

The Reliability Programme

Final report of the Policy Advisory Group



Sandra Burslem

March 2011

Ofqual/11/4829

Contents

| | |
|--|----|
| Summary..... | 3 |
| Introduction..... | 3 |
| Research on Strand 3 | 4 |
| Surveys of perceptions of A Levels and GCSEs..... | 5 |
| Qualitative investigations of perceptions of reliability..... | 5 |
| Quantitative investigation of perceptions of reliability | 8 |
| Communicating about reliability with the public | 9 |
| Engaging with key stakeholders in debating issues related to reliability | 10 |
| Activities undertaken by PAG | 10 |
| Stakeholder perspectives on reliability..... | 10 |
| Ways of understanding and communicating reliability evidence | 11 |
| Handling media stories | 11 |
| Ways of improving public understanding of reliability and increasing public confidence | 12 |
| Implications of findings from the Reliability Programme | 14 |
| Discussions on TAG recommendations | 14 |
| Recommendations from PAG..... | 15 |
| References | 17 |
| Appendix A: Members of the Ofqual Reliability Programme Policy Advisory Group.. | 19 |

Summary

The Reliability Programme Policy Advisory Group (PAG) was appointed primarily to provide support to Strand 3 of the programme – investigating public perceptions of reliability and developing regulatory policy on reliability. The group held two meetings during the second year of the programme to gain a full understanding of the progress made on the programme and to discuss the implications of the findings from the programme for Ofqual and other main stakeholders. In particular, the group's work has focused on exploring ways to improve public understanding of reliability concepts, communicate reliability evidence to the public and increase public confidence in the examinations system, discussing the adequacy and appropriateness of the recommendations proposed by the Reliability Programme Technical Advisory Group (TAG), and making recommendations. The group has made the following recommendations to Ofqual:

- continue work on reliability
- publish reliability reports commissioned
- encourage awarding organisations to generate and publish reliability data
- set up a programme to improve public understanding of reliability and increase public confidence in the examinations system:
 - make technical terms plain so that the public can understand
 - enable the public to understand the assessment process
 - explain factors affecting assessment outcomes and factors that can introduce inconsistency in test scores
 - help the public to interpret reliability evidence
 - engage with the main stakeholders.

Introduction

The Reliability Programme that Ofqual has been undertaking looks at the reliability of results from National Curriculum assessments, public examinations and vocational qualifications in order to develop regulatory policy on reliability (see Opposs and He, 2011). The programme was structured into three strands:

- Strand 1: Generating evidence on the reliability of results from a selection of national qualifications, examinations and other assessments in England through empirical studies.

- Strand 2: Interpreting and communicating evidence on reliability.
- Strand 3: Investigating public perceptions of reliability and developing regulatory policy on reliability.

Two advisory groups were appointed to provide support to the programme:

- The Technical Advisory Group (TAG): This group is made up of educational assessment experts and has been advising work on Strands 1 and 2, including advice on the methodologies to be used and the selection of qualifications, examinations and other assessments to be investigated. They were also responsible for reviewing reports from research projects funded under the Reliability Programme.
- The Policy Advisory Group: This group was appointed to provide advice on work for Strand 3, and to a lesser extent, Strand 2 of the programme (see Appendix A for members of the group). The Policy Advisory Group is made up of representatives from a wide range of stakeholders, including assessment experts, assessment providers, employers, communications experts, teachers, students and parents.

Specifically, PAG has been asked to advise on:

- stakeholder perspectives on the reliability of results from England's test, examination and qualification systems
- ways to understand reliability evidence and communicate it to a non-technical audience
- implications of findings from the programme
- ways to deal with media stories that misinterpret or inappropriately communicate reliability statistics
- ways to improve public understanding of reliability and increase public confidence in the examinations system
- adequacy and appropriateness of recommendations from TAG.

This report provides a brief summary of the work on Strand 3 of the programme, the activities that have been undertaken by PAG and its recommendations to Ofqual.

