

# Modified Question Papers

The Future of Language Accessibility in the UK



October 2010

Ofqual/10/4749

# Contents

1. Introduction.....	2
2. Background .....	4
3. Theories of assessment.....	6
Fairness reviews .....	6
4. Research design.....	9
5. Ofqual Paper Modification Forum .....	11
The first discussion group .....	11
Diversity and inclusion.....	11
Fairness and validity.....	12
High level principles .....	13
Development and production of modified assessment .....	14
Monitoring of paper modification .....	16
The second discussion group .....	17
Risk-assessment strategies.....	18
The third discussion group .....	20
Modified language papers: For all or for none? .....	20
The Support Model.....	24
New rules on language accessibility.....	28
6. Conclusion.....	31
7. References .....	33

## 1. Introduction

This paper aims to support good assessment practice arising from changes to the government's education policy.<sup>1</sup> These changes are likely to result in more 'smart regulation' – a different form of regulation that's based on the cooperation between assessment providers and the regulators.<sup>2</sup> Smart regulation could set the tone for a new kind of guidance on the rules for the language of assessment.

The paper reports on the various stages of the Ofqual Paper Modification Forum (OPMF), which led to the proposal for non-statutory guidance on language accessibility. The objective is to gauge views on, and voluntary support for, the future Ofqual Language Accessibility Online Working Group that's dedicated to unpack the concept of 'appropriate language' for assessment. One of the aims of this group will be to produce guidance on how to control unnecessary linguistic complexity of content-based assessment items, which are regulated by the *National Curriculum Assessments: Code of Practice 2010* (Ofqual, 2010), the *GCSE, GCE, Principal Learning and Project Code of Practice* (Ofqual, 2010), and the *NVQ Code of Practice* (Ofqual, 2006).

In early 2011, we plan to publish non-statutory guides on appropriate language in item development, which may be referenced in revised versions of the *National Curriculum Assessments: Regulatory Framework*,<sup>3</sup> to be published following the Apprenticeships, Skills, Children and Learning (ASCL) Act, and *GCSE, GCE and Principal Learning and Project Code of Practice*. The non-statutory guides will assist test developers in controlling the linguistic source of construct-irrelevant variance<sup>4</sup> that may threaten the validity of assessment. The guidance should also help test developers to move away from how language modification is currently undertaken – non-standardised and very costly. In future, item developers could produce assessments that are free from linguistic features that affect comprehension by all test takers referred to in the Disability and Discrimination Act.

---

<sup>1</sup> The full version of this paper can be found on the Office of Qualifications and Examinations Regulation (Ofqual) website – [www.ofqual.gov.uk](http://www.ofqual.gov.uk).

<sup>2</sup> Where GCSEs and GCEs are concerned, there needs to be involvement from all regulators – Ofqual, the Council for the Curriculum, Examinations and Assessment (CCEA) and the Department for Children, Education, Lifelong Learning and Skills (DCELLS).

<sup>3</sup> Current version available at [www.ofqual.gov.uk/component/search/National+Curriculum+Assessments%3A+Regulatory+Framework/%2F?ordering=&searchphrase=all](http://www.ofqual.gov.uk/component/search/National+Curriculum+Assessments%3A+Regulatory+Framework/%2F?ordering=&searchphrase=all)

<sup>4</sup> See Haladyna, T.M. and Downing, S.M. (2004) "Construct-Irrelevant Variance in High Stakes Testing", *Educational Measurement: Issues and Practice*, 23(1), pp. 17–27. See also [www.ericdigests.org/2000-3/validity.htm](http://www.ericdigests.org/2000-3/validity.htm)

For the OPMF, we gathered information from presentations by the Qualifications and Curriculum Development Agency (QCDA) and interviewed awarding organisation representatives from England, Wales, Northern Ireland and Scotland. We asked whether the processes for language modification to the assessments in question were standardised, and whether there were enough qualified modifiers to review question papers.

The next sections aim to present detailed accounts of the OPMF, and explain the practices that awarding organisations and QCDA used until the beginning of 2010 to modify the language of the National Curriculum assessments, GCE, GCSE and NVQ question papers.

Section two looks at the statistics compiled by Ofqual on access arrangements and special consideration at GCSE and A level.

Section three establishes the theoretical foundations. The theoretical background comes from the disciplinary field of psychology, more specifically, the principle of fairness in test item development.

Sections four and five present the research design for the OPMF, and the results of the three discussion groups.

The conclusion explains why and how we will need to work towards consistency in language modification and modification for the visually impaired.

## 2. Background

According to Ofqual's *Statistics for Access Arrangements and Special Consideration at GCSE and A level: 2008* (Ofqual, 2009),<sup>5</sup> awarding organisations offered over 21,000 modified question papers for the June examination series. Modified question papers allowed candidates with a range of visual impairments and significant language comprehension disorders to demonstrate their abilities.

In 2008, the larger awarding organisations in England modified 21,195 question papers of which only 2,549 were language modified question papers. The large awarding organisations explained that language modified papers were issued on demand. Awarding organisations did not have the personnel to sit at all question paper evaluation committees and propose language modification 'at source'.

Table 1: Modified question papers in 2008

	AQA	Edexcel	OCR
Number of modified question papers	9,248	3,936	8,011
Braille question paper	647	219	529
Examination on coloured paper	0	0	70
Enlarged paper (18 and 24 point bold)	6,304	2,677	4,510
Modified language	454	93	1455
Tactile diagrams	48	15	48
Unmodified A3 QP	1,795	932	1,399

---

<sup>5</sup> Available at [www.ofqual.gov.uk/component/search/Statistics+for+Access+Arrangements+and+Special+Consideration/%2F?ordering=newest&searchphrase=all](http://www.ofqual.gov.uk/component/search/Statistics+for+Access+Arrangements+and+Special+Consideration/%2F?ordering=newest&searchphrase=all)

In the past 20 years, most awarding organisations have used the British Association of Teachers of the Deaf (BATOD) rules on language modification for the deaf. The rules published by BATOD/National Association for Tertiary Education for Deaf People (NATED) were specifically designed for candidates with hearing impairment and seem inadequate for the total number of candidates with a range of speech and language difficulties (SLDs), including dyslexia.

The BATOD guidelines were last revised in 2003 and are still the only guidelines available for language modification in 2010. The guidelines outline the grammatical features of written language likely to prove difficult for deaf learners to understand, and are useful for training mainstream staff and those working with deaf learners. However, with the enforcement of the Disability and Discrimination Act, language modification should meet the needs of many other candidates, not only those with hearing impairment. Organisations therefore need written procedures for language modification.

According to some representatives of awarding organisations, question paper modification for the blind seem to be more consistent in comparison to modification for candidates with hearing impairment that depend on the opinions of professionals, which are not always consistent. The outcomes of language modification for candidates with hearing impairments and SLDs are at times very different, and the rephrasing of questions depends on the individuals proposing the modifications.

Because modifications proposed by two different modifiers are rarely the same, language modification could be judged as unreliable. We will need to work towards a set of objective rules to increase consistency across modification professionals. Some stakeholders suggested that the plain English rules could bring some consistency into language modification.

### **3. Theories of assessment**

From a theoretical viewpoint, educational assessments must rely on a statement of purposes or specifications, which should be consistent with the construct or domain of knowledge to be assessed. A sound assessment framework will spell out the content, skills, processes and diagnostic features of the construct to be measured. Constructs represent a subset of the knowledge to be assessed, such as numbers, geometry and statistics. They also elicit the cognitive processes needed to complete an assessment that requires conceptual understanding, procedural knowledge or problem solving (Linn, 2006: 30). An assessment construct is therefore the logical starting point for developing valid educational assessments that identify the content domain and the cognitive processes used by candidates (American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME), 1990: 18–19).

