Current practices and operational aspects of paper modification in England 2009/10

October 2009
Product code: OFQUAL/09/4530

Contents

Summary 3

Project A Current practices and operational aspects of paper modification in England 2009/10 4

Aims 4

Definitions 4

Access arrangements 4

Reasonable adjustments 5

Special considerations 6

Modified papers 6

Project B Alternatives to test accommodation: new models for pupils with speech and language difficulties 8

Justification 9

The proposal 13

Rationale 15

Background 17

What counts as paper modification? 17

The Test Review Group 18

Paper modification at QCDA 18

Previous reports 19

Paper modification abroad 20

The United States 21

Educational Testing Service view 22

Israel 24

The Netherlands 24

Brief description of research 25

Current practices and operational aspects of paper modification in England 2009/10
Potential impacts of research 28
What to do with research results 28
Method of investigation and timescale (42-day project) 29
Phases of research on paper modification: operational perspective 31
Phases of research explained 32
Bibliography 34
Appendix 1. Specifications for a report by QCDA 38
Appendix 2. Specifications for awarding bodies reports 41

Summary

This document looks at current practices and operational aspects of paper modification in England. It proposed two pieces of research, Projects A and B.

Project A is a 42-day operational piece of research that seeks to equip Ofqual to deal with challenges surrounding the use of modified question papers in assessments to be carried out by a member of the Ofqual Research Team. It has two main deliverables:

- a 9,000-word report on paper modification in England 2009/10
- an Ofqual code of practice on paper modification.

The latter will contain operational, functional, theoretical and technological guidance on the modification of standard papers.

The methodology of Project A includes the management of in-house discussion groups – the Ofqual Paper Modification Forum or OPMF – and interviews with those responsible for the paper modification process in national curriculum tests and GCSEs.

Project B is experimental research externally commissioned by Ofqual. This research addresses the need to guarantee clear and flexible paper design and supports research into how to minimise bias. It involves an experimental pilot with key stage 2 pupils with and without speech and language difficulties.

Project A Current practices and operational aspects of paper modification in England 2009/10

This project adopts the broad definition of *modified papers* established by the Joint Council for Qualifications (JCQ):

> 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.[1]

Aims

This operational research seeks to equip Ofqual to deal with challenges surrounding the use of modified question papers in assessments, which include:

- the need to produce a code of practice on paper modification that would define the principles, methodology and standards that test development agencies (TDAs) and awarding bodies would aim to meet in producing their examinations
- the need to ensure the public that Ofqual's diversity and inclusion policies do not clash with its work to maintain standards for qualifications and assessments and consistently aim to reduce construct-irrelevant barriers to test performance
- the need to assess operational issues involving test design, score interpretation and the monitoring of test modifications so that the Ofqual Management Group can identify where paper modification sits within the organisation.

Definitions

The notion of paper modification is generally affected by its association with 'access arrangements', 'reasonable adjustments' and 'special considerations'. According to the 2009 JCQ report *Access arrangements, reasonable adjustments and special consideration*, the following definitions are current in England.

Access arrangements

These are pre-exam adjustments made for individual candidates, based on evidence of need and the candidate's normal way of working. They exist to ensure all candidates have the same opportunity to be successful in their exams. The QCDA

[1] JCQ (2009) Access arrangements, reasonable adjustments and special consideration, p. 38 www.jcq.org.uk/attachments/published/1096/25.%20AARASC%200910.pdf. The JCQ report was produced on behalf of AQA, Edexcel, OCR, City & Guilds, CEA, WJEC, with effect from 1 September 2009 to 31 August 2010.

Test and exam support website explains that "access arrangements are designed for a small number of pupils that may require additional arrangements to access the tests."[2] QCDA considers access arrangements to be "adjustments that schools must consider prior to the tests, and should be based primarily on normal classroom practice for pupils with particular needs. Schools must have evidence to show that resources are routinely committed to providing this support."[3] Candidates must request access arrangements from QCDA (for independent schools and academies) or local authorities (for local authority maintained schools). Applications for access arrangements made to the local authority are submitted through the QCDA *Test forms* website at www.qca.org.uk/testforms.

Awarding bodies for general qualifications publish their own regulations on the access arrangements that can be provided. In 2009, the JCQ listed 33 types of access arrangements[4] including:

- additional tapes/CD for deaf/hearing impairment
- modified enlarged A4 (18-point bold) for visual impairment
- modified enlarged A4 to A3 (24-point bold) for visual impairment
- modified language paper for substantial comprehension difficulties
- oral language modifier – preceded by psychological assessment carried out by a qualified psychologist, or specialist assessment carried out by a specialist teacher confirming below average reading/comprehension
- prompter
- sign language interpreter.

Modified papers were certainly an important category of access arrangements, with some papers being modified at source by teachers skilled in language modification.

Reasonable adjustments

In September 2007, the Disability Discrimination Act 1995 (DDA) was extended to cover awarding bodies that offer general qualifications, such as A levels and

[2] QCDA *Access arrangements* website testsandexams.qcda.gov.uk/16158.aspx.

[3] QCDA *Access arrangements website* testsandexams.qcda.gov.uk/20069.aspx.

[4] For a synopsis of access arrangements and evidence requirements see JCQ (2009) p. viii to ix www.jcq.org.uk/attachments/published/538/29.%20Access%20Arrangements%20Booklet.pdf.

GCSEs.[5] The Act gave rights to disabled people, those with a "physical or mental impairment which has a substantial and long-term adverse effect on his/her ability to carry out normal day to day activities".[6]

In agreement with the DDA, equal opportunities policies of awarding bodies had to guarantee a fair assessment for all candidates and today it is permissible, when necessary, to make reasonable adjustments for candidates with particular requirements to enable them to access fair assessment and demonstrate attainment. However, any reasonable adjustments made must not invalidate the assessment requirements set out in the specification for the relevant qualification. Reasonable adjustments must not give candidates an unfair advantage compared to candidates for whom reasonable adjustments were not been made. Such arrangements ought to maintain the relevance, reliability and comparability of the assessment.

Special considerations

Pupils whose performance in the tests is affected by extremely distressing circumstances are eligible for special consideration. In such cases schools can apply for special consideration available in the QCDA *Test forms* website.[7] However, special considerations are not appropriate for pupils who have not covered the curriculum or not been prepared for the tests.[8] An application can only be made where a pupil has taken all the tests in a subject. No level can be awarded for a pupil who was absent for one or more of the tests in the subject applied for.

Modified papers

The aim of paper modification in this broad sense is to remove unnecessary barriers to comprehension by getting rid of complicated sentence structures where they are not basic to the question itself. Awarding bodies are slowly "working towards the modification of carrier language in all question papers, so that the standard paper should be suitable for all candidates with substantial comprehension difficulties, irrespective of the reason for the impairment".[9] Some papers had already been modified prior to being produced as the standard paper for all candidates. Yet, the JCQ notes, despite the need for modified papers, technical language and abstract

---

[5] Ofqual, Reasonable adjustments www.ofqual.org.uk/467.aspx.

[6] See definition of disability under DDA www.rnib.org.uk/livingwithsightloss/moneyandrights/knowingyourrights/disabilitydiscriminationact/Pages/definition.aspx.

[7] See www.qca.org.uk/testforms.

[8] See Special consideration, testsandexams.qcda.gov.uk/20279.aspx.

[9] JCQ (2009), p. 38.

concepts underpinning the subject cannot be removed as these are part of the assessment objectives being tested. JCQ and awarding bodies had been involved with small-scale pilots to adapt modified papers to be read on screen.[10]

[10] JCQ (2009), p. 38.

