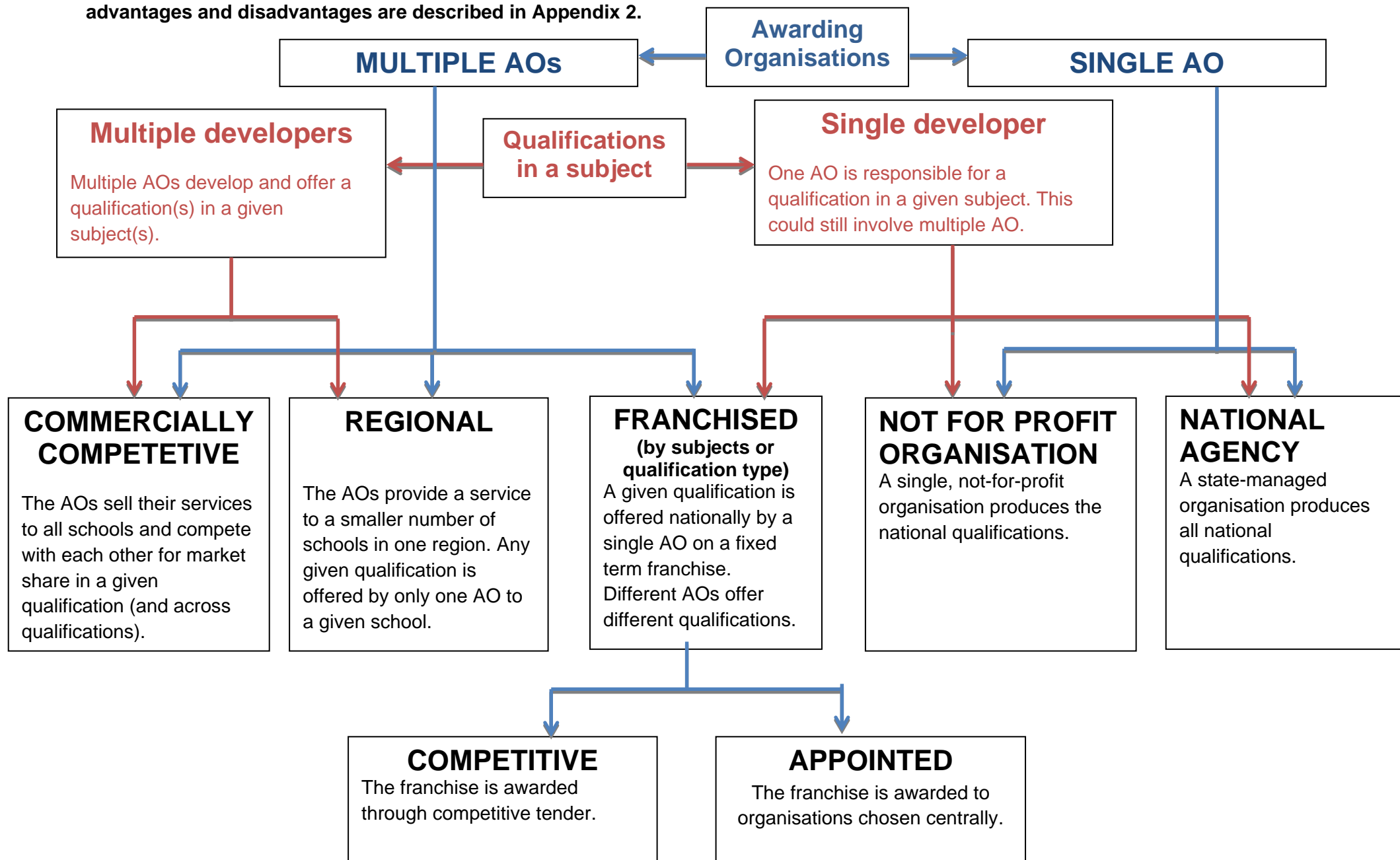
Ensuring accuracy in setting papers, marking scripts, and awarding grades

