A Quantitative Investigation into Public Perceptions of Reliability in Examination Results in England

October 2010

Ofqual/10/4769

Qingping He and Dennis Opposs (Ofqual) and Andrew Boyle (The City and Guilds of London Institute)

Contents

Executive summary

The Office of Qualifications and Examinations Regulation (Ofqual) in England is carrying out a two-year research programme investigating the reliability of results from national tests and public examinations. Strand three of the programme is to gauge public perceptions of unreliability in examination results. Based on findings from previous qualitative studies involving the use of workshops and focus groups, a further quantitative study was conducted on public perceptions of reliability using an online questionnaire survey. 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.

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.

Respondents were sampled from three key stakeholder groups:

- A level teachers
- A level students aged 16–18
- employers.

The main findings from the study are the following:

- 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.

Background

It has been argued that England is a country in which much educational testing takes place (Black and Wiliam, 2005). For example, there are or have been the following major testing occasions in the English system: whole cohort national curriculum tests (NCTs) across various subjects and at various ages, public examinations – including standardised qualifications typically taken at the age of 16 and 18, and large and diverse suites of vocational qualifications, which may be taken by people at all ages in schools, further education and in workplace as part of on-the-job training. Some assessment systems (such as the NCTs and the 16-plus General Certificate of Secondary Education examinations or GCSEs) produce data that is also used to provide public evaluations of institutions and individual professionals, in addition to providing information about individual students' attainments in specific subject areas.

Reliability, in educational measurement terms, refers to the consistency of results on a given measure from repeated measurements under equivalent conditions and is an important indicator of the quality of an assessment. However, there has been little large-scale research to monitor the reliability of results from England's test and examination system and little understanding of the public's knowledge of and attitudes towards unreliability in assessment results. To address this, Ofqual is conducting a two-year research programme. The primary aim of this programme is to gather evidence to inform Ofqual on developing regulatory policy on reliability. The programme is structured into three strands:

- Strand one: Generating evidence of reliability of results from a selection of national qualifications, examinations and tests in England through empirical studies.
- Strand two: Interpreting and communicating evidence of reliability through reviewing measurement theories and models that are used to study reliability and techniques that are used to produce and interpret reliability indices and organising seminars involving leading assessment experts and communications experts to discuss effective ways to communicate reliability information to key stakeholders.
- Strand three: Exploring public perceptions of reliability and developing regulatory policy on reliability through questionnaire surveys, workshops and focus groups and evaluating findings from the programme and from other reliability studies and reviewing current practices adopted internationally.

Although it is generally realised that assessment results contain errors and substantial work has been carried out to study the reliability of assessments, there is considerable variability in how measurement uncertainty is reported in different parts of the world (Bradshaw and Wheater, 2009). While in the United States 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 about the reliability of results to the public (see, for example, AERA et al., 1999 – Standard 2.1; Newton, 2005a, b; Boyle et al., 2009; Phelps et al., 2010).

Newton (2005a, b) also discussed whether and how assessment organisations should communicate with the public about measurement inaccuracy, which includes unreliability (measurement error). Work carried out by Ofqual (2009), Ipsos MORI (2009) and Chamberlain (2010) under Strand three of the Ofqual Reliability Programme suggested mixed views from the assessment community and the public regarding dissemination of reliability information about assessment results. 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; Boyle et al., 2009; Ofqual, 2009).

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

Existing studies on public perceptions of reliability in England

As part of Strand three of the reliability programme, Ofqual commissioned two qualitative research projects from Ipsos MORI and the Assessment and Qualifications Alliance (AQA) to investigate public understanding of reliability and their opinions about the national examination system and measurement error in examination results. The research 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.

Ipsos MORI held two workshops in London and Birmingham in January 2009. Research participants were drawn from the following groups: teachers, students, parents, members of the general public, employers and examiners (Ipsos MORI, 2009). The sessions started with an analogy to an error occurring in medical treatment; this was used as a substantial input to help workshop participants to understand the concepts under discussion (see Ipsos MORI, 2009, for full discussion guides). 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.

The findings suggested 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' (Ipsos MORI, 2009). Sometimes participants appeared to be making a distinction between inherent and preventable error, but other times not. Some research participants stated that their attitude to error depending upon whether the error changed a student's grade or mark. For example, grade-related error was considered more consequential than mark-related error. Participants' views about error could also vary by group and by the perceived cause of the error. 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. There was considerable discussion on 'test-related error'. Although students and the general public were able to debate whether and how examinations can and should sample from curricula, they failed to realise that this would also introduce error in examination results.

Chamberlain (2010) from AQA conducted qualitative research to follow up Ipsos MORI's (2009) work. She collected data from 10 focus groups, with samples of: job-seekers, employees, employers, students taking Postgraduate Certificate in Education (PGCE) courses and teachers. 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" (Chamberlain, 2010).

Chamberlain's prior assumption that respondents would have limited awareness of reliability was confirmed by the data collected. All the participants, except secondary school teachers, lacked awareness of reliability concepts. The secondary teachers had more developed views, often based on experience of dealing with re-marks or appeals. The participants tended to be fairly trusting of the examinations process, trusting in the professionalism and training of subject experts. Once again, secondary school teachers' views differed from those of other groups; some respondents had acted as moderators in order to mediate the influence of external examiners. Participants felt it would be useful for reliability information to be communicated to the public in general terms, but were opposed to specific quantification of unreliability (e.g. via an indication of the amount of uncertainty associated with a grade) on a candidate's examination certificate.

Chamberlain (2010) also suggested the following hypotheses that could be addressed by items in a subsequent quantitative questionnaire survey:

- **Public perceptions of reliability**

  The public rarely think about the reliability of assessment outcomes.

  The public trusts that awarding organisations have systems in place to ensure that candidates receive the grades that they deserve.

  The public trusts that examiners are subject experts.

  The public is more concerned about outcomes that are poorer than expected, than about outcomes that are better than expected.

- **Acceptability of human error**

  The public believes that given the scale of the public examination process, some human error is inevitable.

  The public has a low level of tolerance for typographic errors and errors concerning the distribution of examination materials.

- **Acceptability of measurement inaccuracy**

  Measurement inaccuracy is caused by a collection of fortunate and unfortunate circumstances that apply to assessment as equally as to other aspects of life.

  The public does not perceive random error as 'true' error as it cannot be eliminated from the assessment process.

  The public understands that a different test on a different day may produce a different outcome.

The public believes that measurement inaccuracy impacts only on assessment scores.

The public believes that measurement inaccuracy has little impact on a candidate's grade.

The public does not perceive measurement inaccuracy as relevant to the reliability of assessment outcomes.

Giving the public examples of random error may encourage qualification users to view grades as 'approximations' rather than 'facts'.

- **Reporting reliability**

  The sectors of the population most likely to be interested in and engage with issues of reliability are those who are directly or indirectly involved in educational assessment (e.g. teachers, students, employers, universities).