Project B Alternatives to test accommodation: new models for pupils with speech and language difficulties

Project B is experimental research externally commissioned by Ofqual. In order to be accountable to the public and guarantee best practice in assessment, Ofqual needs to identify innovative models of assessment and support research into how efficiently these models minimise bias. This project addresses the need for valid, reliable and innovative models of assessment for pupils with speech and language difficulties through an experimental pilot involving key stage 2 pupils with and without speech and learning difficulties.

Invitations to quote have been sent to UK researchers with expertise in the field of test development and new models of assessment. Research is due to start in early November and Project B must be delivered at the beginning of March 2010.

Justification

According to JCQ (2009) there are technological developments on paper modification under way. Yet, there is no validity evidence on new approaches to paper modification carried by TDAs and awarding bodies. To address the lack of clear policy on paper modification, Ofqual would benefit from leading a validity research on operational, functional, theoretical and experimental aspects of paper modification irrespective of pupil's level of language difficulty.[11]

At the moment, the Standards Group can only look at regulatory aspects such as how test agencies and QCDA are developing new assessments or how well they are following certain principles of good test designing. Ofqual monitors the QCDA-setting level thresholds but this relates to grading only. The GCSE, GCE, GNVQ and AEA code of practice requires awarding bodies to collect data about access arrangements,[12] broken down by qualification type, for centre-delegated access arrangements and awarding body-approved access arrangements, showing the number of individual candidate applications, by category (set by awarding bodies regulations and guidelines), for access arrangements and the numbers granted.

QCDA has collected data on the number of approved access arrangements at the end of each summer examination series since 2004. In 2007, it published a report on the performance of awarding bodies for general qualifications, with data showing a general increase in the overall number of approved arrangements. The table below shows the combined data collected from AQA, Edexcel and OCR between 2003/4 and 2005/6 and the categories of arrangements should be consistent with those set out in the awarding bodies' regulations and guidelines.

[11] This is the basis for Project B.

[12] See QCA (2007), *QCA review of GCSE and GCE access arrangements from 2004 to 2006*, November, paragraph 7.10, QCA/07/3419 www.ofqual.gov.uk/files/qca-07-3419_access_arrangements.pdf.

Number of arrangements by category made by AQA, Edexcel and OCR between 2003/4 and 2005/6

	2003/4	2004/5	2005/6
GCSE and GCE subject entries	7,442,314	7,442,445	7,620,798
Awarding body-approved arrangements (excluding modified paper questions)	103,818	104,907	125,114
Centre-delegated approved arrangements	81,825	43,869	78,833
Modified question papers	22,037	20,346	20,539

Source: QCA (2007).

An Ofqual report on access arrangements and special considerations applications between September 2007 and August 2008[13] indicated that awarding bodies approved 179,611 requests for arrangements for A level and GCSE candidates. The total number of awarding body-approved arrangements for 2008 increased by 20 per cent compared with the equivalent period for 2007, which was 125,114 (see table above).[14] 'Reader/computer reader' was the most frequently occurring arrangement in 2008 (see table below).

[13] See Ofqual (2009), *Statistics for access arrangements and special considerations at GCSE and A level*: *2008*, Ofqual/09/4121 www.ofqual.gov.uk/files/Access-arrangements-and-special-consideration-report-2008.pdf.

[14] The total number of A level and GCSE question papers distributed in the same period increased by approximately 29 per cent.

Number of awarding body-approved access arrangements by type of arrangement

Type of access arrangement	AQA	Edexcel	OCR	Total
Additional tapes/CD	not available	0	301	301
Alternative venue	1,012	285	678	1,975
Colour naming (by the invigilator)	166	116	166	448
Early opening of question paper (up to one hour before start time)	972	215	310	1,497
Extra time (more than 25%)	1,244	610	873	2,727
Live speaker	104	47	99	250
Oral language modifier	276	19	211	506
Practical assistant	388	209	312	909
Reader/computer reader	**35,376**	**26,272**	**36,056**	**97,704**
Scribe (voice input system)	18,798	11,947	17,565	48,310
Use of signers	248	145	318	711
Transcript of recording	not available	2	99	101
Voice-activated computer	not available	17	0	17
Word processor	9,667	5,979	8,374	24,020
Certificate indications	98	6	31	135

Source: Ofqual (2009).

According to the table above, readers for candidates with particular visual impairments or a learning difficulty accounted for just over half of all approved requests, 97,704 in total. To conclude, it is clear that:

- the number of awarding bodies approving arrangements is increasing

- in 2008 an average of 10.5 per cent of pupils in England's primary schools and 9.3 per cent in secondary schools[15] were classified as having special educational needs (SEN) statement[16]
- there are a substantial number of primary and secondary SEN candidates eligible for key stage 2 tests and GCSEs examinations that may need to be modified.

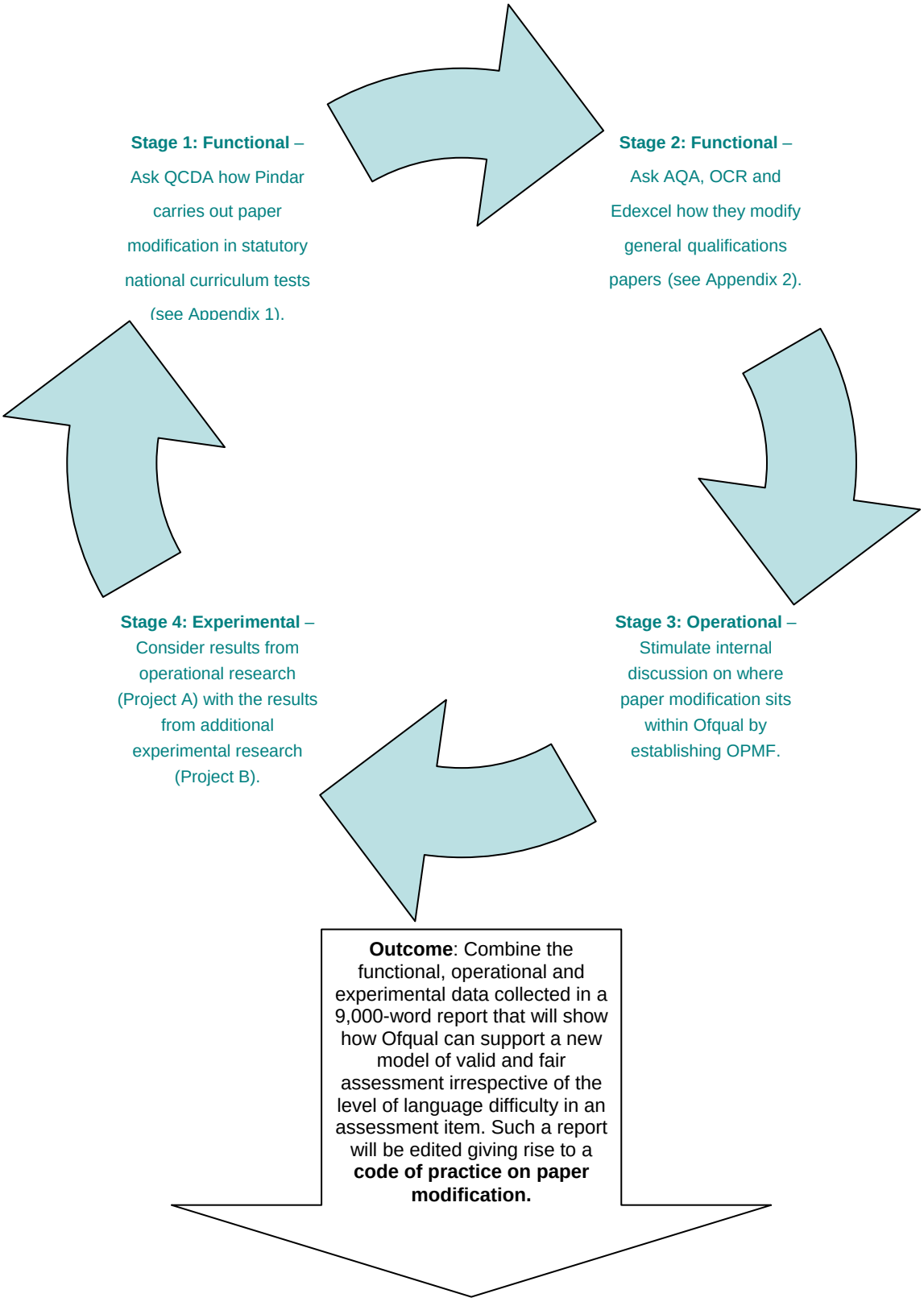
[15] See 'Pupils with SEN, England average' in DCSF website *Achievement and attainment tables 2008* www.dcsf.gov.uk/performancetables.