Research on Strand 3

There is considerable variability in how measurement uncertainty is reported in different parts of the world (Bradshaw and Wheater, 2009). For example, while in the

USA and other countries, test results are sometimes reported as raw scores or scaled scores together with the associated standard error of measurement (Bradshaw and Wheater, 2009; Phelps et al., 2010), in England test and assessment organisations tend to report students' performance levels or grades for public tests and examinations without any indication at all of the likely error-rates involved. However, there has been suggestion that there is a duty to communicate the reliability of results to the public (see, for example, AERA et al., 1999, Standard 2.1; Newton, 2005a, b; Phelps et al., 2010). Newton (2005a, b) discussed whether and how assessment organisations should communicate with the public about measurement inaccuracy. There has been debate about the implications of reporting reliability or measurement error in examination results for public confidence in the examination system (Newton, 2005a, b; Ofqual, 2009).

It is essential to understand the public's knowledge of and attitudes towards unreliability in examination results when developing regulatory policy on reliability in order to improve the quality of the national examinations system and increase public confidence in the system.

Surveys of perceptions of A Levels and GCSEs

Ipsos MORI conducts a survey of perceptions of A level and GCSE that is now in its eighth wave (Ipsos MORI, 2010). The most recently reported wave of the survey was conducted in the winter of 2009, and reported findings based on samples of A level and GCSE teachers, A level and GCSE students and their parents, and the general public. Both the 2009 and the 2010 surveys included questions about the reliability and accuracy of examination results. Large majorities of teachers, parents and students thought that most or all students got the correct grade at GCSEs and A levels. Respondents also gave reasons that they perceived as being likely to cause candidates to get the wrong grade in examinations, which included:

- students performing better or worse than expected in examinations or coursework
- inaccurate marking and poorly designed examination papers.

Qualitative investigations of perceptions of reliability

As part of Strand 3 of the Reliability Programme, research projects were commissioned from Ipsos MORI and from the Assessment and Qualifications Alliance (AQA) to investigate public perceptions of reliability using workshops and focus groups. These studies focused on the following aspects of reliability:

- the assessment process
- factors affecting the performances of students on examinations

- the reliability concepts and measurement error
- the different types of error in examination results: preventable human mistakes versus inevitable random measurement error
- factors contributing to measurement error in examination results
- the level of acceptance towards human error and measurement error in examination results.

The research conducted by Ipsos MORI in January 2009 used two workshops in London and Birmingham to investigate the opinions of different groups about reliability and unreliability (Ipsos MORI, 2009). Research participants were drawn from teachers, students, parents, members of the general public, employers and examiners. The sessions started with an analogy to an error occurring in medical treatment; this was used as a substantial input to help workshop participants understand the concepts under discussion. Researchers understood that giving such substantial input to participants whose opinions and attitudes one was trying to discover ran the risk of biasing them. However, the belief was that participants would probably not have developed views on reliability in test scores and so it was felt important to give them contextualisation of this sort. Some of the findings from the research include:

- There was a demarcation in the minds of the public between inevitable errors in the assessment process and preventable errors. The research participants appeared to accept that a certain amount of error was inevitable in a large examinations system, but they could be intolerant of 'preventable errors'.
- Sometimes participants appeared to be making a distinction between inherent and preventable error, and at other times not.
- Some research participants stated that their attitude to error depended upon whether the error changed a student's grade or mark. They considered grade-related error to be more consequential than mark-related.
- Participants' views about error could vary by group and by the perceived cause of the error. For example, students and teachers could be intolerant of typographical errors in papers, while examiners could be more sanguine, taking the view that what was important was that any mistakes that did occur were rectified.
- There was evidence that students were aware that some inconsistency between human markers was inherent in subjects such as English. However, there were also statements that such inherent error should be minimised or even eliminated.

- Students and the general public were able to debate whether and how examinations can and should sample from the curricula.