  Assessment reliability is of limited relevance to the general public.

  Qualification users should be informed about how errors can occur in the assessment system.

  The public believes that employers should be aware that grades are 'approximations' rather than 'facts'.

- **How to report reliability**

  The public will not appreciate or engage with the publication of reliability statistics.

  The public believes that the publication of a reliability statistic alongside candidates' grades will confuse candidates.

  The public believes that the publication of a reliability statistic alongside candidates' grades will devalue candidates' achievements.

  Any efforts to improve understanding about reliability should start with teachers and students.

Ipsos MORI conducts a survey of perceptions of A level and GCSE each year for Ofqual that is now in its eighth wave (Ipsos MORI, 2010). The most recently reported wave of the survey was conducted shortly after the end of the 2009 examination session 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 (Ipsos MORI, 2010). The majorities of the teachers thought that

most students got the correct grade at GCSE. However, the general public were more sceptical, with more respondents believing that a larger proportion of candidates got the wrong grade. Respondents gave reasons that they perceived as being likely to cause candidates to get the wrong grade in examinations. They listed: students performing better or worse than expected in examinations or coursework. They also mentioned inaccurate marking and poorly designed examination papers.

Studies on trust

Both the Ipsos MORI and the AQA qualitative work (Ipsos MORI, 2009; Chamberlain, 2010) suggested that some participants had limited awareness of reliability concepts. It would be anticipated that some of the research participants might have different attitudes to unreliability in examination results if they had not gone through the workshop/focus group process. It would be expected that in the case when one does not have full knowledge about a situation, an attitude would be influenced by factors like personal experience and approaches to trust and others. A brief literature review on studies of trust was therefore also conducted to provide some insight into factors that affect trust, which would also apply to research into public trust in the reliability of examination results. There has been research undertaken to investigate trust in the government in general, government policies, government statistics, organisations, professionals and individuals (Putnam, 1995; Fukuyama, 1995; Hardin, 2002; O'Neill, 2002; Tschannen-Moran and Hoy, 2000; McLeod, 2002; Lycan, 1999; Coren et al., 1999; Reiss, 2000; Wilmot et al., 2005; Bradberry, 2007; Wilkinson and Pickett, 2009). Trust is at its most straightforward when it is clearly merited, for example, if government officials are competent and honest and perform their functions to the required standard.

A more complex situation is where a negative outcome is inevitable, for example, there will always be crime in modern societies, and there will always be old people who die when in hospital care. Trust in such contexts requires the truster to exercise trust even though he or she knows that a certain number of 'bad events' will happen. For trust to occur in this situation, it is necessary for the truster to perceive that sufficiently few negative events take place and that when they do take place, they cannot be attributed to incompetence, ill will, etc., on the part of the people or organisations being trusted. Such trust may be more difficult to achieve than straightforwardly merited trust for competently-performed successful actions.

The present study

The qualitative investigations of stakeholders' perspectives into reliability discussed previously had elements that sought to 'teach' participants about reliability – the Ipsos MORI (2009) research used a workshop format with a substantial initial input and the Chamberlain (2010) research used vignettes as part of a focus group approach. This might have helped the participants to understand the concept of reliability and the factors that could introduce uncertainty in examination scores and

to develop views on measurement error. The group discussions could also have influenced the opinions of the participants about error in examination results. Furthermore, the small sample size of these studies makes it inappropriate to make any generalisation of the findings. The Ipsos MORI (2010) survey only addressed some narrow aspects of reliability of examination results. The present study seeks to contribute further to a developing understanding of attitudes to reliability and unreliability using an objective questionnaire survey and explores the public's awareness of and opinions about reliability in the following areas:

- 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.

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.

Methodology

Instrument development

The questionnaire was structured into five distinctive topics:

- Topic A: Experiences of, and knowledge and beliefs about the examination system
- Topic B: Awareness of unreliability
- Topic C: Attitudes to unreliability
- Topic D: Views on approaches for improving reliability
- Topic E: Approaches to trust.

This structure of the questionnaire allowed the balance in items between the topics to be controlled and for relationships between topics to be investigated (see Appendix A for a full list of questions in the questionnaire).

The questionnaire had 23 questions, many of which were multi-part, making a total of 80 individual sub-items in total. All sub-items are multiple choice questions with varying number of response options. The majority of the sub-items required respondents to endorse their views on a statement with the degree varying from 'Strongly agree' to 'Strongly disagree'. A few questions also used 'Don't know' as a response category. The statements were also varied so as to contain both positive and negative statements (Pearson and Raeke, 2000). There were no constructed response items in the questionnaire; this was because the Ipsos MORI (2009) and Chamberlain (2010) research exercises had provided substantial qualitative evidence and the current exercise was aiming to derive a quantitative measure. A few questions in the questionnaire had slightly different versions for different respondent groups (for instance, those asking about respondents' personal experience of examinations; see next section on respondents sampling). However, the vast majority of the questions and sub-items were common for all respondents groups, to facilitate comparison between groups.

In addition, many of the questionnaire questions had sources from the research literature. This involved questions which were the replication of questions from other research instruments or the editing of well-known quotes or assertions from the literature for respondents to express their level of agreement to. This grounding of the items in the literature had two benefits: for those items that were based on earlier items, it permitted comparison of the responses to this questionnaire with those to earlier instruments. Where statements from the literature were turned into questionnaire items, it permitted a 'validation' of those statements with stakeholder

groups or a 'reality check'. The origins of the various items in the literature are shown in Table 1.

Table 1: Question topics and basis of questions in literature where appropriate.

Topic	Question number	Question topic	Source of item
A	1	Personal experiences of exam system: I was happy with my exams, etc.	
A	2	View of the national exam system – is it doing a good job?	
A	3	View of the national exam system – I have confidence in the national examinations system.	Adapted from Ipsos MORI (2009)
A	4	Which is more important: comparability between boards, standards over time, being fair to all groups, measurement error?	Newton (2005a) Chamberlain's suggested hypothesis no. 6
A	5	Causes of trust in exam system – expert judgement, statistical procedures, government backing, reputation of exam boards	See the background section of this paper; Chamberlain's suggested hypotheses no. 2 and 3
B	6	Factors that affect exam scores	Chamberlain's suggested hypothesis no. 9
B	7	Sources of inconsistency: what affects scores – differences between markers; luck of the draw in getting the right questions; level setting panel	Newton (2005a)
B	8	The amount of misclassification in different GCSE subjects	