[16] See Special Educational Needs and Disability Act 2001, Chapter 10. This Act requires responsible bodies to take reasonable steps to ensure that disabled pupils are not placed at a substantial disadvantage www.opsi.gov.uk/ACTS/acts2001/ukpga_20010010_en_1.

The proposal

This proposal aims to facilitate regulatory work where there is a need to produce good practice guidelines for test designing that benefit all candidates with or without disabilities, with a focus on speech and language difficulties.

This research design approaches paper modification through a fourfold strategy.


Stage 1: Functional – Ask QCDA how Pindar carries out paper modification in statutory national curriculum tests (see Appendix 1).
Stage 2: Functional – Ask AQA, OCR and Edexcel how they modify general qualifications papers (see Appendix 2).
Stage 3: Operational – Stimulate internal discussion on where paper modification sits within Ofqual by establishing OPMF.
Stage 4: Experimental – Consider results from operational research (Project A) with the results from additional experimental research (Project B).
Outcome: Combine the functional, operational and experimental data collected in a 9,000-word report that will show how Ofqual can support a new model of valid and fair assessment irrespective of the level of language difficulty in an assessment item. Such a report will be edited giving rise to a **code of practice on paper modification.**

Rationale

It is a constitutive part of test development theory that every test design should anticipate the need for *accommodations* for disabilities. Examinees may need changes in test administration procedures based on their legal right to accommodation, which amounts to extending time or changing the allowable methods to respond to test questions. There is no doubt that test manuals should provide options for accommodating students with severe physical or communication difficulties (Braden and Elliot, 2003; Elliot, Braden and White, 2001, p. 539). As the Equalities Review (2007, p. 8) highlights, “many pupils with SEN are not achieving their true potential at primary and secondary school.” In the same view, there must be a path to broader and more equitable access to education and employment. Policy makers and assessment experts are urged to seek better solutions for disabled young people who cannot continue at risk of being not in education, employment or training.

Test fairness depends on a multiplicity of factors such as quality management in test design, efficient administration, coherent scoring, suitable coverage of relevant content, construct validation work and the functioning of test items. A fair test must include items that measure only the skill or ability being scrutinised and avoid introducing multidimensionality into the measurement; for example, candidates with speech and language difficulties might not understand a test item “if important words are not emphasised ... [or] may fail to understand and use abstract concepts (time, space, quantity, emotions). They may experience grammar and syntax difficulties and tend to interpret language literally.”[17] Socioeconomic background, medical and other background characteristics may lead to construct-relevant variance that should not interfere with test performance.

Changes to the testing situation, known as *accommodations* or *modifications*, are a real fairness requirement but modifications are among the most controversial issues in test administration because changes are made to what is considered the ‘standardised’ test condition (Thurlow, Thompson and Lazarus 2006, p. 657). The present context is:

- there are *access arrangements* in place for national curriculum tests and post-16 examinations in England and it is only after centres have tried everything that they offer an exemption to candidates with disabilities and language difficulties

[17] Ofqual (2008), *Oral language modifiers focus group*, February, Ofqual/08/3518, p. 6 www.ofqual.gov.uk/files/Ofqual-08-3518_Oral_language_modifiers_focus_group_11-08-08.pdf.

- at the moment QCDA employs NFER and Pearson Research and Assessment (PRA) to develop the national curriculum tests and Pindar as the modified test agency for statutory tests only; for single level tests, QCDA has a different TDA.

Yet there are problems to be solved regarding the language of examinations. One fundamental question is: Can we achieve an assessment model that is fair for all students with or without disabilities?

The programme leader in monitoring national curriculum assessments at Ofqual understands that in order to guarantee the internal and external validity of test design:

- the regulator needs to give thought to the rubric of tests, i.e. time to be spent on each question, the language of instructions, and format of papers
- there is a clear need for Ofqual to formulate a code of practice on modified papers.

To guarantee good quality modified papers, this research will involve TDAs and awarding bodies in the production of reports, which will support a debate within Ofqual on what modified test development should look like.

Background

What counts as paper modification?

Assessment policies such as 'reasonable adjustments', 'reasonable accommodations', 'fair access by design' and 'oral language modifiers' have been tried and implemented in England to remove some of the barriers that traditional assessment practices imposed on candidates with disabilities.[18] The 2006 Amendment to the Disability and Discrimination Act (1995) fostered even more equality.[19] The UK legislation challenged not so much test developers responsible for designing the national curriculum tests but focused more on awarding bodies responsible for general qualifications. By raising the status of 'competence standards', the DDA challenged the deep-rooted notion of 'academic standards'. The assessment of competence should be flexible, based on new modes of competence measurement. This shake-up increased equal opportunities, independent of particular conditions and background.

At present, an exam centre can make a request for paper modification using the modified papers tool by following awarding bodies' links.[20] Those who need to order modified papers for national curriculum tests can do so from the QCDA *Test orders* website (testorders.qca.org.uk).

According to the JCQ report, seven categories count as 'modified papers':

- A3 unmodified
- A4 modified 18-point bold
- A3 enlargement of A4 modified paper (approximately 24 point)
- braille papers
- modified language
- tactile diagrams
- transcript of listening test/video.

---

[18] JCQ (2009).

[19] Equality and Human Rights Commission link to key legislation: www.equalityhumanrights.com/uploaded_files/code_of_practice__revised__for_providers_of_post-16_education_and_related_services__dda_.pdf.

[20] Appropriate links are: www.aqa.org.uk, www.ccea.org.uk, www.edexcel.org.uk, www.ocr.org.uk and www.wjec.co.uk.

This project concentrates on only one category – modified language – because it is the modification that allows the regulator to research into the *intrinsic validity*[21] of non-standard papers. It is important to rethink the aim of paper modification in this broad sense, which is to remove unnecessary barriers to the comprehension of test and exam items.

The Test Review Group

In 2008 a Test Review Group was part of the QCDA test development process and comprised a group of practitioners who looked at how questions were presented. However, there was no consideration of the needs of particular sectors. According to the JCQ, there were 29 access arrangements available in 2009 for GCSE and GCE qualifications, with paper modification being one of these arrangements.[22] Yet, unlike other access arrangements, applications for modified papers should be made at the beginning of the course[23] because awarding bodies do not automatically modify all their papers. The Ofqual Monitoring Team did not know enough about current practice to implement fair access by design to the GCSEs. There was a need for information on paper modification.