Messick (1984) was of the view that educational achievement assessments reflected not only the psychological constructs of knowledge and skills that were intended to be measured, but invariably “a number of contaminants”. These polluting influences included “a variety of other psychological and situational factors that technically constituted either construct-irrelevant assessment difficulty or construct-irrelevant contamination in score interpretation” (Messick, 1984). Construct-irrelevant variance (CIV) is defined as error variance that arises from systematic error. A useful way to understand systematic error would be to compare it with random error, or the difference between any observed and corresponding true score for each examinee (Haladyna and Downing, 2004).

#### **Fairness reviews**

Since the 1960s, fairness review became a major tool in making fair assessments while finding construct-irrelevant variance sources of difficulty that may interfere with the validity of assessment to the diverse groups of test takers. In the 1970s, psychometricians developed precise methods for studying bias. Later on, Angoff (1993) described a methodology to help identify items that could be unfair.

By the 1990s, fairness reviews were widespread and an expected feature of professional assessment development. Fairness reviews intend to identify and remove invalid aspects of assessment questions that might hinder people in various groups from performing at levels that allow appropriate inferences about their relevant knowledge and skills (ETS, 2009).

Some test development agencies require attention to fairness throughout the life cycle of an assessment. Assessment programmes in the US are generally asked to demonstrate that reasonably anticipated areas of unfairness could be addressed as the assessment was designed, developed, administered and scored, and as the results were used. In addition to fairness review, some test development agencies use differential item (or test question) functioning (DIF) statistics, validation of assessments, and the provision of information about appropriate assessment interpretation and use.

The AERA, APA and NCME observed in their publication that there were different definitions of 'fairness', and diverse interpretations in different social and political circumstances (1999: 73–80).

Zieky (2006: 359) discussed fairness review guidelines in terms of the:

- Maximisation of respect
- Control of the effects of construct-irrelevant knowledge or skills
- Avoidance of controversial materials
- Avoidance of stereotypes
- Respect for diversity in depictions of people.

In 2002, the National Centre on Educational Outcomes (NCEO) began to publish reports on the notion of 'universal design'. This concept came from the field of architecture and quickly expanded to education. In the US, the Center for Universal Design came up with seven elements of universally designed assessments.

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, non-biased items
4. Amendable to accommodations
5. Simple, clear and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

The above elements of universally designed assessments show the extent to which fairness and validity are closely intertwined. In fact, the validity of assessment is

threatened when CIV impacts differently on the performance of subgroups (Abedi, 2006).

The linguistic complexity of assessment items and its impact on the test taker's assessment outcome is one of the most common sources of CIV. Some linguistic features are said to affect comprehension and these include:

- Word length
- Sentence length
- Voice of verb phrase
- Length of noun phrases
- Complex question phrases
- Comparative structures
- Propositional phrases
- Sentence and discourse structure
- Subordinate clauses
- Conditional phrases
- Relative clauses
- Abstract or impersonal presentations
- Negation.

However, identifying potentially problematic linguistic features in assessment is a never-ending task, which requires:

- Linguistic and content experts
- Guides on linguistic modification
- Research on linguistic aspects likely to cause problems to specific groups of test takers.

## **4. Research design**

From October 2009 to April 2010 the Assessment Research Team at Ofqual organised three discussion groups, which formed the OPMF. The discussion groups involved assessment professionals from within the organisation and external stakeholders sharing special interest in test development and language modification. The OPMF adopted the broad definition of modified papers established by the Joint Council for Qualifications (JCQ, 2008):

The modification of carrier language in all question papers, so that the standard papers should be suitable for all candidates with substantial comprehension difficulties, irrespective of the reason for the impairment.

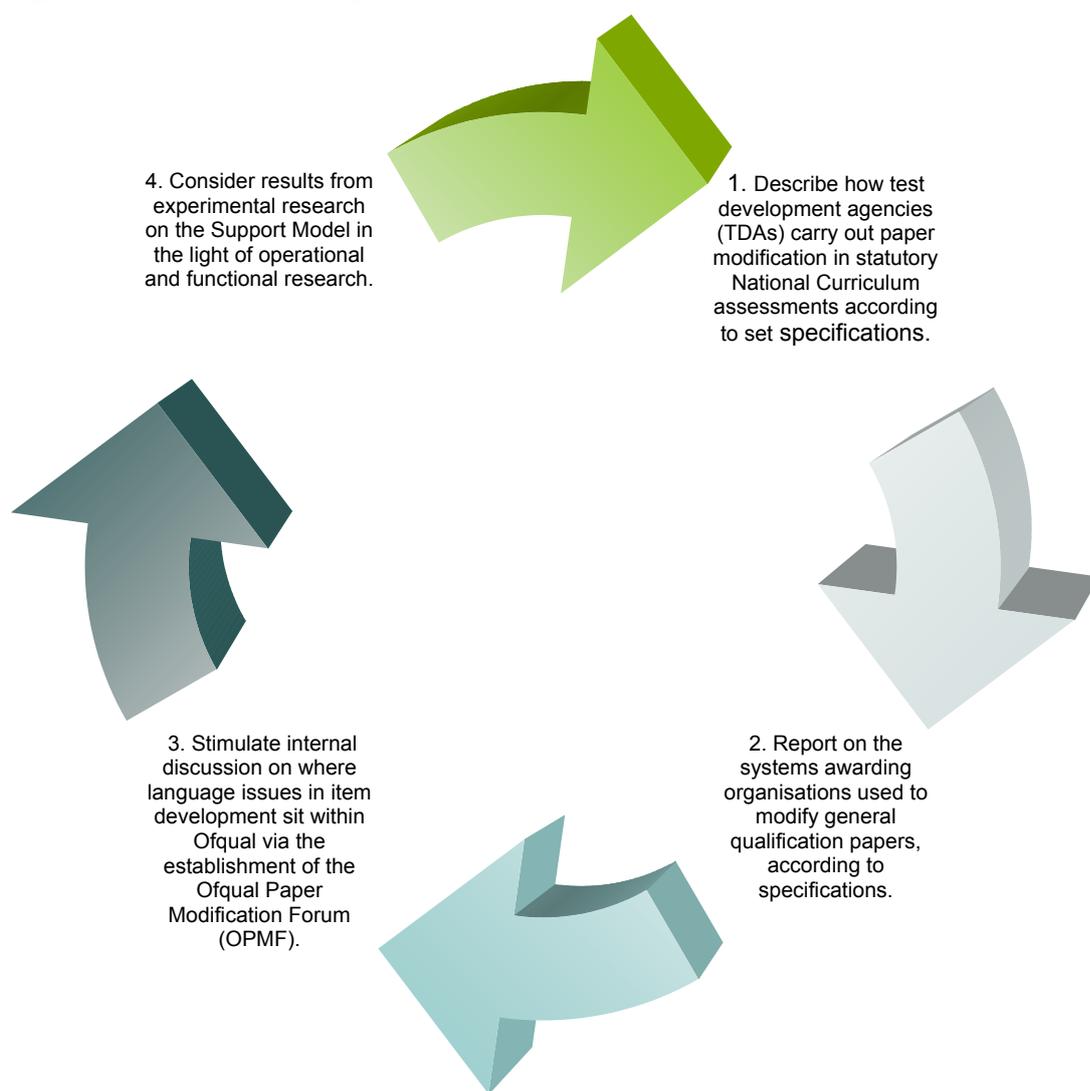
The views gathered during these discussions touched on organisational, experimental and operational issues affecting the language of all the assessments and qualifications regulated by Ofqual. At different points in time, discussions focused on the National Curriculum assessments and nationally recognised qualifications such as GCSEs, A levels, the Diploma and NVQs.