Topic	Question number	Question topic	Source of item
C	9	Attitudes to accuracy in the exams system	Chamberlain's suggested hypothesis no. 7
C	10	Distinction between inherent error versus avoidable mistakes – acceptability of both	Newton (2005a)
C	11	Agent of the inconsistency (board's fault; student not paying attention/revising the whole syllabus; government changing exams system)	Ipsos MORI (2009)
C	12	Consequence of error	Ipsos MORI (2009)
C	13	Acceptability of error in exam scores for different groups	
C	14	The appeals system	Chamberlain (2010)
D	15	Whether informing people about 'bad news' in exams increases or decreases trust	Newton (2005a)
D	16	Indication of amount of uncertainty on an individual's certificate	Chamberlain's suggested hypothesis no. 19
D	17	Options for increasing reliability: extra markers, multiple-choice questions, longer tests, teacher assessment	Ipsos MORI (2009)
E	18	Trust in general	World Values Survey (WVS), 2005–6 Wave, Root Version questionnaire item V23 (WVS, undated)
E	19	Trust in different groups of people	Adapted from WVS questionnaire items V4–9

Topic	Question number	Question topic	Source of item
E	20	Trust in organisations	
E	21	Trust in professions	See background section of this paper
E	22	Trust when you know that some things inevitably go wrong	See background section of this paper
E	23	Attitudes to government statistics	See background section of this paper; Wilmot et al. (2005)

Respondents sampling

The following three groups of stakeholders were chosen to provide samples of respondents to the questionnaire:

- sixth form students studying on A level courses in schools or colleges in England
- school teachers who teach on A level courses in schools or colleges in England
- employers (especially members of staff with responsibility for recruitment).

This was a choice of respondents which placed limits on the research findings; for example, Chamberlain's work (2010) suggested that primary school teachers had a different view to those teaching on A level courses. However, the choice to select these groups was made on the grounds that they would be groups for whom examinations and their reliability would be likely to be at the forefront of their mind. Some other groups could have been selected (say, primary school teachers, parents or university selectors) but their views could be sought – via survey or other methods – at a later date.

The National Foundation for Educational Research (NFER) was contracted to sample respondents and administered the questionnaire to collect data. To obtain responses from teachers and students, samples of institutions were drawn from the NFER's Register of Schools and Colleges. The database contains comprehensive and up-to-date information about each school in England, including basic school information such as size, governance and location, as well as information about

schools' overall levels of attainment. College data is also held on the system along with information about college type and geographical location. The Register of Schools and Colleges is updated regularly using extracts from Edubase and other national data sets, along with live amendments triggered by our ongoing work with schools and colleges. A random sample of 800 institutions was drawn from maintained and independent schools in England with a representative number of sixth form centres and further education (FE) colleges. The sample was stratified by school type and government office region. Based on NFER's experience of inviting institutions to participate in questionnaire surveys, it was anticipated that around 35 percent would agree to participate in the research. The 800 institutions in the sample included only those where they have both year 12 and year 13 students. Two or three A level teachers and five or six A level students were invited to complete the online questionnaire.

The sample of employers was obtained from Experian. It was anticipated that it could be more difficult to engage employers in research and for that reason it was decide to invite representatives from 3,000 companies. It could also be hard to define a representative group and for that reason employers were sought from the full range of those available in the following categories:

- standard industrial classification (SIC) 2007 categories
- business size by turnover
- location by head office (from English counties)
- age of business.

Table 2 shows the achieved sample sizes for the three groups and respondents' demographic information. The error associated with the population estimate of the percentage response to a sub-item based on the achieved sample sizes was estimated to be within ±5.5 per cent, ±5.2 per cent and ±6.8 per cent for teachers, students and employers respectively at a 95 per cent confidence interval.

Table 2: Respondents' basic demographic information.

Respondent group	Total	Male (per cent)	Female (per cent)	Subjects (teaching, studying or studied) (per cent)			
				S1*	S2**	S3***	S4****
Teachers	314	45	55	53	13	12	22
Students	358	45	55	33	12	14	40
Employers	210	62	38	8	13	7	22

*S1: Including English, languages, art, drama, music, history and religion.

**S2: Including technologies, geography and ICT

***S3: Mathematics.

****S4: Sciences.

Data collection and analysis

The questionnaire was administered through the internet. This online delivery method was chosen because it was believed that it would return a robust set of data quickly without the need for manual data entry. It is acknowledged that some methodological sources (Groves et al., 2004) have ascertained bias in sample of respondents to online questionnaires, for instance an over-representation of young people or an under-representation of ethnic minority or working class people. This potential weakness in the research methodology was countered in two ways: firstly, the sampled groups (sixth form students, their teachers and employer-recruiters) would be unlikely to be members of groups that would be averse to completing an online questionnaire. Secondly, the data collection agency provided a detailed description of the achieved sample, allowing any biases to be explicitly evaluated. All data was collected anonymously so that no organisations and individuals could be identified in subsequent analysis.

To facilitate statistical analysis, the response categories in a sub-item were transformed into numerical values, varying from 0 for the weakest category to the number of options minus 1 in the sub-item for the strongest category. For negatively asked question, the transformation was reversed. The coded data were analysed for reliability for each topic for each of the respondent groups, in addition to analysis for some basic descriptive statistics at both sub-item level and topic level. Correlation analysis between the topics for each group was also conducted to investigate how attitudes to assessment error correlate with other attributes of the respondents based on information obtained from individual topics.

Results and discussion

Instrument internal consistency reliability

The questionnaire used in the current study attempts to explore respondents' approaches to trust, knowledge and understanding of the assessment process and factors affecting students' performances on examinations and factors introducing errors in examination scores, as well as attitudes to unreliability in examination results, and is therefore a multi-dimensional instrument. The internal consistency reliabilities of scores represented by Cronbach's alpha for the individual topics for each group were estimated to vary from 0.25 for Topic D (views on approaches for improving reliability) for teachers to 0.85 for Topic E (general approaches to trust) for employers (see Table 3). This indicated that there was greater variation in choosing the response categories of sub-items in Topic D for teachers and employers. Topic D was also the shortest topic, containing only nine sub-items. Except for Topic D for teachers and employers, Table 3 suggests that the topic response data had reasonably adequate internal reliability.

Table 3: Values of Cronbach's alpha for individual topics in the questionnaire.

Group	Cronbach's alpha				
	Topic A	Topic B	Topic C	Topic D	Topic E
Teachers	0.68	0.62	0.62	0.25	0.80
Students	0.64	0.69	0.72	0.58	0.83
Employers	0.81	0.78	0.73	0.58	0.85

Experience of and confidence in the national examinations system

Questions in the first topic of the questionnaire were about respondents' personal experience, knowledge and opinions about the national examinations system. There were 23 sub-items in this topic. Responses to the sub-items from the respondents are discussed below.

Figure 1 illustrates the percentages of respondents who strongly agreed or agreed to a series of statements about their experiences of or beliefs about the examinations system. About 54 per cent of students thought 'Doing exams is unreasonably stressful', while less that 18 per cent of them regarded their experience of the

examinations system acceptable. Girls were more positive than boys. Over 81 per cent of teachers thought most examinations were fair to most students, and over 93 per cent of them thought their students were well-prepared for examinations. About half of the employers thought that students' grades generally reflected their ability. About 14 per cent of the employers suggested that they selected candidates for interview based on their examination results, while about 65 per cent indicated that they sometimes use their own tests to assess candidates' skills.