10. SCORE urges the Committee to consider the quality of examinations papers as well as their accuracy. The poor quality of assessment items degrades the curriculum through wash-back. Increasingly, assessments tend to be in the form of written examinations with items that test what is easy to assess. They concentrate on the lower level domains in Bloom's taxonomy: recall, comprehension and application. Consequently, there is an over-emphasis on these skills in the way the subject is taught. This has a damaging effect on the learner experience because teachers will tend to emphasise the content and techniques that they know are likely to come up in examinations.
11. In paragraph 7 SCORE refers to Ofqual as a 'crash scene investigator' rather than an 'air traffic controller'. The regulator should be responsible for preventing problems arising with examinations in the first place and it is hard to understand how Ofqual can accredit specifications without taking into account the accompanying assessment tools.
12. SCORE therefore calls for there to be a formal requirement of the regulator to review assessment prior to use. The regulator should undertake this review with an expert panel comprising of subject and assessment expertise. Professional bodies and subject associations should be involved in the process – either through direct involvement or through proposing members of subject review panels. This would ensure the appropriate level of demand is demonstrated in all assessment materials and that there is a comparable standard of assessment across equivalent qualifications.
13. To ensure assessment is not dictated by commercial forces the Select Committee may wish to consider a model where one national organisation develops a bank of trialled, quality examination questions. Such a body would have a permanent team of recognised subject experts and subject assessment experts and could exist in a model with more than one AO.

Commercial activities of Awarding Organisations, including examination fees and textbooks, and their impact on schools and learners

14. SCORE is very concerned about the management of conflicts of interest between awarding functions and any other activities AOs (and their related companies) undertake. Good specifications should support effective teaching, learning and assessment, without being influenced or constrained by commercial interests and/or connected activities.
15. In 2010, some AOs marketed unaccredited GCSE science qualifications to schools in order to capitalise the market, leading to possible confusion as to their status. The regulator must have the power to ensure that for those qualifications that need to be accredited, specifications can be marketed only after accreditation.
16. Different qualifications (e.g. GCSEs in English and Science) can and are grouped together by AOs in package deals for centres. The cost of such a grouping, rather than the quality of a particular qualification within it, can affect a centre's choices. There should be regulation to ensure that pricing structures of individual qualifications, and packages, are fair.
17. The relationship between AOs and publishers must be carefully/strictly monitored. This relationship could mean that 'preferred' published resources are pushed in the direction of teachers, even if they are unsuitable for developing a depth of understanding of a subject. There are also issues with Chief Examiners writing text books as there is a perception amongst teachers that these books may contain 'insider information' and this could be seen to be a conflict of interest in terms of the Chief Examiners' role. This is potentially damaging to the teaching and learning of science.
18. AO endorsement of text books means that textbooks are very tightly matched to specific specifications and their associated examinations. Schools usually feel it is necessary to replace entire sets of text books if they changed specifications. This may lead to reluctance to change specifications for financial reasons. Additionally, the relationship between the specifications being developed by AOs and the production of textbooks that support those specifications can, as was experienced in 2010, lead to unresponsiveness by AOs to feedback on their proposed specifications. We strongly recommend that this link is broken between specification/assessment development and the commercial publications that provide resources in support of a specification.
19. The Select Committee should consider recommending the following restrictions on AOs to break the link between them and publishers:
 - A restriction on AOs talking in detail to publishers until after specifications have been accredited – the date for introduction/first teaching would need to be extended by 1 year;
 - A restriction on AOs (or the Department for Education) endorsing particular text books – moving towards more general text books for GCSE science;
 - A restriction on AOs being owned by publishers where there is clear evidence of this having too much influence on qualifications development and the outcome of examinations;
 - A restriction on AOs and current examiners writing textbooks, for instance preventing them both from releasing information about any mutual affiliation they may have.

Appendix 1: A diagram summarising a number of different model for structuring Awarding Organisations (AO) and the way that qualifications are provided. This diagram does not show any one model to be more successful than another and a full break down of advantages and disadvantages are described in Appendix 2.



Appendix 2: A description of the potential advantages and disadvantages of the various models, as set out in Appendix 1, for structuring Awarding Organisations (AOs) and the way that qualifications are provided.