The aims of the discussion groups were to equip Ofqual to deal with the challenges surrounding the use of modified question papers in assessments and the needs to:

- Produce a guide on language accessibility that would define the principles, methodology and standards that test development agencies and awarding organisations would agree to meet in producing examinations
- Ensure the public that Ofqual's diversity and inclusion policies do not clash with its work to maintain standards for qualifications and assessments, and that they consistently aim to reduce construct-irrelevant barriers to test performance
- Assess operational issues involving assessment design, score interpretation and the monitoring of assessment modifications, so that the Ofqual Management Group can identify where paper modification sits within the organisation.

The basic research design for the OPMF can be seen in the diagram below. The outcome of this would be to combine the functional, operational and experimental data to explain how the regulators could support a new model of valid and fair assessment, irrespective of the level of language difficulty in an assessment item. The research results could give rise to a guide on language accessibility, which will be used as an appendix to the current codes of practice.

Figure 2: Research design for the OPMF



## 5. Ofqual Paper Modification Forum

### The first discussion group

The first discussion group of the OPMF took place on 26th October 2009 in Coventry. In attendance were members from the Standards and Policy and Regulation Teams at Ofqual.

Questions addressed at the first discussion group were:

- Where does paper modification sit within the regulators' structure?
- Should Ofqual, CCEA and DCELLS monitor paper modification in future?
- Should the regulators produce a code of good practice on the language of assessments?

The conclusions reached by the discussion group would help Ofqual solve current operational uncertainties, for example where paper modification sits within the regulators' remit, and how paper modification is related to 'systems' and 'qualifications issues'.

### Diversity and inclusion

The first discussion group included a presentation on diversity and inclusion. At the time, the regulators were working towards embedding fairness principles into all aspects of the regulations work they undertake. Key issues were the:

- Existing guidelines for fairness review in test development
- Concept of fairness as validity.

Among the points raised was whether the criteria in certain assessments failed to target the desired construct (or the ability that was being assessed), and in this way, unwittingly introduced factors that were a real challenge for the regulators. For example, **giving marks in GCSE mathematics for general knowledge when it was not the general knowledge that was being assessed**. If the notion of fairness as validity is adopted, valid assessments should assess a desired construct or ability.

There was doubt among participants on how the documents *Fair Access by Design: Guidance for Awarding Bodies and Qualifications Regulators on Designing Inclusive General Qualifications* (Ofqual, DCELLS and CCEA, 2009) and *Fair Access by Design: Guidance for Awarding Bodies and Regulatory Authorities on Designing Inclusive Vocational Qualifications* (QCA, ACCAC and CCEA, 2006) fitted into paper modification. The documents provide guidance for awarding organisations and regulatory authorities on designing inclusive GCSE, GCE and vocational qualifications. They do not provide guidance on language modification or making

reasonable adjustments<sup>8</sup> for candidates who have particular requirements. Rather, their focus is on designing qualifications that are as accessible as possible.

Guidance in *Fair Access by Design* on GCSEs and GCEs include the language and syntax of questions and rubric, which state that they “must be easy to understand”, “papers must be readable”, and “even complex concepts and instructions can be communicated in simple language without compromising standards” (Ofqual, DCELLS and CCEA, 2009: 10). In relation to vocational qualifications, written questions used in examination conditions must “ensure the level of language used matches the level of the qualification” (QCA, ACCAC and CCEA, 2006: 19).

The regulators can only monitor fairness of assessment design if major points on qualifications design become statutory and not just guidance. However, this would lead to the regulators being perceived as too intrusive and therefore more information is needed on the subject.

The meeting organisers had consulted with language assessment specialists who suggested that the best way to monitor fairness of the language of assessment would be through the monitoring of professional assessment modifiers who are affiliated to an association in charge of maintaining the standards of language modification in all assessments, examinations and qualifications in the UK. During the first discussion group, several other ideas came to play. The monitoring of professional assessment and the impact of modification on the fairness of assessment for all test takers were further debated.

### **Fairness and validity**

Another point of discussion was whether we should define fairness as validity. Some of the participants at the first OPFM suggested that greater stress on fairness and too much modification could lead to a ‘devalidation’ of assessment. Extensive alteration could lead to the paper becoming too neutral. By trying to become too meaningful or relevant for all test takers, assessments could end up becoming meaningless for many.

Assessments could become over-modified to the extent that interference could impact on validity. Guidance principles could be introduced to avoid bias rather than

---

<sup>8</sup> Reasonable adjustments are “arrangements which are provided in advance of an examination or assessment to allow attainment to be demonstrated by candidates with either a permanent or long-term disability or learning difficulty, or a temporary disability, illness or indisposition” (QCA, ACCAC and CCEA (2006) *Fair Access by Design: Guidance for Awarding Bodies and Regulatory Authorities on Designing Inclusive Vocational Qualifications*, page 35).

[www.rewardinglearning.org.uk/docs/regulation/fair\\_access\\_final100406.pdf](http://www.rewardinglearning.org.uk/docs/regulation/fair_access_final100406.pdf)

over-modify an assessment. Therefore fairness should be considered on a practical level as well as in terms of modification.

There was doubt among participants as to whether an assessment could be fair, but not valid, or whether an assessment could be valid, but not fair. To a certain extent, one does not mean the other. There is a need to consider bias as well as modification – modification should not be considered on its own. The place of modification should at least be to make clear to the learner what is being asked and answered.

Paper modification could have a positive impact on construct relevance and assessment appearance. Yet mark schemes should be tight and tied to construct and response. In the view of one participant, we should not define fairness as validity when considering National Curriculum assessments. The group suggested that perhaps the concept of fairness should be applied to general qualifications only, and rather than modifications for individuals, all papers should be made accessible.

There was no doubt among participants about the need for 'fairness' when papers are being developed, but some participants were unsure whether validity in itself was enough to define fairness. There seem to be a number of other issues under the fairness umbrella. Participants were interested in looking at other documents that are used to inform the development of assessments.

### **High level principles**

In considering how Ofqual fairness review guidelines should apply to the National Curriculum assessments and GCSEs, it was evident that the regulators should promote learners' access to fair assessment and that their position should not be too prescriptive. The regulators need high level principles as guidance to promote the place of modification and handle the elimination of bias as much as possible, considering both internal and external assessments.

Given the basic role of language in assessment, the regulators should consider the full range of regulated qualifications. A possible strategy would be to look at integrating a rule (not prescriptive or detailed) at a high level to require awarding

---

<sup>9</sup> Group referred to ETS (2009) *ETS Guidelines for Fairness Review of Assessments* – [www.ets.org/Media/About\\_ETTS/pdf/overview.pdf](http://www.ets.org/Media/About_ETTS/pdf/overview.pdf). The relevant documents in the UK are Ofqual, DCELLS and CCEA (2009) *Fair Access by Design: Guidance for Awarding Bodies and Qualifications Regulators on Designing Inclusive General Qualifications* – [www.rewardinglearning.org.uk/docs/regulation/fair\\_access\\_final\\_gcse\\_gce.pdf](http://www.rewardinglearning.org.uk/docs/regulation/fair_access_final_gcse_gce.pdf) – and QCA, ACCAC and CCEA (2006) *Fair Access by Design: Guidance for Awarding Bodies and Regulatory Authorities on Designing Inclusive Vocational Qualifications* – [www.rewardinglearning.org.uk/docs/regulation/fair\\_access\\_final100406.pdf](http://www.rewardinglearning.org.uk/docs/regulation/fair_access_final100406.pdf).

organisations to give modification and bias in assessment due consideration. Some participants at the first discussion group did not recommend that the regulators promote a code of practice, as they felt it would be too prescriptive.