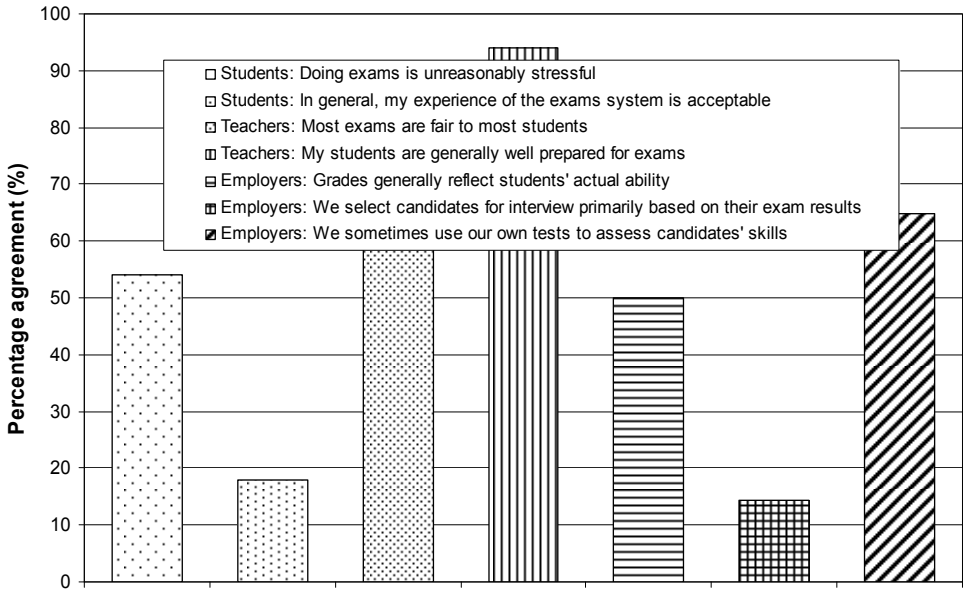

Figure 1: Experiences of and beliefs about the examinations system.

Figure 2 shows the percentages of respondents from each of the three groups who selected 'Strongly agree' or 'Agree' to the statements 'In general, students/I get the grades they/I deserve in exams' and 'To what extent do you agree or disagree with the statement ''I have confidence in the national examination system."' About 89 per cent of the teachers felt that their students got the grades they deserved, and about 66 per cent of the employers thought that students got the grades they deserved. These findings are broadly in line with the findings from the 2010 Ipsos MORI survey and the qualitative studies by Ipsos MORI (2009) and Chamberlain (2010). In contrast, only about 17 per cent of the students felt that they got the grades they deserved, which is substantially lower than the percentages for the teachers and the employers. It is also noticed that this figure for students is substantially lower than that for a similar question asking about whether students taking A levels get the grades their performance deserves by Ipsos MORI (2010). However, the two

questions were phrased differently in the two studies: in the Ipsos MORI survey the question for students was 'Most students taking A levels get the grades their performance deserves', while the question for students used in the present investigation was 'In general, I get the grades I deserve from my exams.' While the Ipsos MORI survey was asking about students in general, the present study was asking about the student himself/herself, and this could have resulted in the difference in opinions between the two samples. The samples used in the two studies were also different in terms of size and composition. The sample size of students for the present study was substantially larger than that for the Ipsos MORI study. The methods used in the two studies were also different. While the Ipsos MORI survey used telephone interviews in its investigation, the present study used an online objective questionnaire survey.

About 62 per cent of the teachers showed confidence in the examinations system, which is lower than the percentages from the 2010 Ipsos MORI survey about views on the accuracy of GCSE grades. The percentages of students and employers who had confidence in the system were substantially lower than that for teachers, with only 42 per cent and 39 per cent respectively. It would be expected that teachers would be more confident in the examination system than students and employers as they use the system more than students and employers and are more familiar with the system.

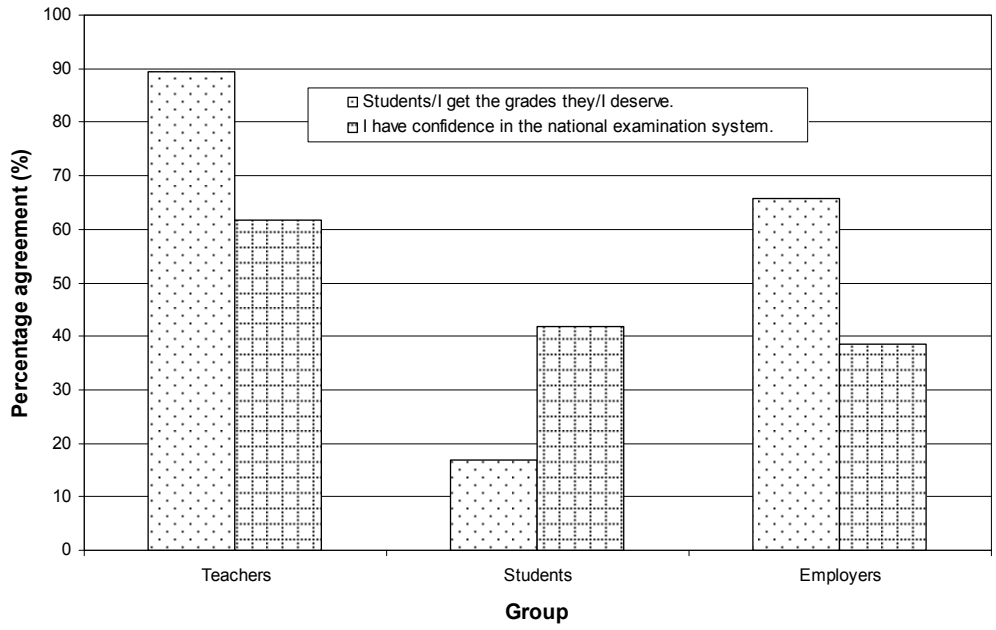

Figure 2: Confidence in the national examinations system.

When asked about their opinions about the performance of the national examination system, 26 per cent of teachers, 25 per cent of students and 18 per cent of employers felt that the system was doing either a very good job or a good job (see Figure 3). However, about 61 per cent of teachers, 57 per cent of students and 48 per cent of employers thought that the system was doing a good job but needed improving. Again teachers trusted the system more than students and employers. About 12 per cent of the teachers, 14 per cent of the students and 23 per cent of the employers thought that the systems was not doing a good job and should be reformed.