Paper modification at QCDA

In spring 2007 the national curriculum assessments monitoring team began to work on focusing on the criteria of minimising bias. TDAs and inclusion group meetings raised a series of questions that the QCA would discuss further with the NAA. These meetings were organised by TDAs and were different from other test review meetings that were organised by the NAA. The TDAs' inclusion meetings relied on a panel composed of a mix of specialities:

- specialists in hearing impairment (HI)
- English as an additional language (EAL)
- SEN.

Other areas of expertise included: communication difficulties, psychological problems and dyslexia.

The main purposes of the TDAs' inclusion meetings were:

---

[21] For a definition of 'intrinsic validity' see Pollitt *et al.* (2008) *Improving the quality of GCSE assessment*. www.qcda.org.uk/libraryAssets/media/Improving_the_Quallity_of_GCSE_Assessment_final.pdf.

[22] JCQ (2009), pp. 4–26.

[23] Ibid, p. 3.

- to discuss the standard national curriculum assessments during their development to get feedback from experts about whether the test materials are accessible to pupils with SEN
- to ensure that test materials are as accessible as possible to pupils with specific assessment needs by gathering feedback on the materials during the development process.

The main outcomes from the inclusion meetings were that its records were shared with NAA and its first quality assurance pre-test reports would include a summary of any changes recommended by the TDAs based on feedback from the expert panels.

In 2008 NAA's inclusion panels looked at tests for the 2009 sessions (key stage 3 science – 22 June 2007; key stage 3 English – 11 July 2007; key stage 2 mathematics – 8 February 2008) which Sarah Lambert (Co-ordinator, QCA) attended. Later on, when Ofqual had been announced, Lambert's unpublished report *Ofqual report to NAA on observations of inclusion panel* summarised issues from each meeting. The *National curriculum assessments: Regulatory framework* outlined five common criteria that the NAA should consider in all processes throughout the delivery of assessments. One of these criteria was 'minimising bias' in the process and, specifically, the arrangements for inclusion groups. Specialists in EAL and HI had recommended that TDAs simplify language to avoid the inconsistencies sometimes found in national curriculum tests.

Previous reports

The Ofqual website has links to a collection of documents on paper design and modification such as a report by Nick Peacey and Lindy Peacey (2007).[24] This is a literature review commissioned by QCA to the Special Educational Needs Joint Initiative for Training (SENJIT) on attempts to minimise bias within national curriculum assessments at key stage 2 and key stage 3. It covered the history of the modifications for those with impairments/SEN. It also explained reasonable adjustments and their justifications in relation to the disability discrimination legislation. The authors suggested that further consideration should be given to:

- extra time as an adjustment
- areas such as autism, speech, language and communication and mental health which have had less attention in the past

[24] Peacey and Peacey (2007), *Minimising bias in assessment for students with special educational needs and disabled students: reasonable adjustments in written national curriculum tests at key stage 2 and key stage 3* www.ofqual.gov.uk/files/Minimising_bias_report.pdf.

- adjustments where an adult provides the interface to the presentation and response
- professional development for teachers implementing reasonable adjustments
- the government's commitment to accreditation for all SEN co-ordinators in schools.

They concluded that the regulator should strive towards making all its assessments as inclusive as possible, rather than emphasising 'access' through reasonable adjustments.

The document *Fair access by design*[25] gave guidance for awarding boards and regulators of external qualifications in England, Wales and Northern Ireland on designing inclusive GCSE and GCE qualifications to promote fair opportunities for all candidates. As the document explained, "Plain language and clear presentation in specifications and assessment materials are in the interest of every candidate. They promote fair access for all groups, including those with hearing, visual or learning difficulties. They also help where alternative means of communication – including technology – permit more candidates to show their knowledge and skills."[26]

Paper modification abroad

QCA's International Unit published a briefing on assessment arrangements for disabled students in 25 countries and a selection of states, territories and provinces.[27] The focus of the briefing was on assessment in secondary education and disabilities, with some references to dyslexia and primary education. The countries included in this briefing were: Australia, Canada, Czech Republic, Finland, France, Germany, Greece, Hungary, Iceland, Italy, Japan, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, New Zealand, Portugal, Singapore, Slovenia, South Africa, Spain, Sweden, Switzerland and the United States. The briefing included sections on control, terminology and mainstreaming (structured country by country) and changes to assessment arrangements (a comparative analysis). At the time "several countries were conscious of issues relating to equity and validity."[28] Equity was emphasised in terms of equality of opportunity. In Finland, for example, assessment methods enabled students to demonstrate their capabilities and

[25] See QCA (2005), *Fair access by design* at www.ofqual.gov.uk/files/pdf-05-1941-fair-access-design-final-v2.pdf.

[26] Ibid, p.4.

[27] Pepper, D. (2007), *Assessment for disabled students: an international comparison*, QCA briefing document, International Unit, Regulation & Standards Division, 25 September.

[28] Pepper (2007), p. 11.

performance as well as possible. In Singapore, special arrangements were designed to ensure students were not disadvantaged by factors irrelevant to the assessed constructs. In Australia, the purpose of special consideration procedures was to give all students the opportunity to show what they knew and could do.

Validity – the extent to which what is assessed really matches what one intends to assess, for example reading ability – was also emphasised, notably in regard to assessments of reading and writing. In Texas no assistance was permitted in the reading of test papers that assessed reading or writing. In Queensland there were restrictions on special considerations in reading and writing tests (eg the test was not produced in braille and the use of a scribe, reader, interpreter, dictionary or voice recognition technology was not permitted). In Australia, the states and territories had agreed that, in the interest of fairness to all students, certificates would show what the student did do, not what the student might have done in other circumstances. In Alberta, schools made requests for the use of specific accommodations and the ministry granted these requests on a case-by-case basis to assure both validity and fairness. In Hungary, the award of additional marks to disabled students was seen as providing a significant advantage over other students. Legislation in Norway and Slovenia said that accommodations should not lead to an 'unfair' advantage.

The United States

According to the United States Department of Education in 2002, the number of students with disabilities aged 6 to 17, served under the Individuals with Disabilities Act during the 2000/1 school year, was 5,496,329. The second largest category was that of students with speech and language impairment corresponding to 1,088,863 cases or 19.3 percent.[29] It is understandable therefore that since 1987 there has been an acute interest in paper modification in the United States. In the 1999 edition of the *Standards for educational and psychological testing*, the American Psychological Association pointed out the importance of the Centre for Equity and Excellence in Education Test Database.[30] The purpose of this database was to help test users to select suitable tests for their needs, especially when candidates were to sit high-stakes tests where the results define their progress in the educational ladder and job market. Today, over 50 educational and child-centred organisations in the United States and Canada have position statements about high-stakes use of standardised tests and many of them touch upon test design.

[29] Koenig, J.A. and Bachman, L.F. (eds) (2004), *Keeping score for all: The effects of inclusion and accommodation policies on large scale educational assessment* (National Academies Press, United States).

[30] Centre for Equity and Excellence in Education r3cc.ceee.gwu.edu/standards_assessments/EAC/HOME.HTM.

Educational Testing Service view

Researchers from the Educational Testing Service (ETS) proposed in 2003 an 'Evidence-centred assessment design (ECD) for individuals with disabilities' (Hansen, Mislevy and Steinberg, 2003). Their research question was: Does the accommodation change the construct? They also aimed to develop reusable structures and explore computer-based tools to validate test modifications and test accommodations. Their ECD approach involved domain analysis, domain model and conceptual assessment framework (CAF) based on:

- a student model (the learner characteristics that one wishes to assess)
- an evidence model (procedures for task scoring and for updating beliefs about student-model variables
- a task model (specifications for the task performance situation).