Chamberlain from AQA (2010) conducted qualitative research to follow up Ipsos MORI's (2009) work. Chamberlain collected data via ten focus groups, with samples of job-seekers, employees, employers, students for Postgraduate Certificate in Education (PGCE), and primary and secondary teachers (74 participants in total; 28 male and 46 female). Like Ipsos MORI, Chamberlain designed her research with the assumption that she would have to take steps to mitigate participants' lack of knowledge of key elements of the reliability concept. Chamberlain used vignettes as a technique to introduce reliability to her research participants. The vignettes were very short stories or scenarios involving fictional characters in specific dilemmas which were related to the research context and relevant to the lives and educational experiences of the participants. Main findings from the study include (Chamberlain, 2010):

- *With the exception of the secondary school teachers, the participants had limited awareness of the concept of reliability. Participants were able to recognise forms of human error in the assessment process but often failed to envisage how this might impact on the reliability of their assessment outcomes.*
- *The participants struggled to see how measurement inaccuracy (Newton, 2005) could be termed 'assessment error' and how it could impact on the reliability of outcomes. Instead, they suggested that measurement inaccuracy was an inevitable part of life, and that to draw attention to its impact on assessment outcomes would not be beneficial.*
- *The participants had rarely questioned the reliability of the assessment process or their assessment outcomes, and showed a significant amount of trust in the system to award them the 'right' outcomes. Some participants had experiences of re-marks or appeals. This appeared to make them more questioning of the accuracy of their results than other participants, but seemed to do little to undermine their trust in the assessment system as a whole. The secondary school teachers spoke extensively about their experiences of challenging students' results, and demonstrated their awareness of how errors could occur.*
- *Participants tended to trust examiners to assess their work fairly, believing that examiners are professional and well trained subject experts. The participants could recognise, however, that some subjects require more interpretation than others, and thus that the reliability of marking could be variable. The secondary school teachers tended to be less trusting – many acted as moderators themselves in order to mediate the influence of external examiners, and to gain a better understanding of assessment criteria to pass on to their students.*

- *On the whole, the participants suggested that they would like to be more informed about assessment reliability, but only through a better understanding of how the assessment process works i.e. knowing what happens to a candidate's script after the candidate has completed the examination. There was a notable lack of support for any quantification of reliability and, in particular, publishing a reliability statistic alongside a candidate's grade. The secondary school teachers were particularly emphatic that any initiative to enhance understanding of reliability should begin with teachers and students, and not with parents or the public at large.*

Results from the research indicated that reliability is a difficult concept to comprehend. The author suggested that 'a qualitative approach to reliability that focuses on students and teachers may be a possible way forward in enhancing the dissemination of reliability information' (Chamberlain, 2010).

Quantitative investigation of perceptions of reliability

Based on findings from the qualitative studies involving the use of workshops and focus groups by Ipsos MORI (2009) and Chamberlain (2010), a further quantitative study was conducted on public perceptions of reliability using an online questionnaire survey (He, Opposs and Boyle, 2010). The questionnaire was structured into five distinctive topics to measure different aspects of respondents' knowledge of and attitudes towards unreliability in examination results:

- knowledge of and experience in the examination process and confidence in the national examinations system
- understanding of factors that affect the performances of students on examinations and factors that introduce uncertainty into examination scores
- attitudes towards different types of assessment error (including human mistakes and measurement inaccuracy)
- approaches for improving reliability
- approaches to trust in general.

Respondents were sampled from three key stakeholder groups: A level teachers, A level students aged 16–18, and employers. Data collected is also used to investigate:

- how attitudes to unreliability are related to knowledge and understanding of the reliability concept
- how attitudes to unreliability are related to confidence and belief in the examination system and approaches to trust

- how confidence and belief in the examination system are related to trust.

Main findings from the study are as follows:

- There was substantial variability in the understanding of reliability concepts and attitudes to unreliability in examination results among the respondents, both within group and between groups.
- The majority of the respondents from the three groups appeared to understand the assessment process and the factors that affect students' performances on examinations.
- To a degree, the respondents also understood the factors that could introduce uncertainty in examination results.
- The respondents showed various degrees of experiences of the examination process and acceptance of measurement error in examination results.