Participants identified a conflict between the regulators' remit to encourage innovation in assessment and their duty to produce sound guidelines for awarding organisations that wanted certainty.

There should definitely be a very high level of requirement for fairness when assessments and examinations are developed. However, the first discussion group was reluctant for Ofqual, CCEA and DCELLS to develop a detailed document such as the Educational Testing Services' (ETS) *Guidelines for Fairness Review of Assessments* (ETS, 2009). It was felt that this should be part of quality assurance and not for the regulators to get involved with in great detail. The regulators will need to know that procedures are in place, but will not need to set down detailed rules.

### **Development and production of modified assessment**

Participants discussed whether the regulators should engage with the development and production of modified assessment. After a presentation on the development process of National Curriculum assessments, one participant concluded that Ofqual should not get involved with the development and production of 'modified assessment' (new terminology suggested by the group), but should again tackle the issue from a much higher perspective of the principles and outcomes of paper modification.

It was suggested that professionals working within standards, monitoring, policy and regulation at Ofqual should make sure that they know more about the initial stages of the process and master the principles of paper modification. We should also carry out further research into the end of the process – the outcomes of paper modification and their impact on learners. That is to say, the regulators should only deal with the beginning and the end of the paper modification process.

During the next National Curriculum assessment season in 2010, the first discussion group put forward that we should carry out a survey with readers and amanuensis, asking the following question: Do you think assessment adaptations allowed learners to show their knowledge and skills in the best possible light? We should also find out more about 'expert advice' during the development process of National Curriculum assessments, asking the following types of questions:

- How do you use expert advice to amend assessment items?
- Who attends the expert advice meetings?
- What happens with comments made by expert advice groups?

- How are suggested changes to standard assessments incorporated into the modified versions of the National Curriculum assessments?

In order to answer these questions, we will need to know which guidelines test development agencies use to modify National Curriculum assessments.

From a regulatory policy angle, we need to review all aspects of assessment development to guarantee valid and reliable assessments. A key question would be: What resources will Ofqual need to review all those aspects? The numbers of learners sitting the National Curriculum assessments who also have statements of special educational needs (SEN) will need to be established. The regulators will need the annual figures published in Ofqual's *Statistical Bulletin: Access Arrangements for GCSE and GCE: June 2009 Examination Series* (Ofqual/10/4700) to distinguish cohorts and compare them.

Vocational qualifications were seen as an interesting case because competence standards cannot be adjusted in vocational qualifications, but methods of assessment can be – so long as the assessment still tests the competence standards.

In relation to assessment modification within the regulators' remit, a balanced approach to assessment modification within Ofqual, CCEA and DCELLS should separate the principles of assessment modification from its outcomes. The outcomes of paper modification should be followed by professionals responsible for maintaining the standards of assessment, but the principles of assessment were certainly a matter for the regulatory organisations as a whole. The regulators' CEOs should own the principles of assessment modification because the value of fairness would sit at the highest level within the organisations. To show how much relevance the issue of language accessibility had within Ofqual, CCEA and DCELLS, the principles of assessment would need to be discussed at a corporate level.

However, there will be the consequences of assessment modification on areas such as manageability, scoring, reliability of assessment results, comparability of assessments, and so on. These practical outcomes of assessment modification should be dealt with by those engaged with the National Curriculum assessments, awarding organisations and those responsible for qualifications. All those involved would need to develop a methodology to judge the outcomes of assessment modification. The responsibility towards for example assessment modification at Key Stages 2 and 3 would lie within a very specific area of work within Ofqual.

If there is a case for visible monitoring of the development and implementation of modified assessments, this work will have to be carried out by professionals involved with National Curriculum assessments, and by those working directly with general qualifications monitoring. A skills gap affecting the regulators could arise and training

of additional staff to carry out the duties related to assessment modification will be needed to resolve the issue. Therefore, one of the main problems with monitoring the language of assessments will be related to costs. A sustainable approach to assessment modification will be sought.

### **Monitoring of paper modification**

Opinion about the need to monitor paper modification was highly divided among the first discussion group, but there was more support for monitoring than not. Some participants judged that in future, Ofqual, CCEA and DCELLS should monitor paper modification because the regulators should be seen as upholding the values of inclusiveness.

From a communications strategy perspective, the regulators will first establish the current picture of assessment modification through a benchmark exercise and decide whether there is a case for monitoring test development and assessment modification. If there is evidence that things are going well, the regulators shall not interfere with the current ways of modifying assessment. The regulators do not want to stray into delivery areas and will avoid crossing the barriers of their remit. However, if evidence suggests that things are not going well, the regulators will have a stronger case for visible monitoring.

More cautious, other participants were of the opinion that the regulators should not monitor assessment modification and that organisations should be confident that there is an adequate system in place, and trust that such a system will be followed.

For those in favour of the regulatory monitoring of paper modification, Ofqual, CCEA and DCELLS can only support future actions to control unfairness in assessment if they are able to produce common criteria for good practice on minimising bias across all assessments, examinations and qualifications. Such common set of criteria should guarantee consistency of standards over time.

In relation to the development of assessment for the National Curriculum assessments in England for disability groups, it was suggested that QCDA should have a separate programme on assessment modification alongside the standard assessments. We need to know more about QCDA's Test Review Group and the Teachers' Panels, which are designed for the review of National Curriculum assessment materials, and increase monitoring of these reviews to embed diversity and equality principles.

As Ofqual does not have an expert on assessment modification internally, the regulators will need to rely on external expertise to be able to monitor assessment modification of GCSE and GCEs with competence. It was recognised that there needs to be a system in place to support awarding organisations' actions in future.

The regulators are to think carefully about what they are trying to achieve with more monitoring.

In 2009, it was not clear who was responsible for the monitoring of modified assessment development. It is therefore crucial to establish consistency of monitoring across the National Curriculum assessments, GCSEs and GCEs.

### **The second discussion group**

The second discussion group of the OPMF included a joint presentation by QCDA's Director of Test Development and a representative of a modified test agency on *The Principles and Outcomes of Modifying the National Curriculum Tests*. The presentation covered details of access arrangements, ranging from scribes, Braille, enlarged print, oral/language modifier, modified large print, additional time, mother tongue translation, dictionaries, individual rooms, live speakers, prompter, tape recorders, readers, coloured paper/ink, reading cards, word processors, supervised rest breaks, early openings, modified language papers and mark scheme amendments.

It became clear that some modifications were only applicable to GCSEs, some to National Curriculum assessments and some to both. In the case of National Curriculum assessments, the access arrangements should reflect the classroom practice and not compromise what is being assessed. The group analysed examples of National Curriculum assessment questions and discussed how they could be modified.

In light of the practice of the National Curriculum assessment modification process, there was doubt on whether changes to question wording, question paper layout, content and the mark scheme would impact on the comparability of assessments. There were also questions about the existing principles for paper modification followed by awarding organisations. It seemed that there is no shared approach on the language of assessment across National Curriculum assessments and general qualifications, yet paper modifications are considered necessary to make assessments accessible to all.