Hansen *et al.* (2005) later found that it was possible to integrate thinking about accessibility, task design and validity – all in a framework of sharable terminology, concepts and knowledge representations. The authors developed a model for reasoning about thousands of disability cases. Each case involved a specific student profile and a specific variant of a test.[31]

In 2005, Stephen Sireci, a psychometrician with many publications on accommodations and paper modification in the United States,[32] noted that test accommodations were often given to students with disabilities as one means of removing construct-irrelevant barriers to proper measurement of their knowledge, skills and abilities. Yet, the practice of test accommodation was still controversial because the effects of accommodations on test performance tended to be inconsistent due to:

- the wide variety of accommodations
- the various ways in which they were implemented
- the heterogeneity of students eligible to an accommodation.

However, Sireci, Scarpati and Li (2005) pinned down two consistent findings on extended time and oral accommodations:

---

[31] See Hansen *et al.* (2005), 'Accessibility of tests for individuals with disabilities within a validity framework', *System*, 33, p. 124.

[32] See Sireci (2001, 2002, 2003, 2008).

- extended time tended to improve the performance of all students, although students with disabilities tended to exhibit relatively greater score gains
- oral accommodations on mathematics tests were also associated with increased test performance for some students with disabilities.[33]

Addressing validity issues and test accommodations, Sireci (2008) argued that national assessments in the United States aimed to include all students in the sampling frame. Yet accommodations to improve measurement of knowledge, skills and abilities of students with disabilities produced "the ultimate psychometric oxymoron – an accommodated standardized test".[34] Looking at accommodations for reading tests, Sireci stressed how some types of accommodations on these tests were particularly controversial. He identified the most suitable occasions for providing accommodations and the best practice in reporting scores from accommodated test administrations. Key points in his findings were:

- read aloud and other oral accommodations to reading tests were likely to change the construct measured
- the principles of universal test design, which suggest building tests with greater content validity and more flexible administration conditions, should be considered for future development of reading tests
- both qualitative and quantitative approaches should be used to determine whether a particular test accommodation changes the construct measured
- testing agencies should develop clear definitions of the constructs measured on a test, as well as potential sources of construct-irrelevant variance; these definitions would help test users better evaluate the utility of the test and this would facilitate understandings of how accommodations could alter the construct
- ultimately, accommodation and score-reporting decisions should be made on a case-by-case basis.[35]

[33] Sireci, Scarpati and Li (2005), 'Test accommodations for students with disabilities: An analysis of the interaction hypothesis', *Review of educational research*, 75, No. 4, pp. 457–490.

[34] Sireci, S.G. (2008), 'Validity issues in accommodating reading tests', *Jurnal Pendidik dan Pendidikan*, 23, pp. 81–110, www.usm.my/education/publication/JPP23-5_VALIDITY%20ISSUES.pdf.

[35] Ibid, p. 105–6.

Israel

The National Institute for Testing & Evaluation (NITE) of Israel develops and administers psychometric entrance tests (PET) to institutions of higher education. NITE has a special unit with the responsibility of evaluating clinical diagnostic reports and providing test accommodations to applicants recognised as learning disabled. Most accommodations focus on modifications in the administration of the exam (for example, extended test time, extended breaks between chapters or enlarged test forms). When developing test items, NITE only considers the removal of excessive language burden when it improves the quality and clarity of the item for all examinees. NITE does not adapt test items to any specific population. Yet, NITE does provide test accommodations to applicants who were recognised as learning disabled. NITE researchers (Oren and Even, 2005) published papers on validity and fairness issues in granting test accommodations to students with language difficulties on the psychometric test. One of these works presented an alternative solution: the use of a computerised adaptive version of the PET. This alternative was adopted at NITE to provide test accommodations while maintaining the validity of the test.

The Netherlands

CITO, a leading testing and assessment company based in the Netherlands, also addresses issues on test modifications and accommodations. In its test at the end of primary education, there is a black and white version (for colour-blind students), a version in larger font, a spoken version and a version in braille. There is an external report (in Dutch) about the optimal font size for dyslexic students and this is the font size used in CITO's larger font version. In September 2008, Anton Béguin, the Director of CITO's Measurement and Research Department, explained in an interview[36] that research into the spoken version was very limited, but there was some research planned for the construction of test versions for students with learning disabilities. A main component of this research would be the identification of item drift (different response behaviour on specific items) between student classified with learning disabilities and other students.

[36] Interview with Anton Béguin by email with Ofqual's Fatima Carvalho in September 2008.

Brief description of research

Given the above justifications and background, it is clear that there are contrasting theories, models, beliefs, doubts and concerns surrounding paper modification. This project will address some key aspects of paper modification in the following way.

Stage 1: Ask for (or commission) a 8,000-word report describing how Pindar carries out paper modification in statutory national curriculum tests according to set specifications so that Ofqual can use the results of such a report to draw regulatory *principles of good testing design* from current practices in modified TDAs in England. This report will cover the principles and process of paper modification, as well as quality assurance, quality control, test takers' characteristics, validation framework and relation with regulator. The specifications for this report are set out in Appendix 1.

Stage 2: Demand (or commission) reports from the three main awarding boards – AQA, OCR and Edexcel – on the systems they use to modify general qualifications papers, according to set specifications. Such reports will focus on the systems awarding bodies use to modify general qualifications papers and those systems they are developing to use in future. According to JCQ some GCSE and GCE papers had been modified prior to being produced as the standard paper for all candidates. Yet, other papers had been modified on demand. The awarding bodies' reports will explain the differences in approach between universally modified papers and on-demand modified papers as well as the principles and theories guiding their practice of paper modification. Furthermore, awarding bodies' reports will also explain the validation methods used to sign off modified papers. Deliverables: three 8,000-word reports. The specifications for these reports are set out in Appendix 2.

Stage 3: Stimulate internal discussion on where paper modification sits within Ofqual through the establishment of the Ofqual Paper Modification Forum (OPMF) within the framework of the Diversity and Inclusion Strategy/Policy Group. Allocate one meeting of the forum to invite specialists in language modification (such as Mike Gutteridge, ESOL Examinations Consultant) to discuss technical aspects of paper modification. The OPMF will function as a focus group that will read the following documents:

- report from Pindar
- reports from awarding bodies

- report from experimental research commissioned by Ofqual.

The OPMF will meet three times (half-day sessions), with moderation by the main researcher. Guest specialists in paper modification will participate in two OPMF sessions and will help the forum to achieve consensus on the following questions:

- What are Ofqual's responsibilities in relation to paper modification (with a particular focus on national curriculum tests, GCSEs and A levels?
- What responsibilities on paper modification lie with QCDA and awarding bodies?

Deliverable: Minutes of three focus groups.

Stage 4: Consider results from the above functional and operational research merged with the results from additional experimental research on alternatives to test accomodation. This final report will review work that Ofqual/QCDA has developed on paper modification previously in the light of:

- how Pindar carries out paper modification in statutory national curriculum tests
- which systems awarding bodies use to modify general qualifications papers and which strategies they are developing for future use
- conclusions from the experimental research
- recommendations by the OPMF on the regulator's views on paper modification (with a particular focus on national curriculum tests, GCSEs and A levels)
- consensual view on which responsibilities regarding paper modification should remain with QCDA and awarding bodies.

Deliverable: A 9,000-word Ofqual report for the Ofqual Management Group on the current practices and operational aspects of paper modification in England, 2009/10.