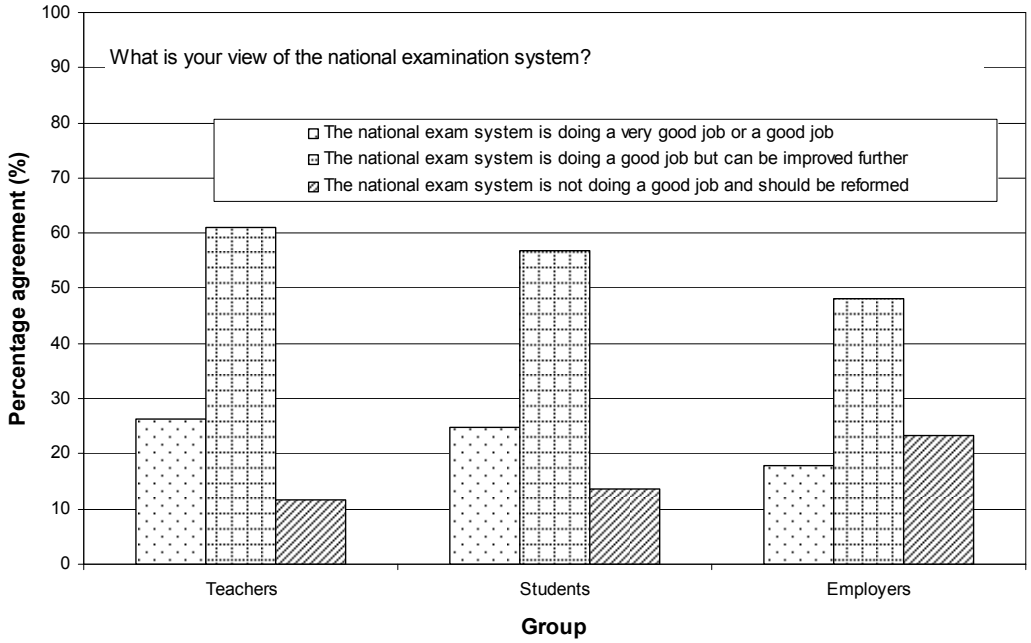

Figure 3: Views of the national examination system.

Opinions were sought of the respondents on a range of features that the examinations system must possess, including demands of question papers, question papers free from typographic errors and mistakes, comparability of results from different examination boards, consistency of standards over time, consistency of results from the same examinations, fairness for all people in society, and people getting the results they deserve. Over 95 per cent of teachers regarded all those features important for the examination system, which were generally slightly higher than the percentages of students and employers (see Figure 4). About 87 per cent of students thought it important for examination papers to be demanding, and 82 per cent of them thought consistency of results from the same examinations important.

About 81 per cent of employers thought fairness of examinations for all people important.

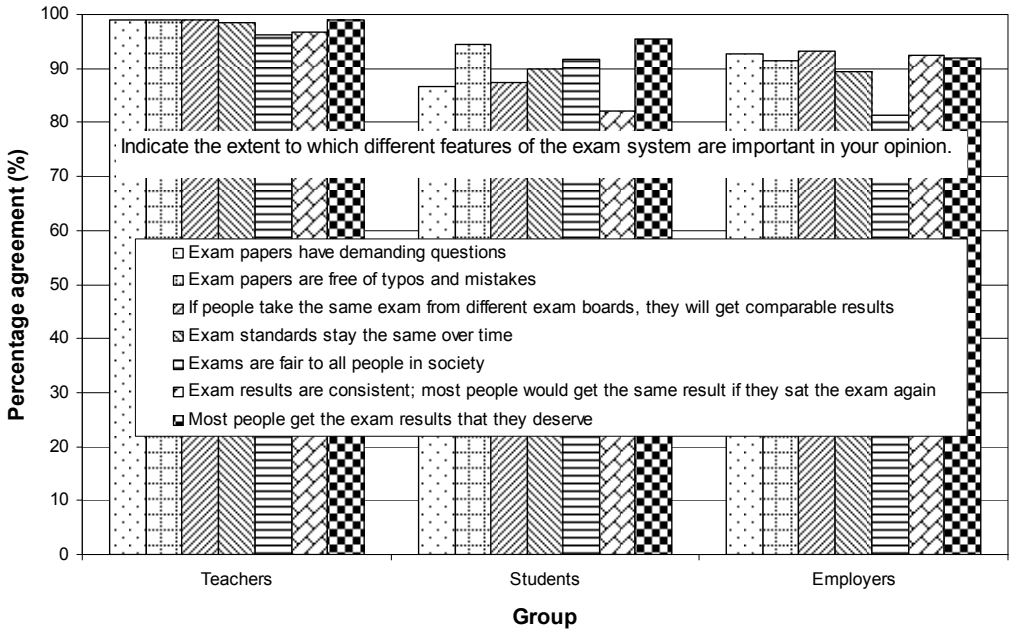

Figure 4: Views on the importance of features of examination system.

Figure 5 depicts percentages of respondents from the three groups who regarded the various factors listed in the questionnaire as important for creating trust in the examination system. The vast majority of the respondents thought all the four factors shown in Figure 5 important in creating trust. The endorsement rate for 'Subject experts making sure that exams measure the right things and they are at the right level' and 'Exam boards have the necessary expertise and experience' was over 91 per cent. In terms of use of statistical procedures in awarding, the endorsement rate was over 80 per cent for teachers and students and about 74 per cent for employers. These findings were generally consistent with the findings from Chamberlain's work, which suggested that the participants involved in the focus group discussions trusted the examination system because they believed that awarding organisations had systems in place to ensure that candidates received the grades they deserved (see previous discussions and Chamberlain, 2010).

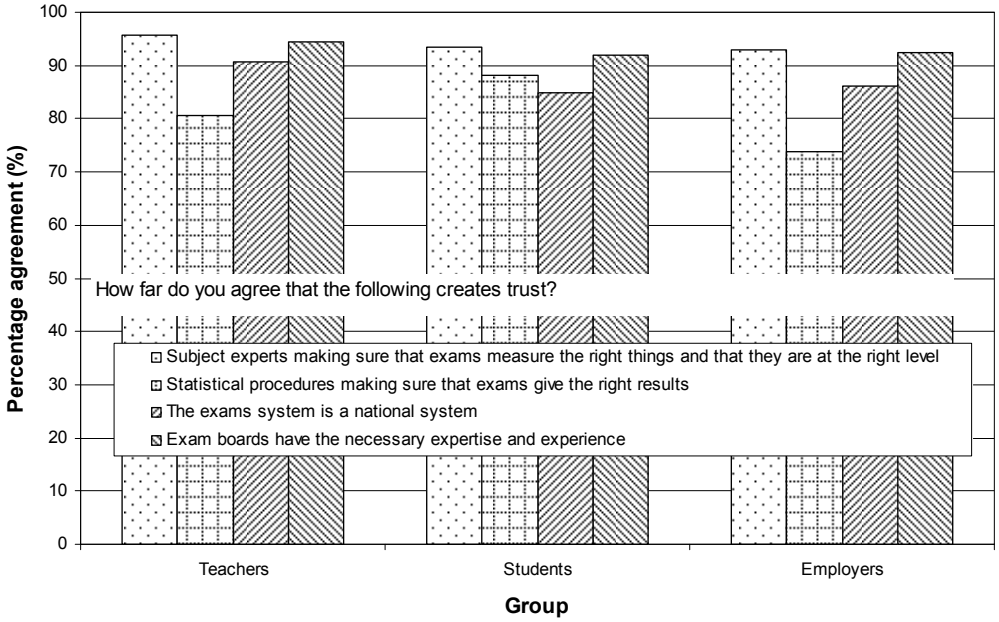

Figure 5: Views on factors that create trust in the examination system.