The level of tolerance of the respondents for measurement uncertainty to some degree was positively correlated to the level of belief about the examination system, knowledge of aspects of unreliability and approaches to trust.

Communicating about reliability with the public

Boyle, Opposs and Kinsella (2009) conducted research looking at issues with communicating unreliability in test scores to the public. These researchers suggested two reasons for the difficulty in communicating the reliability concepts with the public:

- the concept of reliability is complex and hard to explain succinctly
- unreliability seems like an intrinsically bad news story.

They cited two sources of evidence for these reasons:

- Firstly, literature describing the media environment that surrounds examination results in England is summarised, which gives a history of assessment organisations' attempts at communicating with the public and is used to make suggestions for how such bodies might communicate better.
- The second source of evidence is the findings from the 2009 Ipsos MORI work (2009) discussed above, which provides the researchers with an initial feel for the tolerance that different sectors of the public have for different sources of measurement inaccuracy in examination results.

The researchers then conclude by suggesting ways to improve each of the issues with unreliability as a media story:

- The problem of complexity is addressed by allowing people to interact with the message via multiple media, using varied analogies and so on.
- In terms of the negativity of the story, the suggested response is not to try to make this into a good news story. Rather, the aspiration is to communicate the message that many assessment results contain an element of unreliability to the public in a manner that allows people to become more sophisticated users of those results.

Engaging with key stakeholders in debating issues related to reliability

Ofqual has held discussion groups at various public events to raise public awareness of reliability and to engage with the public to discuss reliability related issues.

Activities undertaken by PAG

The group met twice during the second year of the programme, once on 10th March 10th, 2010 and the other on 10th November 2010. The main objectives of these meetings were to:

- gain full understanding of the progress of the Reliability Programme
- understand stakeholder perspectives on the reliability of results from England's test, examination and qualification systems
- explore ways to understand reliability evidence and communicate it to a non-technical audience
- discuss implications of findings from the programme
- discuss ways of dealing with media stories that misinterpret or inappropriately communicate reliability statistics
- explore ways to improve public understanding of reliability and increase public confidence in the examinations system
- discuss the adequacy and appropriateness of recommendations from TAG
- make recommendations to Ofqual.

Stakeholder perspectives on reliability

Results from the various studies on public perceptions of unreliability in examination results indicated that the public generally understood the assessment process but only had limited knowledge about reliability concepts. They also seemed to understand some of the factors that could affect test or examination results and

sometimes make a distinction between inconsistency in results caused by human mistakes and that which is intrinsic to the measurement process. Some of the public also understood the level of inconsistency in test scores could vary between assessments. The level of acceptance for unreliability in examination results varied between different stakeholder groups.

One of the problems that makes it difficult for the public to understand the concept of reliability is that the technical meanings of the words such as “reliability/reliable” and “error” in the context of education assessment are different from their meanings in daily use. Although the daily use of ‘reliability’ is generally associated with the idea of consistency of occurrence of the same event, it is constantly implied as an ‘either/or’ concept. However different levels of reliability can exist in results from educational assessments. The word ‘error’ in its daily use is constantly associated with human mistakes, while its technical meaning in educational measurement implies deviation of scores on a test from some notional number when the measurement procedure is repeated.

The fact that most candidates only take the same examination once might also make it difficult to associate examination results with the concept of reliability.

Ways of understanding and communicating reliability evidence

It was recognised that the reliability concept is difficult for the public to comprehend. Although there is ample literature on the themes of reliability, most of it is quite technical. Education was seen to be the key for the public to understand reliability concepts in the context of educational assessment. The use of plain language and examples to explain the assessment process, the meaning of technical terms, the factors that affect test scores and factors that introduce inconsistency in test scores would be useful. The use of analogies from other disciplines, such as medical diagnosis, might also be useful. See also later discussions.