1. Multiple developers versus a single developer of a qualification in any subject:

- A model which includes multiple developers of a qualification in any given subject has the potential to offer schools (although not necessarily learners) more choice. It also spreads the risk in the system and avoids single points of failure. However, the model requires additional regulation (including the development of criteria) to ensure parity. It presents difficulties for subject communities to engage with multiple bodies. Furthermore the nation's assessment expertise is spread thinly. This model also increases the length of the development cycle and there is a duplication of effort which in return increases the cost to the school (no advantage of scale).
- Conversely, a model in which a single body develops specifications for a given qualification offers many advantages (and could still exist with multiple AOs). Subject expertise can be concentrated to ensure the best teams develop qualifications. It also allows for effective engagement with the subject community as efforts can be focused in one qualification. This model allows for greater links to be forged with educational and assessment research to help facilitate innovation. Regulation would be easier; it can be based on quality of the qualification and its assessment tools rather than on comparability. Furthermore there are no destructive drivers (such as commercial competition for the market share) on standards and quality. On the downside, by not spreading the risk across multiple developers this model allows for a single point of failure to occur (although whatever failures do occur will affect all learners so no one will be any more disadvantaged). There is also no immediate market incentive to maintain quality and there is the possibility that qualifications might stagnate – though this could be overcome by working with not-for-profit organisations where the interest in the quality of a qualification is intrinsic and goes beyond commercial considerations. Such a model will involve a large number of candidates and could prove difficult to run logistically.

2. Multiple AOs versus a single AO:

- Multiple AOs have the potential to offer a range of supporting resources. The competition between AOs may also prevent stagnation in the system. It also spreads the risk across the system. Such a system does however require a regulatory framework and risks the comparability between qualifications. Whilst qualifications have not stagnated in the current system, there have been very few changes for the better in the assessment tools. See paragraph 3 for model options within multiple AOs.
- A single AO would remove risks of comparability between different specifications within a qualification. Such a model would be likely to increase cost-effectiveness and also reduce duplication of effort. Market pressures would not

exist and, as with a single developer of a qualification, it would be easier for professional bodies/learned societies and the wider subject community to engage with qualification development. Potential risks include stagnation and a reduced emphasis on innovation, a lack of choice and potentially a large number of candidates. See paragraph 4 for model options within a single body.

3. Commercially competitive AOs versus regional AOs:

- Commercial competition provides an incentive for AOs to keep the costs to schools down. However, market pressures encourage the system to focus on costs and accessibility rather than standards, the quality of examinations, its assessment tools or the learning it engenders. Qualifications are offered in suites with some included as loss-leaders. Furthermore, the regulation in place has to take account of commercial sensitivities weakening its power as a regulator – it would be risky for the regulator to make statements that might damage market share.
- A model using regional AOs offers the potential for qualifications to be tailored to regional educational resources. It does carry substantial risks particularly on regional comparability and routes for progression across the country; it offers a lack of choice and has the potential to create regional differences. It would also be difficult to implement as more schools become independent of any local or regional control. Furthermore this divisive model could cause problems in applying to universities (if a whole region is favoured or not) and when moving schools. SCORE does not see this as a realistic option.

4. Centralised state body versus not-for-profit single organisation versus franchised system:

- A centralised state body removes 'market pressures' completely and potentially allows a direct focus on standards. However, it does present concern on who would regulate the state. There may also be implications from having direct political control on qualification development (e.g. introduces party ideology into assessment system) and the potential changes in 'ethos' on the cycle of elections could affect the stability of any such model.
- A not-for-profit organisation again removes market pressures completely but, as with a centralised body, raises concern on who will examine the examiner. One possibility could be to set up a system of peer review as a form of regulation. This could also include a stakeholder review or a steering group (convened by a Professional Body where they exist for a subject or where there isn't one, made up of people with real knowledge and understanding of the qualification).
- Some of the risks above might be mitigated by developing a system in which there are multiple AOs but only one holds the franchise for a qualification or suite of qualifications. Under this model market pressures have the potential to drive up standards as the franchise (and commercial return) would be awarded for excellence of the qualification and assessment tools. The requirements of a franchise could also drive innovation by stipulating the need to develop more than one version of the qualification (i.e. the 'B' specification would not reduce market

share for the AO). Stagnation or continual change is, however, a risk depending on the franchise period. It also raises questions on who should have the responsibility for selecting particular franchises. Furthermore a school would have to deal with a number of different AOs (unless all the administration is centralised). Paragraph 5 outlines the models available within a franchise system.

5. Competitive franchise versus appointed franchise:

- A competitive franchise potentially takes out 'market pressures' with a focus on creating high quality qualifications. There would however be upheaval whenever a franchise holder is changed. There may be an advantage to the existing franchise holder as they will tend to have greater subject expertise and experience. There is also the possibility that different qualifications will cost different amounts (bigger entries etc) and it may not be within an organisation's interest to develop a more expensive qualification which offers little return. Another risk is there may be no bids for minority qualifications which could result in a loss of potentially good qualifications.
- An appointed franchise system is harder to rationalise as it removes competition entirely and begs the questions on what basis organisations would be selected.