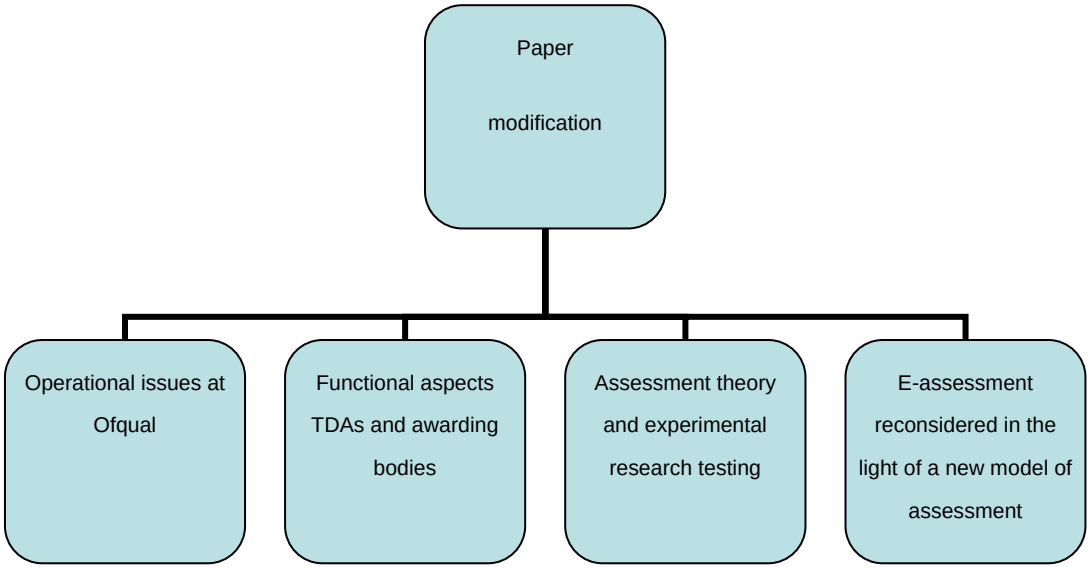
Outcome: A 40-page code of practice on paper modification, which will draw on a final 9,000-word report combining the research outcomes from Project A Current practices and operational aspects of paper modification in England 2009/10, and

Project B Alternatives to test accommodation: new models for pupils with speech and language difficulties.

The code of practice will have a section on the development of new modification strategies for all national tests and examinations.

Potential impacts of research

The impact of the research of Project A derives from its ability to connect different dimensions of paper modification:



According to the diagram above, Project A will not only collect credible information on current practices and operational aspects affecting the monitoring of paper modification, but it will also produce a final report connecting operational, functional, theoretical and experimental aspects of paper modification. Such a holistic approach to paper modification will be viable because by the end of Project A, its complement – Project B Alternatives to test accommodation: new models for pupils with speech and language difficulties – will have tested the potential of a new model of assessment for producing valid assessment for all test takers, with and without speech and language difficulties. The combination of two research projects on operational and experimental facets of paper modification will produce a credible statement connecting the practice of paper modification with a new assessment model that advocates enhanced educational experiences. That is to say, Ofqual will be able to support a new form of assessment, which possibly involves e-assessment, not as a mere innovation but as a tool for test fairness based on the theory of assessment.

What to do with research results

The research results from Projects A and B can be easily applied by Ofqual:

- the Ofqual code of practice on paper modification, containing operational, functional, theoretical and technological views on how to produce accessible papers, will help the regulator to support TDAs, awarding bodies and the assessment community because it will define the notion of ‘high-quality question papers’

- the views of the OPMF will help Ofqual to solve current operational uncertainties, for example where paper modification sits within the regulator and how paper modification is related to 'systems' and to 'qualifications issues'
- Ofqual may take advantage of Projects A and B to accelerate the uptake of e-assessment, acting as a mediator between e-assessment specialists and fairness stakeholders to support the full development of a new model of assessment by 2012, in the following way:

2010	**2011**	**2012**	**2013**
Implement the new models of assessment in the context of teacher assessment across 10 schools in the UK	Fully develop a version of the new model of assessment	Test a fully developed version of the new model of assessment using a large cohort – 200 key stage 2 pupils and 200 GCSE candidates with speech and language difficulties	Implement the new model of assessment as standard teacher assessment for pupils with speech and language difficulties across the UK and adopt the support model as the main paper modification procedure for GCSEs and GCEs

Method of investigation and timescale (42-day project)

Qualitative strategy – one of the two major approaches to research methodology in social sciences – involves document and bibliographic desk research to explain knowledge, attitudes and beliefs, etc. Qualitative methods may also involve in-depth understanding of human behaviour by means of different data collection methods – direct observation, interviews, focus groups, brainstorming, role play, simulation, and case study. In this project all data collected will be organised and analysed using a number of approaches and techniques for scrutiny. Such techniques include interpretation and development of arguments to explain the reasons behind paper modification.[37] Key bibliographic references will inform this research:

[37] For an approach to qualitative methodologies see the *Qualiti* website that is part of the ESRC National Centre for Research Methods (NCRM), based at Cardiff University. The mission of the NCRM is to provide a strategic focal point for the identification, development and delivery of an integrated national research, training and capacity building. www.cardiff.ac.uk/socsi/qualiti/index.html.

- Downing and Haladyna (2006, p. 5) identified 12 steps for effective test development[38] – this is a key reference for those working in paper design and modification, the model for test development is based on the *Standards for educational and psychological testing*[39] the intent of which is "to promote the sound and ethical use of tests and to provide a basis for evaluating the quality of testing practices"[40]
- the influence of test-taker background factors[41] (Kunnan, 1994, 1995 and 2000, Milanovic and Saville, 1996, Shohamy, 2000, Huhta, Kalaja and Pitkänen-Huhta, 2006) and the need to exclude irrelevant factors from test construct
- test development issues of validity, reliability, impact and practicality according to ESOL Research and Validation Group (Weir and Shaw, 2005, Taylor, 2005, Green and Jay, 2005, Khalifa, 2005 and Blackhurst, 2005).

[38] See Downing S. and Haladyna, T. (2006), *Handbook of test development* (Lawrence Erlbaum Associates London), Table 1.1., p. 5.

[39] See American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) (2004), *Standards for educational and psychological testing*.

[40] Ibid, p. 1.

[41] On variance introduced through test-taker background, see Kunnan (2000) and Shohamy (2000).

Phases of research on paper modification: operational perspective

The number of days of research from August–December 2009 are as follows.

Number of days	1–5 August	6–10 August	11–15 August	16–20 September	21–25 October	26–30 November	31–42 December 2009 – March 2010
Bibliographic search and acquisition (August 2009)	5						
Bibliographic review		5					
Contact QCDA and awarding bodies attending JCQ meetings			5				
Collect other data and opinion on paper modification (ESOL)				5			
Receive and review Pindar and awarding bodies' reports					5		
Confirm data collected from in-depth interviews						5	
Create, manage, moderate and write minutes of OPMF (September to February 2010)				15			
Data analysis and report writing (November to March 2010)						12	

Provisional key milestones for the projects are given below.