Understanding of factors affecting performances on examinations and factors causing unreliability in examination results

Questions in the second topic of the questionnaire were about respondents' awareness of reliability issues, including questions about understanding of factors that can affect students' performances in examinations and factors that can introduce errors in examination scores. There were a total of 12 sub-items in this topic.

The studies conducted by Ipsos MORI (2009) and Chamberlain (2010) indicated that the research participants generally understood the many factors that could affect the performances of students on examinations. This was further confirmed by findings from the present study. Figure 6 shows the percentages of respondents from the three groups who selected either 'Strongly agree' or 'Agree' for the five statements about factors that could influence a student's score on an examination. All groups showed a similar pattern in the level of endorsement for the statements. In general, all the five factors listed in the questionnaire were regarded as important in influencing students' performances on examinations. Of these factors, knowledge about the subject and preparedness of the student were regarded as most important by all respondents (with endorsement rates over 91 per cent). 'How well the student feels on the day' and 'Who marks the question paper' were regarded as less

important than knowledge about the subject and preparedness, with endorsement rate varying from 57 per cent for employers to 66 per cent for teachers and students.

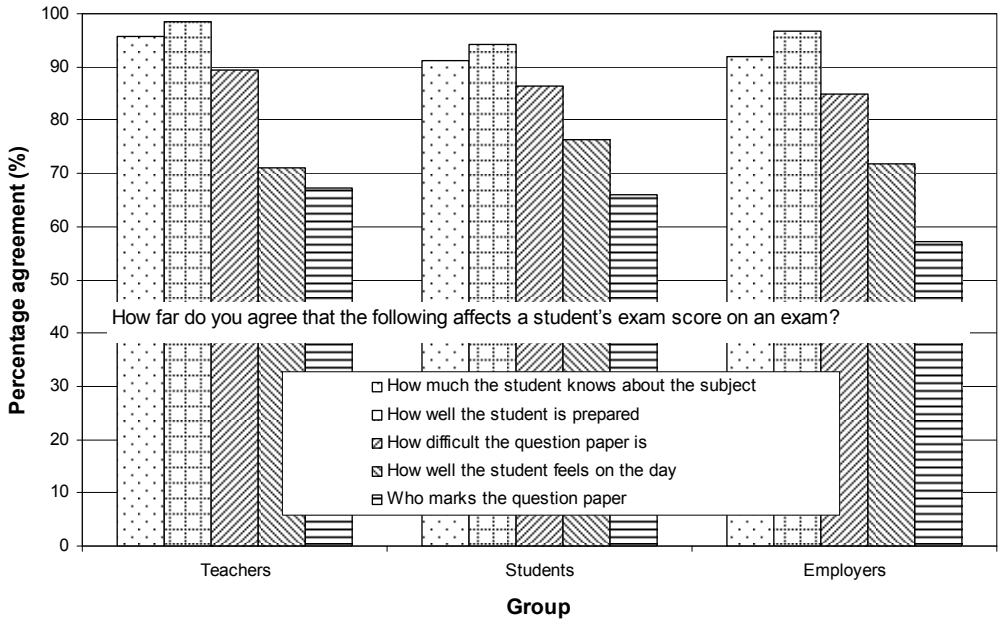

Figure 6: Understanding of factors that affect students' scores on an examination.

The study by Ipsos MORI (2009) suggested that students were aware that some inconsistency between human markers was inherent in subjects like English. However, there were also statements that such inherent error should be minimised or even eliminated. Although the workshops were guided, some participants were still not quite clear about the factors that could introduce inconsistency in examination scores if the examination procedure was repeated (see previous discussions). Results from the present investigation seemed to indicate that the majority of the respondents to some degree understood the main sources of error in examination scores. Over 58 per cent of the respondents selected either 'Strongly agree' or 'Agree' for the statements about factors that could cause inconsistency in examination results if the examination procedure was repeated (see Figure 7). About 75 per cent of the teachers and 85 per cent of the students thought 'Test questions (e.g. if a different test had been set, the student might not have been disadvantaged by the wording of an essay question)' an important error-contributing factor. Over 71 per cent of both teachers and students felt that 'Marking inconsistency (e.g. if a different marker had been assigned, the student might have achieved a different result)' could introduce error in examination results.

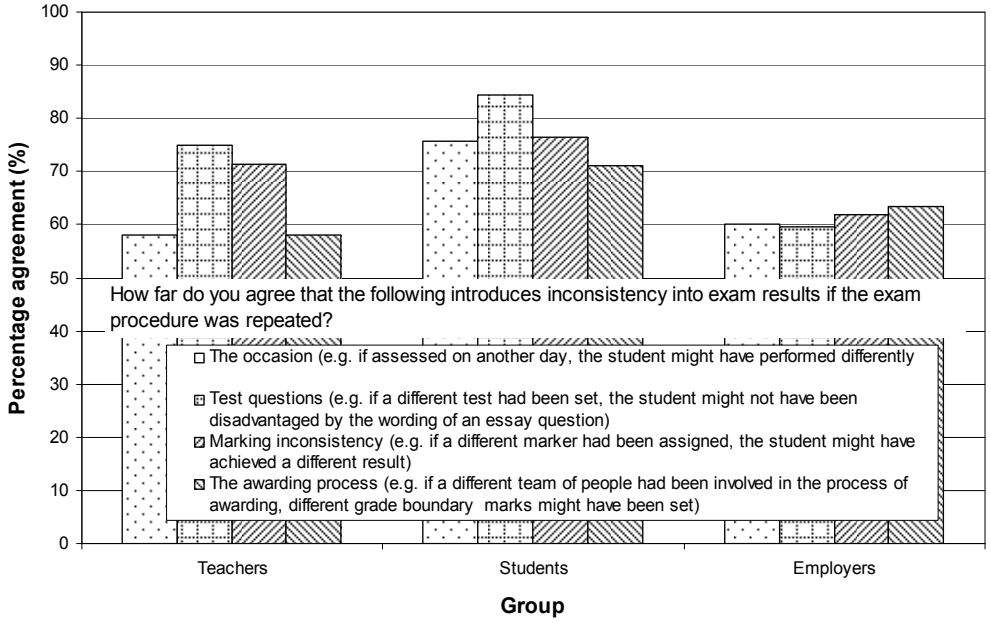

Figure 7: Understanding of factors that can introduce uncertainty in examination results.