Handling media stories

As a regulator, Ofqual needs to be open and transparent, and this includes publishing results from research projects it commissions. Further, one of the aims of publishing reports under the Reliability Programme was to raise public awareness of reliability and stimulate national debate. However this has drawn a few negative headlines from the media, involving misinterpretation or inappropriate communication of reliability statistics, which could undermine public confidence in the assessment system rather than generate debate about reliability, although some of the headlines generated were also a natural reaction from the media. Whilst it was realised that the media always want to generate headlines that are interesting to their audience, Ofqual needs to develop a media handling strategy to alleviate the impact from such negative headlines on public confidence. This would include explaining the purposes of the Reliability Programme, making technical terms plain so that the general public

can understand easily, setting appropriate context for interpreting findings from the research, and expressing Ofqual's views.

It was generally agreed that some further explanations of the results that are being reported would be required. Particularly, the context within which the reliability information is interpreted should be clearly set in plain language, since what technical experts may know the less well informed general public may not. It was realised that some sentences from a report might be picked up and used by the media as negative headlines. It was suggested that reports could be reviewed by someone outside of Ofqual before publication to identify places where potential negative headlines may lie.

It was recognised that while the Reliability Programme would enable Ofqual to have a better understanding of the reliability in results from public test and examinations and that the programme only started to collect evidence, as a public body Ofqual has the duty to communicate with the public.

It was realised that these negative headlines might undermine the confidence of the public in the qualifications system. But once the public fully understand the concept of reliability and the associated issues, their confidence can be increased. Educating the media about reliability might also be required.

Ways of improving public understanding of reliability and increasing public confidence

There was a consensus that since the general public has only limited knowledge about reliability, engaging with the public and education would be the key to improving public understanding of reliability concepts. This may be a long-term process and could involve:

- Explaining technical terms using layperson language so that the public can understand. There is too much packed into the terms like assessment, measurement, reliability and the associated various indices (Cronbach's alpha coefficient, standard error of measurement, classification accuracy, and so on), random error, system error, human error, and others. These need to be unpacked for the general public to comprehend.
- Enabling the public to understand the assessment process.
- Enabling the public to understand the many factors that can influence the performance of a test-taker on a test and consistency of test scores under repeated measurements. These factors would include the particular question paper that the test-taker took on the day of the examination, the particular day of the examination and the particular examiner who happened to mark his/her script, in addition to his/her actual ability in the subject area being tested.

Variability in some of the factors will inevitably exist and is intrinsic to any assessment system when the measurement procedure is repeated and this will result in inconsistency of test scores, although the degree of such variability may be reduced to some degree.

- Enabling the public to understand that the level of inconsistency in test or examination scores from repeated measurements will be inevitable and will vary from assessment/subject to assessment/subject. This is because the level of control on the factors that can introduce inconsistency in test scores varies between assessments/subjects. As tests and examinations normally sample contents and skills from the entire curricula, different areas will be covered in different tests or examinations. Assessments use tasks of different formats to assess different types of knowledge and skills to ensure validity, and some tasks can be marked more consistently than others. Although awarding organisations (AOs) try to improve assessment reliability as much as they can (for example through improving quality of question papers and marker training), there are however certain limitations on what can be done to improve reliability. Improving reliability should not compromise assessment validity.
- Helping the public to make sensible interpretation of reliability evidence, which would involve clearly setting the context for interpretation and explaining the meaning of the numbers associated with reliability indices. For example, it would be helpful to explain what it is meant by Cronbach's alpha being 0.90 or classification accuracy being 95 per cent for outcomes from a specific assessment and how these figures would change for different populations, different assessments, and different grading systems. It would also be useful to make it clear that most reliability indices are for a group of candidates not for individuals.

To increase public confidence in the examinations system, Ofqual needs to make clear that:

- What Ofqual does is to ensure the quality of the assessments and qualifications it regulates and to safeguard learners' interests.
- Reliability is a complicated abstract concept, and reliability measure is only one indicator of the quality of an assessment. Reliability can vary from assessment to assessment. Interpretation of reliability measures requires an understanding of the concept of reliability and the nature of the assessment.
- The Reliability Programme aims to gain a better understanding of the reliability of results from assessments in England in order to improve the quality of the qualifications systems further.