Key milestone (provisional)	**Date**
Formulated request for information on paper modification, TDAs contacted by QCDA	7 August 2009
Requests for information accepted	7 September 2009
Bibliographic search concluded	10 September 2009
Bibliographic review concluded	27 September 2009
Reports received from QCDA, OCR, AQA and Edexcel	30 October 2009
Data analysis concluded	30 December 2009
Management group OPMF debate concluded	20 February 2010
First draft of report concluded	28 February 2010
Second draft of report concluded	16 February 2010
Final report delivered	26 March 2010

Phases of research explained

Fatima Carvalho will spend 42 days completing this research project, which will involve the following activities:

- search and read bibliography (five days)
- collect information on paper modification in England (five days)
- interview Angela Hopkins, the head of test development in the former NAA, who will have an excellent overview of the potential issues and implications of paper modification from a practical viewpoint (one day)
- ask Angela Hopkins to contact Pindar to request a description of the practice of paper modification in statutory national curriculum tests

- participate in the JCQ monthly business meeting to present this project and request from each awarding board a report on their usual methods of paper modification (four days)
- organise the OPMF and moderate three halfday meetings to stimulate a debate on the following topics: ‘Where does paper modification sit within the regulator’s structure?’, ‘Should Ofqual monitor paper modification in future?’ and write Report A on main conclusions (14 days)
- write minutes of each OFMF meeting (one day)
- receive reports requested from awarding boards and Pindar on the current practices for paper modification and write report on the current state of paper modification in the national curriculum key stage 2 tests and general qualifications (GCSEs)
- write a 9,000-word report (combining information from test agencies and awarding bodies) and results from Project B Alternatives to test accommodation: new models for pupils with speech and language difficulties, to inform Ofqual of:
    - current practices of paper modification for the national curriculum tests, GCSEs and A levels
    - ongoing systems to monitor paper modification and assessment bias at Ofqual
    - how current practices and monitoring systems would be affected by the development of new assessment models (12 days).

Bibliography

Ahmed, A. and Pollitt, A. (2002), 'The support model for interactive assessment', Conference Paper, IAEA Hong Kong, September, see the Our Publications section of www.camexam.co.uk.

American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) (2004), *Standards for educational and psychological testing.*

Blackhurst, A. (2005), 'Listening, reading and writing on computer based and paper-based versions of IELTS', *Research notes*, 21, August, pp.14–17.

Braden, J. P. and Elliot, S. N. (2003), 'Accommodations on the Stanford-Binet intelligence scales', *Stanford-Binet intelligence scales,* fifth edition, Assessment bulletin number 2.

Downing, S.M. and Haladyna, T.M. (2006), *Handbook of test development*. (Lawrence Erlbaum Associates, London).

DCSF *Achievement and attainment tables 2008* website (formerly performance tables), www.dcsf.gov.uk/performancetables.

Downing, S. and Haladyna, T. (eds) (2006), *Handbook of test development*, (Lawrence Erlbaum Associates, London).

Elliott, S. N., Braden, J. P. and White, J. L. (2001), *Assessing one and all: Educational accountability for students with disabilities*. Arlington, VA: Council for Exceptional Children (ERIC Document Reproduction Service No. ED458746).

Equalities Review (2007), *Fairness and freedom: The final report of the Equalities Review*, archive.cabinetoffice.gov.uk/equalitiesreview.

Green, T. and Jay, D. (2005), 'Quality assurance and quality control: Reviewing and pre-testing examination material at Cambridge ESOL', *Research notes*, 21, August, pp. 5–7.

Hansen, E.G., Mislevy, R.J. and Steinberg, L. (2003), 'Symposium session: Assessment design and diverse learners: Evidentiary issues in disability, language, and non-uniform testing conditions', Chicago, Illinois. unpublished work. Hansen, E.G., Mislevy, R.J., Steinberg, L., Lee, M.J. and Forer, D.C. (2005), 'Accessibility of tests for individuals with disabilities within a validity framework', *System*, 33, pp. 107–133.

Huhta, A., Kalaja, P. and Pitkänen-Huhta, A. (2006), 'Discursive construction of a high-stakes test: the many faces of a test-taker', *Language testing*, 23, July, pp. 326–350.

Khalifa, H. (2005), 'Are test taker characteristics accounted for in main suite reading papers?', *Research notes*, 21, August, pp.7–10.

Koenig, J. A. and Bachman, L.F. (eds) (2004), *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessment*. National Academies Press. United States.

Kunnan, A. (1994), 'Modelling relationships among some test-taker characteristics and performance on EFL tests: an approach to construct validation', *Language testing*, 11, No. 3, pp. 225–250. ltj.sagepub.com/cgi/content/abstract/11/3/225.

Kunnan, A. (1995), *Test taker characteristics and test performance: A structural modelling approach*. University of Cambridge Local Examinations Syndicate, Cambridge University Press.

Kunnan, A. (ed.) (2000), *Fairness and validation in language assessment*. University of Cambridge Local Examinations Syndicate, Cambridge University Press

Linn, R. (2000), 'Assessment and accountability', *Educational researcher*, 29, No.2, pp. 4–16.

Milanovic, M. and Saville, N. (1996), 'Introduction. Performance testing, cognition and assessment', *Studies in language testing*, Volume 3 (Cambridge).

Ofqual (2008), *Oral language modifiers focus group*, February, Ofqual/08/3518, p. 6 www.ofqual.gov.uk/files/Ofqual-08-3518_Oral_language_modifiers_focus_group_11-08-08.pdf.

Ofqual (2009), *Statistics for access arrangements and special considerations at GCSE and A level: 2008*, March, Ofqual/09/4121, pp. 3–4 www.ofqual.gov.uk/files/Access-arrangements-and-special-consideration-report-2008.pdf

Oren, C. and Even, A. (2005), 'The fairness and validity of higher education selection systems for students with disabilities', *National Institute for Testing & Evaluation (NITE), NCME Annual Meeting, Montreal, Canada*, April.

Peacey N. and Peacey L. (2007), *Minimising bias in assessment for students with special educational needs and disabled students: reasonable adjustments in written national curriculum tests at key stage 2 and key stage 3* www.ofqual.gov.uk/files/Minimising_bias_report.pdf.

Pepper, D. (2007), *Assessment for disabled students: an international comparison*, QCA briefing document, International Unit, Regulation & Standards Division, 25 September.

Pollitt, A., Ahmed, A., Baird, J., Tognolini, J. and Davidson, M. (2008), *Improving the quality of GCSE assessment*, QCA, 2008.

QCA (2005), *Fair access by design*, www.ofqual.gov.uk/files/pdf-05-1941-fair-access-design-final-v2.pdf.

QCA (2007), *QCA review of GCSE and GCE access arrangements from 2004 to 2006*, November, QCA/07/3419.

Shohamy, E. (2000), 'The relationship between language testing and second language acquisition, revisited', *System*, 28, Issue 4, December, pp. 541–553.

Sireci, S. G. (2001), *Equating non-standard and standard administrations of the SAT*. December. Unpublished opinion paper submitted to the Blue Ribbon Panel on Flagging.

Sireci, S. G. (2003), *Unlabeling the disabled: A psychometric perspective on flagging scores from accommodated test administrations.* Center for Educational Assessment Research Report No. 502*.* Amherst, MA, School of Education, University of Massachusetts.

Sireci, S. G. (2008), 'Validity issues in accommodating reading tests', *Jurnal Pendidik dan Pendidikan*, 23, pp. 81–110, www.usm.my/education/publication/JPP23-5_VALIDITY%20ISSUES.pdf.

Sireci, S. G. and Gonzalez, E. J. (2003), *Evaluating the structural equivalence of tests used in international comparisons of educational achievement.* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois, 12–14 April.

Sireci, S. G. and Khaliq, S. N. (2002), *An analysis of the psychometric properties of dual language test forms.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, Los Angeles, 2–4 April.

Sireci, S. G., Li, S. and Scarpati, S. (2003), *The effects of test accommodations on test performance: A review of the literature.* Center for Educational Assessment Research Report No. 485. Amherst, MA, School of Education, University of Massachusetts.

Sireci, S. G., Scarpati, S. E. and Li, S. (2005), 'Test accommodations for students with disabilities: An analysis of the interaction hypothesis', *Review of Educational Research*, 75, No. 4, pp. 457–490.

Taylor, L. (2005), 'Using qualitative research methods in test development and validation', *Research notes*, 21, August, pp. 2–4.