Participants had reservations against how the modifications are made, so that the modifications did not advantage those that sit modified papers. Papers should be as accessible as possible to all learners who have the need for modification. Modification should only be carried out when absolutely necessary.

Some participants were more concerned with the security of modified language assessments in terms of the reliability of using signers and interpreters when allowing access arrangements. Other participants were certain that signers are trustworthy, but still worried about what exactly is being signed.

The second discussion group was unsure about the information that was available for and used by awarding organisations. Some participants argued that at every question paper evaluation committee (QPEC), awarding organisations had already established the style and approach to question paper items. The regulators could not monitor these initial conceptual developments. The regulators can only review the research and information that awarding organisations used for reference, and need to be involved and aware of current practice to be able to regulate this.

The regulators should know more about the language design and modification process with the monitoring of national GCSEs and A levels. The responsibilities of the regulators in this area need to be reviewed and defined in order for them to give a proportional response to problems. It is not clear how much the regulators should be involved with assessment modification. However, participants agreed that the regulators do need to develop a policy to monitor assessment modification.

### **Risk-assessment strategies**

Another key topic that was discussed was how new policies on paper modification should be founded on sound risk-assessment strategies. While identifying potential hazards and threats posed by the regulation and monitoring of assessment modification, some participants argued that the regulators needed to:

- act proportionally to present modification issues
- get detailed information on what awarding organisations are currently doing
- avoid concentrating only on the process of modification.

The scope of the discussion groups will have to expand to encompass NVQs. It will not be a good model of intervention to consider only the National Curriculum assessments and general qualifications, as the objective is to benefit all learners. Training for writing good papers was identified as an area in need of further development.

It was not clear whether the regulators should concentrate on better modifications for small percentages of learners, or concentrate on good assessment language that is beneficial for the largest percentage of test takers. Pitching only at small percentages could be controversial.

Some participants noted that in the 1980s, new strategies on item development were designed specially to include boys and get them to perform to the same standards as girls. This new policy increased the standards of the language used in test items. While looking at emerging issues on diversity and inclusion, regulators will need to back any interference in the field of paper modification with expertise on the topic. Regulatory risk should be proportionate to the amount of expertise gathered on the topic and equality in delivery must be ensured.

Ofqual, CCEA and DCELLS's regulatory expectations should be balanced and developed in consultations with BATOD, Royal National Institute of Blind People (RNIB) and JCQ. Different pressure groups should also support any possible new guidelines for language accessibility.

How advice on training new accredited modifiers is produced will need consideration, as modification is not a science, but an art. The regulators will have the duty of raising the issue of the shortage of accredited modifiers with the government, and inform it of the skills available for the next generation.

### **Modification of standard papers**

The modification of standard papers was not supported by all participants in the second discussion group. Some argued that the regulators should:

- Support more modification at source
- Require clear monitoring outcomes
- Promote more clarity on assessment specifications
- Support good provision of modification because some modifications were misleading to test takers
- Explain to test developers that by changing one aspect of assessment questions, they could be affecting other aspects and this interfered with the validity of assessment items.

There were three reasons to support modification at source.

1. Demand: there was already a tendency for greater demand on modification in England and every test taker would have a specific difficulty.
2. Diversity of modifications: test development agencies and awarding organisations could end up modifying assessments for nearly the whole cohort.
3. Costs: there was capability to do so, but what would be the costs involved?

There are only 40 accredited modifiers in the UK – how could this pool of experts be increased? Participants suggested that the regulators should strive to reduce the need for modification. Assessment specifications needed to be even clearer. Looking at assessments as if they were buildings, rather than adding steps to address different needs, ramps could be used in every building for general accessibility. These ramps were a metaphor for more accessible language. The regulators should avoid remedial responses.

Braille and modified large print will always be required for visually impaired candidates, but other difficulties and groups would all benefit from modification at source and universally designed assessment questions accessible to all test takers. This applied to National Curriculum assessments, general qualifications and NVQs. The National Curriculum assessment modifiers already use a model of language development in their approach to the language of assessments.

### **The third discussion group**

The third discussion group of the OPMF involved participants from:

- AQA
- BATOD
- BCS, The Chartered Institute for IT
- BECTA
- British Dyslexia Association
- Cambridge Exams (CamExam)
- CCEA
- Chartered Institute for Educational Assessors (CIEA)
- City & Guilds
- DCELLS
- OCR
- QCDA
- RNIB
- Scottish Qualifications Authority (SQA)

Participants also included an English for speakers of other languages (ESOL) modifier and representatives from Ofqual for most operational areas.

### **Modified language papers: For all or for none?**

The third discussion group attended a presentation by a specialist in language accessibility on *Modified Language Papers: For All or for None?* This presentation put forward the argument that when we talk about accessibility, we usually think of

very practical situations: lifts, ramps, travel, Braille. Yet language disability is more hidden.

Only since 1974 has there been any formal concept of educational inclusion, which was enshrined in the 1981 Warnock report. In a remarkably short time, children who were previously considered non-educable were accessing the full school curriculum, and wanting to demonstrate their knowledge and skills through the same assessments and examinations as their peers. The increasing demand for access arrangements and modified examinations in part celebrated this achievement.

In 2010, awarding organisations were producing about 10,000 different examination papers a year, whereas QCDA produced only nine Key Stage 2 papers. Modifications for learners requiring modified print included Braille, modified large print and enlarged print. That is to say, modified papers were only part of the access arrangements.

Access arrangements set by JCQ for general qualifications and by assessment and reporting arrangements in National Curriculum assessments included oral language modifiers, readers, lip speakers, scribes, and at Key Stage 2, translators and interpreters. There was a great demand on time and resources for awarding organisations and for the examination centres and inspectorate.

Politically sensitive issues about unfair advantage or disadvantage were not unusual. It's hard to find any research or evaluation that objectively evaluated the effectiveness of access arrangements, or whether they advantaged or disadvantaged any candidates unfairly.

In Key Stage 2 assessments, there is built in modification and guidance for communicators. In Key Stage 4 assessments, there is some built in (modified at source) some bolt on (modified on completion and on demand) modification. Key Stage 2 was developed and modified over a two-year cycle whereas Key Stage 4 did the same over a 24-week cycle.

Written papers prepared for Key Stage 4 were historically only available for the deaf, but the Equality Bill and Discrimination Act enshrined in law the principle of meeting candidates' needs. Therefore, modified papers should be available to any candidate able to provide evidence of a level of need for modified language. The degree of hearing loss was no indicator of the degree of language impairment. Candidates could have similar degrees of language impairment with very different degrees of hearing impairment, or no hearing impairment at all.

Demand for language modification could rise from 1:1000 candidates to an unknown number. It is worth noting that examinations were designed to discriminate between candidates, not so that everyone could get an A\*. What the government, schools and

awarding organisations wanted to ensure was that the discrimination was on the basis of a candidate's subject skills or knowledge, not on any other factors.

Considering that the accessibility of written papers was often assessed in terms of the reading level, all kinds of issues could affect readability, for example motivation, print style, text layout and illustrations. Predictability studies that use readability indexes are not very helpful in improving the language of assessments. There are difficulties in finding out the readability of short assessment questions because readability scales are usually based on passages of continuous prose of about 100 words, useful for the selection of source materials. Moreover, readability scales do not take into account the order of words in a sentence – they look only at the vocabulary. Yet changing the order of words could dramatically change the demand of a question. For example, words commonly used in simple mathematical problems were known to affect performance. The different arrangements of the words made the questions easier or more difficult for the pupils even though the mathematical processes were the same. But how can the predictability of language be achieved? Possibilities include:

- Using word order to convey meaning
- Simpler sentence structures
- Removal of embedded clauses, putting the most important word first.