Table 4 shows the percentage of respondents selecting different grade misclassification categories for different GCSE subjects for the question 'In your view, what is the proportion of students who have a GCSE grade that does not reflect his/her actual ability?' Mathematics and science were generally regarded as having a smaller amount of grade misclassification than English for all three groups. Over 50 per cent of teachers and students perceived grade misclassification in GCSE mathematics was less than 30 per cent. It is noticed that about a third of teachers and employers select the 'Don't know' option.

Table 4: Percentages of respondents who thought a proportion of students had a GCSE grade that did not reflect his/her ability.

Percentage misclassification	GCSE mathematics (per cent)			GCSE science (per cent)			GCSE English (per cent)		
	T*	S**	E***	T	S	E	T	S	E
<10	24	25	15	23	17	15	16	15	12
10–20	19	19	18	15	21	16	23	18	15
21–30	12	13	13	15	16	14	16	17	12
31–50	7	13	13	8	15	11	7	18	13
>50	3	14	12	5	13	11	9	16	22
Don't know	35	16	29	35	18	34	31	17	27

* T – Teachers; ** S – Students; ***: E – Employers

Attitudes towards unreliability and assessment error

The concept of reliability is relatively abstract and difficult to comprehend, and the term standard error of measurement is frequently used to interpret reliability measures. Newton (2005a) provides some alternative definitions of the different types of error in assessment results and discusses the different sources that cause those errors. He refers to 'measurement inaccuracy' as 'the variety of ways in which any set of assessment results will always depart from the mythical ideal of perfect accuracy for all students'. 'Measurement inaccuracy' is a broadly conceived notion and includes reliability, validity and comparability deficits. It can be contrasted with 'human error', which includes 'head-slappingly obvious mistakes'. Human errors are, for practical purposes, inevitable in large-scale testing programmes, but they are not inherent as a matter of principle – in contrast to measurement inaccuracy. Newton (2005b) uses the overarching term 'assessment error' to include both 'measurement inaccuracy' and 'human error'. Here the part of measurement inaccuracy associated with unreliability is referred to as the random measurement error. Newton's work (2005a, b) presents several challenges to research investigating attitudes to unreliability in examination scores. Firstly, it suggests that it would be worth investigating whether respondents' attitudes to inherent measurement inaccuracy (or measurement error) differed from attitudes to human error. Further, the relative breadth of Newton's concept of 'measurement inaccuracy', as compared to the

current research's focus on reliability and unreliability can be noted. It may be worth comparing respondents' attitudes to unreliability with their attitudes to other sources of inaccuracy, such as invalidity, lack of comparability and so on. The third topic in the questionnaire attempted to address these issues.

Questions in Topic 3 asked about respondents' attitudes towards unreliability in examination results, including their tolerance for human mistakes and inevitable measurement uncertainties. There were 16 sub-items in this topic.

Figures 8 and 9 illustrate percentages of respondents who were intolerant of error in examination grades and inaccuracy in the assessment system and those who recognised and accepted the difference between inevitable inherent variability in examination results and avoidable human mistakes in the examination system from the three groups. A contrast between Figure 8 and Figure 9 is noticed. While about 63 per cent of teachers and students selected 'Any level of error has to be unacceptable – even just one candidate getting the wrong grade is entirely unacceptable' on one hand, over 50 per cent of them also selected 'There's a difference between an avoidable mistake – like a typo on a paper – and something inevitable like inconsistency between two markers', suggesting tolerance for error. This inconsistency may reflect the weak relationship between knowledge about reliability and attitudes to unreliability and is consistent with findings from the Ipsos MORI research (see previous discussions). Employers were more intolerant of error than teachers and students.

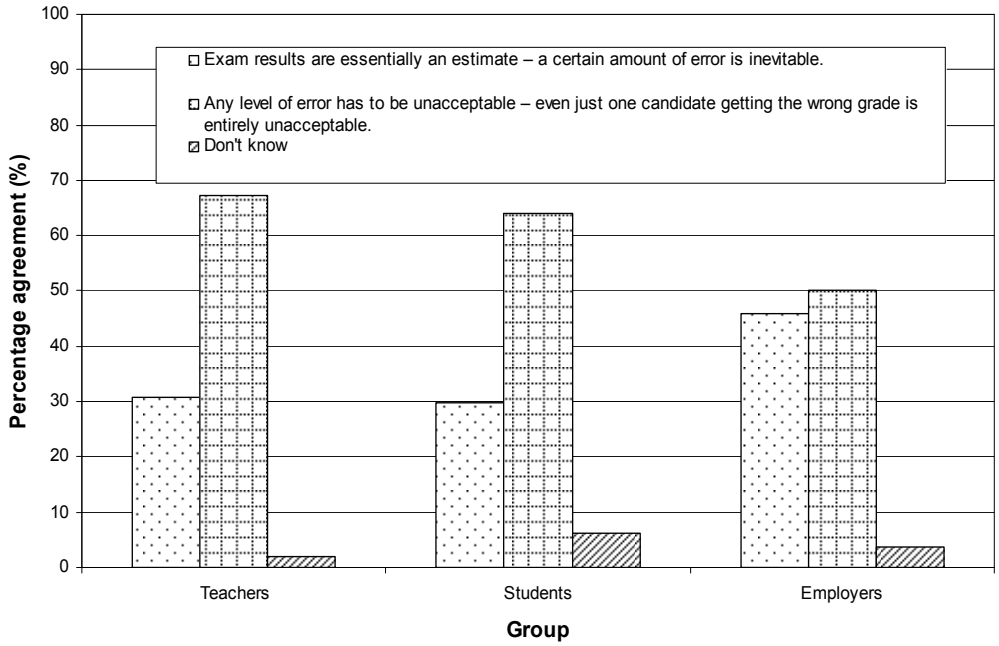

Figure 8: Attitudes to inaccuracy in examination results.