Ofqual ensures that awarding organisations have appropriate procedures in place to ensure assessment reliability.

Implications of findings from the Reliability Programme

The Reliability Programme has made substantial progress in the following areas:

- Generating evidence of reliability in results from a range of tests, examinations, and other qualifications in England.
- Interpreting and communicating reliability evidence.
- Exploring public perceptions of reliability of examination results.

Findings from the programme have provided important information on the reliability of results from a range of assessments and how reliability is understood by both assessment professionals and the general public. This has put Ofqual in a position to develop regulatory policy on reliability for the assessments it regulates in order to improve their quality further, and to develop approaches for improving public understanding of reliability and increasing public confidence in the national qualifications system.

Discussions on TAG recommendations

It was realised that although plain language was important for improving public understanding of reliability concepts, there was still a need for technical language to be used in the technical arena. Based on the work conducted under the Reliability Programme, studies carried out elsewhere and practices adopted in other countries, the Technical Advisory Group (TAG) had proposed a series of recommendations for potential reliability policy (see Baird et al., 2011). Some of the recommendations were discussed by PAG, including:

- Where possible, reliability statistics for the qualification as a whole should be published. It was realised that this could be difficult in practice for qualifications consisting of modular units or components.
- Reliability statistics should be collected for a range of assessment types. A web-based database could be created to collect reliability statistics from different assessment types.
- All awarding organisations should document and publish their standard-setting practices. It was realised that some awarding organisations have already done this and there has not been much work carried out in this area under the Reliability Programme.

- Suites of qualifications should have a minimum level of reliability at subject level. It was realised that some awarding organisations may not have the necessary technical expertise to implement this.
- Having greater control of assessment formats in workplace-based assessments. Workplace-based assessments present a variety of issues such as observations of live performance, the use of a variety of tasks, individualised skills internally assessed, unlimited attempts by candidates, and others, which could introduce inconsistency in results. The introduction of some standardised tasks could exert more control on variability. It was however realised that in National Vocational Qualifications (NVQs), there are a lot of controls in place to ensure that assessments are carried out in a consistent way.

It was noticed that teacher assessment (TA) is an important area where little work has been done. It was realised that if teacher assessment, including coursework and controlled assessment, is carried out to a high standard the reliability of qualifications as a whole would be enhanced.

Recommendations from PAG

Based on analyses of the results from the programme and recommendations made by TAG, the group proposed the following recommendations to Ofqual:

- To continue work on reliability. Although the group realised that substantial progress has been made by the programme, further work would be beneficial. Work in the area of teacher assessment, workplace-based assessment, and construct validity of assessment would be of particular interest and importance.
- To publish reliability reports already commissioned. The various reports produced under the programme represent a very useful resource for a range of audiences, including researchers from academic institutions and awarding organisations, policy-makers and educational practitioners.
- To encourage awarding organisations to generate and publish reliability data.
- To set up a programme to improve public understanding of reliability and increase public confidence in the examinations system, by working with the awarding organisations to:
 - Make technical terms plain so that people can understand.
 - Enable the public to understand the assessment process.

- Explain factors affecting assessment outcomes and factors that can introduce inconsistency in test scores.
- Help the public to interpret reliability evidence.
- Engage with main stakeholders, maybe starting with teachers and students in schools. Other stakeholders such as parents, employers, local education authorities and training agencies could also later be involved.
- Enable the public to understand that the Ofqual Reliability Programme investigates the reliability of assessment outcomes and aims to develop regulatory policy on reliability in order to improve the national examinations system further.