With language being the tool by which assessments are understood, test developers need to use a lower level of language if they are assessing subject skills and knowledge either orally or through reading/writing, at which candidates will be confident. It is nonetheless appropriate to assess the understanding of subject-specific language (such as acute angle, perimeter) specifically taught in the subject context. Problems arise when words have a more general meaning or a different meaning in different contexts, as test takers often resort back to the first meaning they have learnt of a phrase. However, subject-specific vocabulary could also have different meanings in different contexts, for example the word 'volume' could mean book, loudness, amount of space.

GCSE assessment items rely on a wide variety of instructions or command words. Test takers must understand up to 300 different command words to be able to carry out the tasks. Assessments at Key Stages 1 to 3 use a much smaller variety of command words, but the knowledge of such words by younger learners could be a major issue. Command words have different meanings in different contexts and if taken literally, could cause significant misunderstanding, particularly for candidates on the autistic spectrum. Examples of confusing command words would be:

- Discuss the merits of...

- Illustrate your answer...
- Carry out an experiment...
- Trace the development of...
- Find...
- Outline...

However, simplifying command words could disadvantage a candidate if the words in the examination had been carefully chosen by an examiner to elicit a particular response. For all the above reasons, language modifiers need both subject expertise and language expertise.

There are certainly not enough teachers of the deaf who have historically done this work. Involvement with question specification meetings could help question writers to acquire the skills of writing questions using accessible English. However, even experienced modifiers can sometimes 'get it wrong' in cases where examiners refuse the proposed modifications because they consider that the proposals change the nature of the question, or give too much help to the learner. Debate has often been centred on what constitutes a technical or subject-specific term.

If a skilled modifier can get it wrong after careful consideration, what would be the risks of a decision made at leisure? How much more can oral communicators and sign interpreters who have been working in real time with candidates in the examination centre on the examination day do? Each intervention could subtly change the task and interpret the examiner's intentions.

With all the above difficulties, the regulators will look at the issues from a different perspective. If the availability of skilled modifiers was reducing, and with the pressures of time and financial resources for the awarding organisations, perhaps better use could be made of the skills and resources to train examiners to write accessible questions. It would be imperative for modifiers to have access to mark schemes, as working without them was likely to make the process redundant. Also, if modifiers were asked to modify papers at a later stage, they should make a note to the examiner to this effect.

It seems simpler to start by asking about what is being assessed before the paper is written. This would enable standardisation of question language. However, if question setters were to modify language at source, then many more skilled professionals would be required throughout the UK. Braille papers would require a similar approach. A set of skills would need to be set for professional modifiers, but it would be the responsibility of the regulators to know the criteria and to incorporate the comments and questions from language accessibility debates.

Participants were asked if it would be manageable to create a new certificate on language modification. They answered “yes”, this would be essential as the UK needed a much bigger pool of language modifiers. In 2010, language modification only needs the BATOD rubber stamp – it would be better to have a qualification and possibly CIEA involvement. Awarding organisations should indicate whether they would only want to use language modifiers who have the qualification or certificate, and the regulators should lead on this.

### **The Support Model**

During the third discussion, CamExam assessment researchers provided feedback on the Support Model research commissioned by Ofqual. The aim of this study, conducted by CamExam, was to test the suitability of the Support Model methodology for use with learners with SLDs. The Support Model assessed a learner’s ability based on the amount of support they needed to complete a task. Support was given in the form of systematically delivered prompts.

Prompts were categorised as:

- Reading prompts
- Understanding prompts
- Activation prompts
- Writing prompts
- Affective prompts.

The prompts allowed the researchers to use the Support Model to assess the learner’s understanding of the important concepts in a subject, without their SLDs getting in the way. The scoring system could be altered to include different categories of demand as ‘construct relevant’.

All the pupils in the study completed the task and ended with a positive sense of achievement.

Ahmed and Pollitt’s (2010, unpublished) research design assumed that in general, there was threat to validity when one put examination tasks into words to communicate to pupils. The Support Model dealt with this by allowing an interaction, so that pupils understood fully what the task was and could show their understanding, even when the task was not clearly communicated via the question. Learners with SLDs were likely to be more affected than other learners by construct-irrelevant language demands.

During the research, the prompts were standardised, so that every pupil received or could receive exactly the same science-related prompts. The assessor could give each pupil every prompt they might need and they were therefore assessed in a standardised way.

The number of participants was small because the purpose of the qualitative experiment was not to generalise any findings, but to understand how test takers reasoned during a supported assessment experience. The sample included seven pupils aged between 10 to 11 years in years 5 and 6 of a primary school in Cambridgeshire with a speech and language centre attached to it. Each pupil had a different profile of SLDs: three were classified as 'mild' and four as 'moderate'.

The materials used were three questions from past papers in Key Stage 2 science. The pupils, working individually in a room, showed evidence of understanding the science concepts and were able to carry out the tasks, but had difficulty reading and interpreting the tasks and putting their answers into appropriate words. Three of the pupils with moderate SLDs scored higher on science than on communication for all three questions. This indicated that their science understanding was good, but that they would have found it difficult to show this in a traditional examination setting without communication support. Two other pupils, one 'moderate' and one 'mild', showed this pattern on two out of the three questions.

The results of this experimental research confirmed that pupils were able to use the prompting system successfully, and that it allowed them to show science-related knowledge and understanding that otherwise may have been masked by their SLDs. Pupils using the Support Model were likely to get higher scores than they would have done without it. Ahmed and Pollitt (2010, unpublished) noted that there is a bias against all pupils who are not given this kind of language support. The solution would be to find a method of establishing a fair equating, so that all pupils are assigned the correct level whether or not they have this language support.

Until the automated Support Model is fully developed, the system would need a human interpreter for each pupil and assessors will have to be trained briefly on how to use the prompts. In order to remove the need for a human interpreter, the researchers hoped to develop a computed adaptive assessment system that would evaluate pupils' responses and select the appropriate prompt. Such an automated system could work only in on-screen environments and this would severely limit opportunities for use.

Other perceived limitations of the Support Model are outlined below.

- In terms of the manageability of assessment, developing the Support Model methodology for large-scale summative tests in the UK would require IT for everybody. It is difficult to see how this would work for tests that are sat on the

same date and time nationwide. One participant noted that he would be sceptical about using the model for large entry subjects. The Support Model would have to be IT based, as it placed too much demand on teachers to systematically deliver prompts on an individual basis. The Support Model, as presented, would mean having one teacher per pupil.

- It was not clear whether it was possible to measure other variables in the prompts, such as the ability to learn and length of learning speed for each pupil. The Support Model scored pupils according to the number of prompts required. Participants considered that activation would be difficult for some pupils – could different learning techniques affect the pupil’s ability to activate that learning? Researchers agreed that different learning techniques would affect the pupil’s ability and therefore affect their score in ‘activation’. Despite all this, the CamExam study showed that the Support Model does help and pupils would be in a better situation than if they were faced with a written paper and scored no marks.
- Participants noted that it would be useful to go a step ahead and be able to access a profile of each pupil’s progress, know what kinds of prompts they used and how successful they were. At the time, the scores were set out to rank the order of the pupil’s achievement, but the researchers agreed that a profile of pupil response would be very useful and of great value formatively.
- There was a need to add voice-over to the prompts delivered on computer. In Scotland, some of the digital papers with voice-overs were working well, so there would be no problems in principle. However, this will require further investigation and details such as accents will need to be taken into account.
- The cost to develop structured prompts for all pupils sitting for example single level tests was likely to be high and could well be impractical. Participants were sceptical about this, both in terms of it working and cost. Some participants were sceptical that any automated system would ever be able to recognise spelling mistakes. Moreover, the full automation of the computer-based procedure would take more than a year of further development.
- It seems unfeasible to use the Support Model in a large-scale operation as every pupil at some point could require a prompt. There was not a one-to-one scale in the model. It was not clear whether there was a need for one teacher to deal with 10 pupils or more. Disruption to others was not yet considered and it did not consider the effect of overhearing prompts and the impact on other pupils.
- Difficulties would begin to occur when an answer could be better/more/less explained and the Support Model broke down.