Thurlow, M. L., Thompson, S.J. and Lazarus, S.S. (2006), 'Considerations for the administration of tests to special needs students' in Downing, S. and Haladyna, T. (eds), *Handbook of test development*, (Lawrence Erlbaum Associates, London).

Weir, C. and Shaw, S. (2005), 'Establishing the validity of Cambridge ESOL writing tests: towards the implementation of a socio-cognitive model for test validation'. *Research notes*, 21, August.

Appendix 1. Specifications for a report by QCDA

Length: 8,000 words

Focus: The process of paper modification in statutory national curriculum tests 2009

Section 1: Principles and Process: What is the theoretical framework Pindar uses for paper modification?

1.1 What Pindar does when it needs to cater for candidates with Special Needs such as Speech and Language Difficulties (SLDs) sitting the NC tests?

1.2 How your question paper production process looks like?

1.3 Do you follow a guide to paper modification? (Please enclose an example)

Section 2: Quality assurance: How Pindar manages the modified pre-test process?

2.1 Modified pre-testing process: how response validity is enhanced?

2.2 How Pindar targets potential modified pre-testing candidates for specific examinations?

2.3 How many pre-test, and trail papers of modified tests Pindar gets completed each year?

2.4 Can you provide examples of modified papers pre-test versions?

2.5 What Pindar does with pre-test feedback?

2.6 How Pindar manages activities and resources to improve benefits to Special Needs candidates?

2.7 Are pre-test candidates administered an anchor tests (made up of items of known difficulty) and background questionnaires? Can you provide us with one case study?

2.8 How do you estimate the difficulty of each modified pre-test version in relation to the established NC test Common Scale?

2.9 How do you analyse pre-test results?

Section 3: Quality control – Initial inspection of the test material.

3.1 What are the stages of your routine checks?

3.2 Pre-editing: when commissioned materials are initially submitted by item writers, how do you proceed in the first stages of the modified tests editing processes?

3.3 What your routine checks are intended to do when you carry out your initial inspection of the test material? (e.g. ensuring that all test materials meet the test specifications; suggesting appropriate changes to materials requiring amendments).

3.4 Which considerations do you give to the following features of modified test materials - topic, topicality, level of language, suitability for the task, length, focus of text, style of writing, focus of task, and level of task?

3.5 How do you check that modified papers meet intended standards?

3.6 Please explain your procedures to generate material of high quality.

3.7 What the usual procedures to judge this material against established standards for quality control? How do you use feedback results to refine the process?

3.8 What guidance do you give to item-writers on revising items and altering texts?

3.9 How do you define an unsuitable item type for candidates with SLDs? Can you give examples?

3.10 At the editing phase, how texts and selected items are scrutinised, and approved for pre-testing? Are items amended or sent back for further revision?

Section 4 Test taker characteristics

4.1 What test taker characteristics are taken into account at the modified papers pre-testing stage?

4.2 What did you do to remove construct irrelevant barriers to test performance while maintaining the integrity of the construct being measured by the NC tests 2009?

Section 5 Validation framework

5.1 What validation framework do you use for NC tests tasks?

Section 6 Relation with regulator

What is the input of regulatory code of practice in the development of modified papers?

What kind of guidance, at which level of detail, do you expect from the regulator so that test development agencies can develop more efficient paper modification?

What went wrong in the past?

What can be improved?

Appendix 2. Specifications for awarding bodies reports

Section 1 Awarding body regulations and guidelines on paper modification (when there is additional information to that published by JCQ)

Section 2 Theory underpinning exam paper modification in this organisation

2.1 Is a modified version of the standard test always designed for Special Education Needs (Deaf, Blind, Speech and Language Difficulties and other disabilities)?

2.2 Description of current practices on exam paper modification.

Section 3 Details on modified exam paper development

3.1 What is your systematic guidance for modified paper development activities in terms of (a) construct; (b) desired test interpretations; (c) test formats; (d) major sources of validity evidence; (e) purpose; (f) desired inferences; (g) psychometric model; (e) quality control.

3.2 Content definition: What is the essential source of content-related validity evidence (if different from content definition in the standard exam paper)

3.3 Exam paper specification: How do you develop a defensible sample of content domain and desired item characteristics (if different from standard exam paper)?

3.4 Item development: How do you develop effective stimuli and formats? Is there training for item writers and reviewers? How do you carry out effective item editing. How you detect construct irrelevant variance (CIV) flaws? Do you take notice of what are the recall strategies, alternative strategies, abstraction and spatial representation required by each item?

3.5 Modified paper designing and assembly: How do you select items for modified papers? How do you carry operational sampling? What pre-testing considerations you make? Are mark schemes designed in a different way? Do test developers consult unit databanks to ensure that there aren't existing tests which could have been used? Do modified paper developers follow a different rubric of the exam or do they decide on how questions will be presented?

3.6 Are modified paper developers aware of conventions for presentation (type face and size, use of headings and tables etc.) to aid readability.

3.7 Standards: how do you check that the standard of each paper is suitable for the range of candidates with Speech and Language Difficulties?

Section 4 Other validity questions

4.1 Do you carry out market research consulting labour market intelligence and evidence of learner demand for the modified papers?

4.2 How do you check whether the assessment pose extra burden to students?

4.3 Are modified test being developed by experts in the relevant subjects? Do they specialise in test development?

4.4 Are there clear guidelines on paper modification for examiners writing questions and mark schemes?

4.5 When writing modified papers do you consider what will be the cognitive resources used in the question answering process and what are the factors controlling students' ability to interpret questions?

4.6 Do you consider whether there will be problems in question papers for those students with limitations of working memory?

4.7 When test items are imbedded in real-life and meaningful contexts, do you consider whether these items pose problems for test takers?

4.8 Are paper developers familiar with the qualification or unit's learning outcomes

4.9 Are paper developers aware o f how many steps will be required to solve a question? Do test developers realise that a large number of steps can over-load working memory and information will likely to be lost?

4.10 How do you check whether the question paper discriminate effectively among candidates?

4.11 Are choices from optional questions offered? Do such optional questions make comparable demands on candidates?

4.12 What are your procedures to carry out a final validation of the modified paper and mark scheme?

4.13 Do you submit you modified items to a question databank in the format required by the qualifications and assessment regulators?

Section 5 Reliability

5.1 Is the reliability of the modified paper questions statistically checked?

5.2 Which statistical analyses are carried out (e.g. Rasch Measurement Model - probability of success on a question depends upon two variables: the difficulty of the question and the ability of the candidate).

Section 6 Fairness

6.1 Are pre-test results checked by an Equalities Panel?

Section 7 Future Developments in Paper Modification

7.1 Can you describe and enclose any recent research on the future paper modification?

7.2 In future, how paper modification will be dealt with by your organisation?

7.3 How do you intend to integrate technological innovation with paper modification?

Current practices and operational aspects of paper modification in England 2009/10

First published by The Office of the Qualifications and Examinations Regulator in 2009.

© Qualifications and Curriculum Authority 2009

Ofqual is part of the Qualifications and Curriculum Authority (QCA). QCA is an exempt charity under Schedule 2 of the Charities Act 1993.

Reproduction, storage or translation, in any form or by any means, of this publication is prohibited without prior written permission of the publisher, unless within the terms of the Copyright Licensing Agency. Excerpts may be reproduced for the purpose of research, private study, criticism or review, or by educational institutions solely for education purposes, without permission, provided full acknowledgement is given.

Office of the Qualifications and Examinations Regulator
Spring Place
Coventry Business Park
Herald Avenue
Coventry CV5 6UB

Telephone 0300 303 3344
Textphone 0300 303 3345
Helpline 0300 303 3346

www.ofqual.gov.uk