References

- Baird, J., Black, P., Bèguin, A., Pollitt, A. and Stanley, G. (2011). *The Reliability Programme: Final report of the Technical Advisory Group*. Coventry, UK, Ofqual. Available online at: www.ofqual.gov.uk/research-and-statistics/research-reports/92-articles/20-reliability.
- Boyle, A., Opposs, D. and Kinsella, A. (2009) 'No News is Good News? Talking to the public about the reliability of assessment'. Paper presented at the 35th International Association for Educational Assessment (IAEA) Annual Conference in Brisbane, Australia, 13–18 September, 2009. Available online at: www.ofqual.gov.uk/files/2009-09-iaea-no-news-is-good-news.pdf.
- Bradshaw, J. and Wheeler, R. (2010) *International Survey of Results Reporting*. Coventry, UK, Ofqual. Available online at: [www.ofqual.gov.uk/files/Ofqual_10_4705_International_Survey_of_Results_Reporting_08_03_10_\(2\).pdf](http://www.ofqual.gov.uk/files/Ofqual_10_4705_International_Survey_of_Results_Reporting_08_03_10_(2).pdf).
- Chamberlain, S. (2010) *Public Perceptions of Reliability*. Coventry, UK, Ofqual. Available online at: www.ofqual.gov.uk/files/Ofqual_10_4708_public_perceptions_reliability_report_08_03_10.pdf.
- He, Q., Opposs, D. and Boyle, A. (2010) *A Quantitative Investigation into Public perceptions of Reliability in Examination Results in England*. Coventry, UK, Ofqual. Available online at: www.ofqual.gov.uk/files/2010-10-20-public-perceptions-of-reliability.pdf.
- Ipsos MORI (2009) *Public Perceptions of Reliability in Examinations*. Coventry, UK, Ofqual. Available online at: www.ofqual.gov.uk/files/2009-05-14_public_perceptions_of_reliability.pdf.
- Ipsos MORI (2010) *Perceptions of A levels and GCSEs – Wave 8*. Available online at: www.ofqual.gov.uk/files/2010-03-ofqual-perceptions-of-a-levels-and-gcses-wave-8.pdf.
- Newton, P.E. (2005a) 'The Public Understanding of Measurement Error'. *British Education Research Journal*, 31, 419–42.
- Newton, P.E. (2005b) 'Threats to Professional Understanding of Assessment Error'. *Journal of Education Policy*, 20, 457–83.
- Ofqual (2009) *The Reliability Programme – Technical seminar report*. Coventry, UK, Ofqual. Available online at: www.ofqual.gov.uk/research-and-statistics/research-reports/92-articles/20-reliability.

Opposs, D. and Q. He (2011) *The Reliability Programme: Final report from*. Coventry, UK, Ofqual. Available online at: www.ofqual.gov.uk/research-and-statistics/research-reports/92-articles/20-reliability.

Phelps, R., Zenisky, A., Hambleton, R. and Sireci, S. (2010) *On the Reporting of Measurement Uncertainty and Reliability for U.S. Educational and Licensure Tests*. Ofqual, Coventry, UK. Available online at: www.ofqual.gov.uk/research-and-statistics/research-reports.

Appendix A: Members of the Ofqual Reliability Programme Policy Advisory Group

| | |
|---------------------|---|
| Jo-Anne Baird | Professor of Educational Assessment, also Chair of Ofqual Reliability Programme Technical Advisory Group |
| Mike Baker | Journalist |
| Paul Black | Professor in Educational Assessment, also member of Ofqual Reliability Programme Technical Advisory Group |
| Dame Sandra Burslem | Deputy Chair of Ofqual |
| David Butler | Chief Executive, National Confederation of Parent–Teacher Association (NCPTA) |
| Jackie Dean | Technical & Professional Development Consultant, Seven Trent Water |
| Dr Ken Flint | Admissions Tutor, Warwick University |
| Anysha Ihuomah | English Secondary Students Association (ESSA) |
| Jessie Seal | English Secondary Students Association (ESSA) |
| Edward Winter | Professor of the Physiology of Exercise |

We wish to make our publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2011

© Crown copyright 2011

| | |
|--|-------------------|
| 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