- In mathematics assessments, it would be difficult to agree on appropriate prompts and prompts that should be penalised, as the link between technical information and the language of the question was profound and impossible to separate.
- There would be issues around timing if the model was used with lower-tier GCSE papers. Could timing be built into the computer-based model? The CamExam researchers' response was that if the Support Model was automated, a non-response from a test taker after a while would trigger a prompt from the computer. There would need to be a balance if many test takers found that the system took too long and they become tired of dialogue with the prompter. Another problem was that often there were many subjects being examined on the same day.
- There were some reservations about the automated system being a substitute for a human presence to provide the prompts, as a computer would not be able to identify why the pupil was not responding, for example subtle inferences in tone and body language. There is therefore a need to investigate issues further.
- The effects of the familiarity of the prompter to the pupil need to be considered. Although the method would be standardised in summative assessment, it could still raise some issues if the prompts were delivered by somebody known to the pupil.

Despite initial limitations relating to the reliability of the model when used in summative assessment, the Support Model method seemed to have given a more valid assessment of the science-related knowledge and understanding of a small group of pupils with SLDs because language-related construct-irrelevant variance was removed. The model could be used in any subject with a clear, correct answer and it was also perceived as a better tool for overcoming reading difficulty than readability formulae.

The advantages of using the Support Model were greater as a formative teaching tool because the prompts were what a good teacher should be using in the classroom (dynamic assessment). The Support Model could be applicable to standardisation scripts that could be given to teachers/assistants in the classroom at Key Stages in subjects other than English and mathematics, as a tool for promoting reliability of teacher assessment. This would be an interesting concept to follow up in Ofqual's reliability work.

The third discussion group considered the Support Model to be sound and valid as formative assessment because it was based on the use of standardised prompts that had been theoretically grounded and systematically generated. This would be a

significant improvement on any approach that either pre-determined the wording modification before any interaction with the pupil or was reactive, but non-systematic.

Other advantages of the Support Model that were highlighted related to:

- Classroom assessment. The Support Model's interactive approach could answer to individual needs, becoming useful in classroom assessment
- GCSE controlled assessments. The Support Model could be treated as a good and valuable addition for controlled assessments
- Greater reliability of teacher assessment. The Support Model could prove the reliability of teacher assessment.

After evaluating the results and listening to the third discussion group, Ofqual's representatives recognised that the Support Model would be best suited to small-scale formative teacher assessment.

### **New rules on language accessibility**

The following questions were raised at the third discussion group:

- Would it be possible for test developers to avoid unnecessary barriers to all types of assessment?
- What kind of rules would future non-statutory appendices to the Regulatory Framework for National Assessments (RFNA) and *GCSE, GCE and AEA Code of Practice* (Ofqual/09/4151) need to set?

According to the participants, test development agencies and awarding organisations have a duty to provide accessible assessments to all candidates. Together with the regulators, they would need to identify these barriers and justify them in legal terms.

A first step would be for assessment providers to spell out subject criteria. *Fair Access by Design: Guidance for Awarding Bodies and Qualifications Regulators on Designing Inclusive General Qualifications* (Ofqual, DCELLS and CCEA, 2009) has been accepted as a crucial document on language accessibility, but it is only meant to be used as a non-statutory guide. A new (unpublished) version of *Fair Access by Design* includes details on designing qualifications, source materials and marking schemes, as well as case studies and information on e-assessment.

DCELLS chairs the Access to Assessment and Qualifications Advisory Group on behalf of the regulators and one of its members, together with other participants in the discussion group, suggested some points that would need to be considered in developing any code of practice.

- Not merge each step

- Give awarding organisations the duty to have in place procedures to design assessments that are accessible and valid
- Include a glossary on the notes, using plain English
- List items that are difficult and create barriers to accessibility. Inform different groups on the completion of such a list
- Include obligation of awarding organisations to note 'moderation'
- Feed back to modifiers to avoid variation between modifiers.

To avoid making the same mistakes that have been identified by awarding organisations, it will be necessary to:

- Find more modifiers
- Create training in language modification
- Create a system to monitor modification
- Have someone at every QPEC meeting who is a specialist in language or an examiner qualified in language.

An appendix to the code of practice would also need to address the remedial action necessary for those who have specific language deficit and those for whom English is an additional language. There would be a need to address the reliability of systems that can be applied consistently to learners in any centre/school, so that results are comparable, given the same starting point. The key concept would be equitability between those who are receiving support/access arrangements and those who are not, so that the level of demand is equitable.

In order to promote fairness and provide guidance on paper modification consistently for all learners, participants agreed that a guide on language accessibility would need to take into account that:

- There is a need for stability in design and structure of papers. Awarding organisations and test development agencies would like to follow simple guidelines on language accessibility
- Idiosyncrasies of individual modifiers should be avoided and a central resource for post-delivery evaluation secured
- Assessment providers would require practical help and would like Ofqual to facilitate agreement in this area with action

- Guidance should identify barriers to potential groups and address functional skills in ICT
- The notion of 'plain English' will need to be defined
- It is important that the language 'engages' learners.

*Fair Access by Design: Guidance for Awarding Bodies and Regulatory Authorities on Designing Inclusive Vocational Qualifications* (QCA, ACCAC and CCEA, 2006) already includes sections on the readability of written questions, with vital points on language accessibility. A revised version of *Fair Access by Design* for general qualifications and for vocational qualifications will need to:

- Define the command words. There is concern that this may not be possible because different people have their own pet definitions ('chapel' syndrome)
- Provide guidance on rubric
- Provide guidance on typographical issues
- Deal with syntax/semantics.

A new guidance issued by the regulators should consider succession planning. As there is a lack of expertise, participants suggested that a professionalisation of the role is required. There would be a need for a control body of trained or accredited modifiers.

## **6. Conclusion**

In 2010, awarding organisations, regulatory bodies, other associations and institutes for the deaf, the visually impaired and dyslexic people showed support for the production of non-statutory codes on language accessibility and modifications to question papers.

A key point that emerged from the OPMF was that awarding organisations were not satisfied with the number of qualified language modifiers available in the UK. Awarding organisations expected new steering forward from the regulators. They wanted the regulators to exert leadership by:

- Developing new standardised systems to review the language of question papers
- Writing a guide on language accessibility
- Supporting the creation of a new certificate in language modification.

In relation to the modification of the National Curriculum assessments, more research was required to investigate the impact of modifications on pupils from some of the six equalities strands:

- Disability
- Gender
- Race and ethnicity
- Sexual orientation
- Religion and belief
- Human rights.

The OPMF showed support for a new guide on language accessibility to standardise language modification and modification for the visually impaired. A few specialists held views on how to address language modification and other modification in more effective ways in England, Wales and Northern Ireland (see the full version of this paper).

Any guidance produced by the regulators should not take for granted that paper modification for the blind seemed to be more consistent. The regulators should take into account that according to most awarding organisations, there were issues with the language modification for candidates with hearing impairment because this was

slightly more reliant on professionals' opinions and such opinions were not always consistent.

Language modification could risk being particularly unreliable, and in future, any modification proposed by two different modifiers should converge. The regulators would need to work towards consistency in language modification and modification for the visually impaired. To increase consistency, individual modifiers will need to base their work on a set of objective rules and undergo special accreditation on universally accessible language for assessment.

## 7. References

- Abedi, J. (2006) *Language Issues in Test Development*. In Downing, S. and Haladyna, T. (2006) *Handbook of Test Development*, pp. 377–398. London, Lawrence Erlbaum Associates.
- AERA, APA and NCME (1999) *Standards for Educational and Psychological Testing*. Washington, AERA, APA and NCME.
- AERA, APA and NCME (1990) *Standards for Educational and Psychological Testing*. Washington, AERA, APA and NCME.
- Ahmed and Pollitt (2010) *Alternatives to Test Accommodation: New Models for Pupils with Speech and Language Difficulties. The Support Model*. Unpublished.
- Angoff, W.H. (1993) *Perspectives on Differential Item Functioning Methodology*. In Holland, P.W. and Wainer, H. (Eds.), *Differential Item Functioning*, pp. 3–24. Hillsdale, NJ, Lawrence Earlbaum.
- Black, J. (2002) *Critical Reflections on Regulation*. London, Centre for Analysis of Risk Regulation, London School of Economics and Political Science.  
[www.lse.ac.uk/collections/CARR/pdf/DPs/Disspaper4.pdf](http://www.lse.ac.uk/collections/CARR/pdf/DPs/Disspaper4.pdf) . Accessed October 2010.
- Educational Testing Service (ETS) (2009) *ETS Guidelines for Fairness Review of Assessments*. [www.ets.org/Media/About\\_ETTS/pdf/overview.pdf](http://www.ets.org/Media/About_ETTS/pdf/overview.pdf). Accessed October 2010.
- Educational Testing Service (ETS) (2002) *Standards for Quality and Fairness*. [www.ets.org/Media/About\\_ETTS/pdf/standards.pdf](http://www.ets.org/Media/About_ETTS/pdf/standards.pdf). Accessed October 2010.
- Haladyna, T.M. and Downing, S.M. (2004) ‘Construct-Irrelevant Variance in High Stakes Testing’, *Educational Measurement: Issues and Practice*, 23(1), pp. 17–27.
- JCQ (2008) *Access Arrangements, Reasonable Adjustments and Special Consideration*. [www.jcq.org.uk/attachments/published/538/25\\_AARASC\\_0910.pdf](http://www.jcq.org.uk/attachments/published/538/25_AARASC_0910.pdf). Accessed October 2010.
- Levine, M. E. and Forrence, J. L. (1990). “Regulatory Capture, Public Interest, and the Public Agenda. Toward a Synthesis”. *Journal of Law Economics and Organization* 6: pp.167–198.
- Linn, R.L. (2006) “The Standards for Educational and Psychological Testing: Guidance in Test Development”. In Downing, S.M. and Haladyna, T.M. *Handbook of Test Development*. New Jersey, Lawrence Erlbaum Associates.

Messick, S. (1984) "The Psychology of Educational Measurement". *Journal of Educational Measurement*. 21(3), pp. 215–237.  
[www.jstor.org.ezproxy.sussex.ac.uk/stable/1434780](http://www.jstor.org.ezproxy.sussex.ac.uk/stable/1434780) . Accessed October 2010.

NFER. *Assessment: A Discussion of Validity*.  
[www.nfer.ac.uk/nfer/research/assessment/about/validity.cfm](http://www.nfer.ac.uk/nfer/research/assessment/about/validity.cfm) . Accessed October 2010.

Nonet, P. and Selznick, P. (1978) *Law and Society in Transition: Toward Responsive Law*. New York, Octagon Books.

Ofqual (2010) (Ofqual/10/4733) *National Curriculum Assessments: Code of Practice 2010*.  
[www.ofqual.gov.uk/component/search/code+of+practice/%2F?ordering=&searchphrase=all-content](http://www.ofqual.gov.uk/component/search/code+of+practice/%2F?ordering=&searchphrase=all-content). Accessed October 2010.

Ofqual (2010) *Statistical Bulletin: Access Arrangements for GCSE and GCE: June 2009 Examination Series*. [www.ofqual.gov.uk/.../2010-02-03-access-arrangements-in-gcse-and-gce-exams-2009.pdf](http://www.ofqual.gov.uk/.../2010-02-03-access-arrangements-in-gcse-and-gce-exams-2009.pdf). Accessed October 2010.

Ofqual, DCELLS and CCEA (2010) (Ofqual/10/4718) *GCSE, GCE, Principal Learning and Project Code of Practice*. [www.ofqual.gov.uk/files/code-of-practice-2010.pdf](http://www.ofqual.gov.uk/files/code-of-practice-2010.pdf). Accessed October 2010.

Ofqual, DCELLS and CCEA (2009) *Fair Access by Design: Guidance for Awarding Bodies and Qualifications Regulators on Designing Inclusive General Qualifications*. [www.rewardinglearning.org.uk/docs/regulation/fair\\_access\\_final\\_gcse\\_gce.pdf](http://www.rewardinglearning.org.uk/docs/regulation/fair_access_final_gcse_gce.pdf) . Accessed October 2010.

Ofqual (2008) (Ofqual/09/4121) *Ofqual's Statistics for Access Arrangements and Special Consideration at GCSE and A level: 2008*. [www.ofqual.gov.uk/.../Access-arrangements-and-special-consideration-report-2008.pdf](http://www.ofqual.gov.uk/.../Access-arrangements-and-special-consideration-report-2008.pdf). Accessed October 2010.

QCA (2006) (QCA/06/2888) *NVQ Code of Practice*. [www.ofqual.gov.uk/files/qca-06-2888\\_nvq\\_code\\_of\\_practice\\_r06.pdf](http://www.ofqual.gov.uk/files/qca-06-2888_nvq_code_of_practice_r06.pdf). Accessed October 2010.

QCA, ACCAC and CCEA (2006) *Fair Access by Design: Guidance for Awarding Bodies and Regulatory Authorities on Designing Inclusive Vocational Qualifications*. [www.rewardinglearning.org.uk/docs/regulation/fair\\_access\\_final100406.pdf](http://www.rewardinglearning.org.uk/docs/regulation/fair_access_final100406.pdf) Accessed October 2010.

Thompson, S. J., Johnstone, C. J., and Thurlow, M. L. (2002) *Universal Design Applied to Large Scale Assessments (Synthesis Report 44)*. Minneapolis, MN, University of Minnesota, National Center on Educational Outcomes.  
<http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>. Accessed October 2010.

Zieky, Michael (2006) *Fairness Review in Assessment*. In Downing, S. and Haladyna, T. (2006) *Handbook of Test Development*, pp. 359–376. London, Lawrence Erlbaum Associates.

Ofqual wishes to make its publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2010

© Crown copyright 2010

Office of Qualifications and Examinations Regulation	
Spring Place	2nd Floor
Coventry Business Park	Glendinning House
Herald Avenue	6 Murray Street
Coventry CV5 6UB	Belfast BT1 6DN

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346

[www.ofqual.gov.uk](http://www.ofqual.gov.uk)