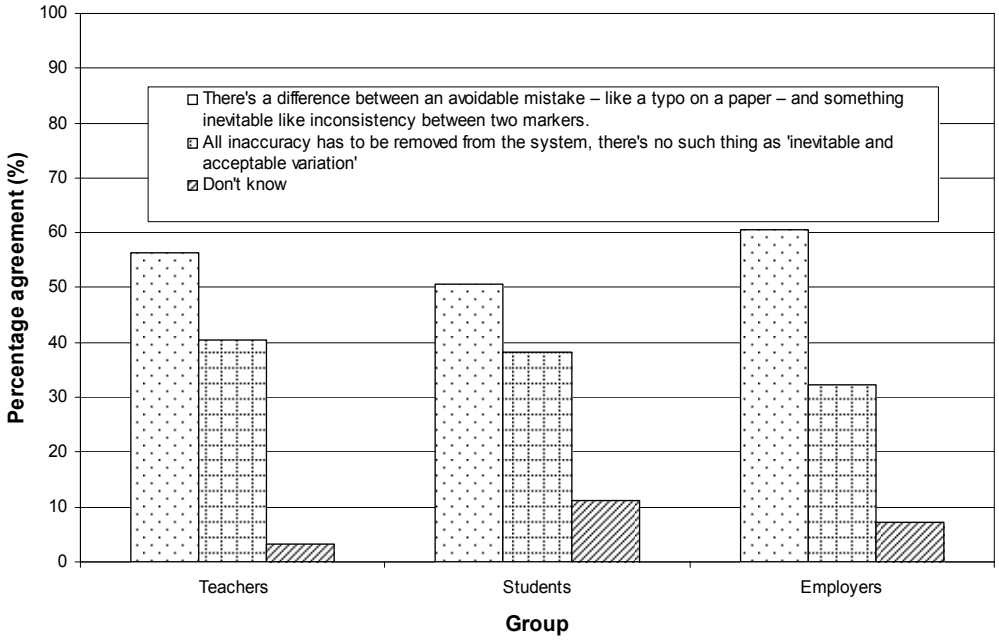

Figure 9: Attitudes to inherent error and unavoidable mistakes.

Ipsos MORI's research (2009) indicated that some participants stated that their attitudes to error depended on whether the error changed a student's grade or mark. They considered grade-related error to be more consequential than mark-related. These findings are supported by findings from the present study (see Figure 10). Less than 49 per cent of the respondents from all the three groups agreed that 'Error in the mark a student receives which does not affect a grade overall is not a cause for concern', while over 90 per cent agreed that 'Error which results in a student receiving a different grade to the one they deserve is serious.' Over 86 per cent of the respondents from the different groups felt that 'Error that changes a grade C to a grade D in a GCSE exam is particularly important.'

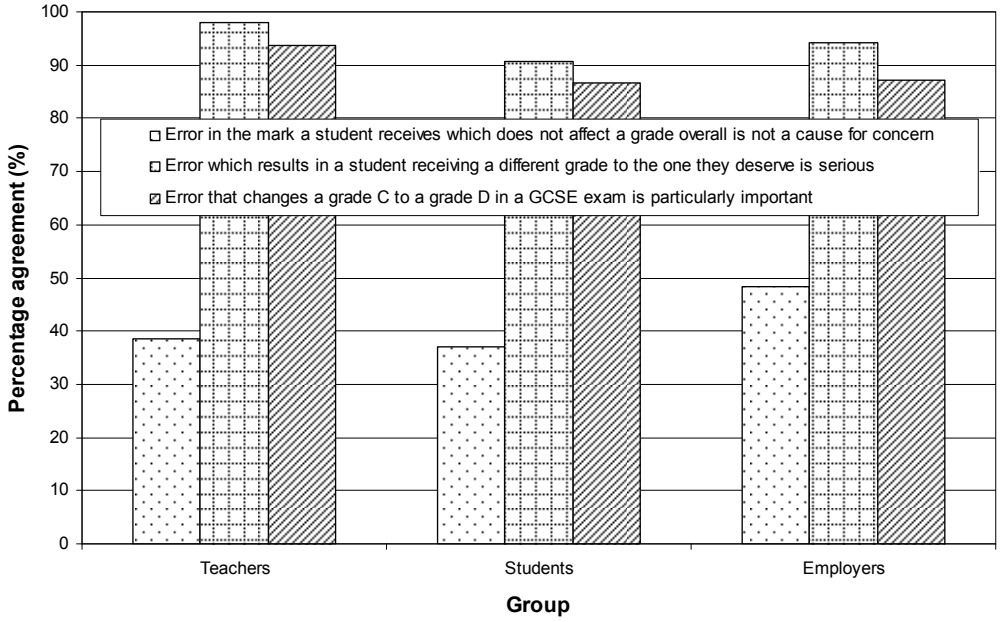

Figure 10: Influence of the effect of error on examination results on attitudes.

Figure 11 shows the percentages of respondents from the three groups who selected either 'Strongly agree' or 'Agree' for the statements 'The performance on the day of an exam can be affected by feeling stressed or unwell, but it is ''just one of those things'', 'Students need to be held accountable for how they perform on the day of the exam' and 'Exam boards should do everything they can to minimise inconsistency from their processes.' Over 76 per cent of the teachers and employers endorsed all the three statements. Unsurprisingly, only 53 per cent of the students agreed that 'Students need to be held accountable for how they perform on the day of the exam', while 78 per cent of the teachers and 77 per cent of the employers endorsed the statement. Over 94 per cent of all the respondents agreed that 'Exam boards should do everything they can to minimise inconsistency from their processes.' These findings again are generally consistent with the findings from the qualitative studies by Ipsos MORI (2009) and Chamberlain (2010).

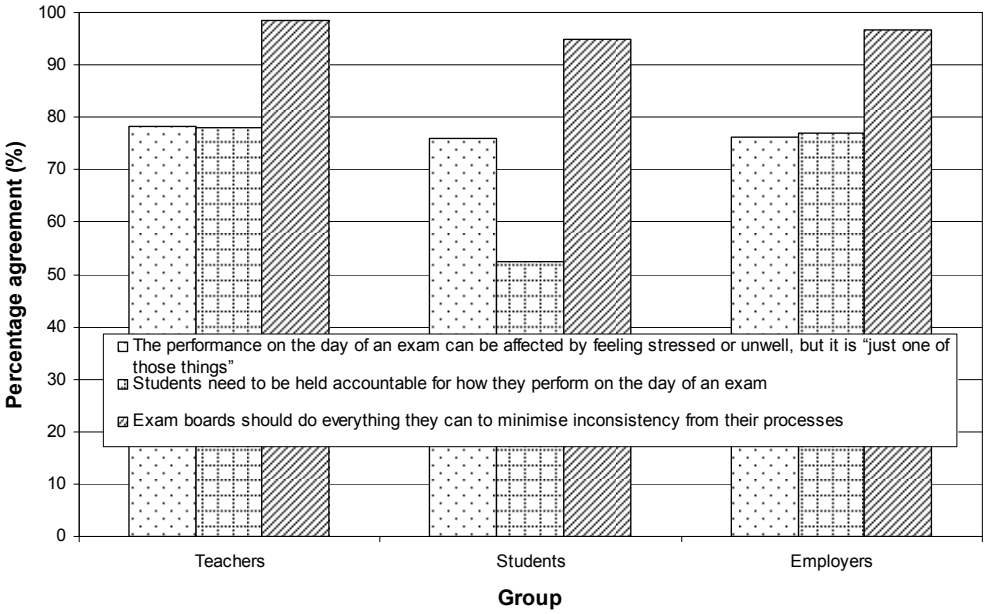

Figure 11: Responsibilities for error in examination results.

Influence of whom error in examination scores may affect on the views of assessment error of the respondents was also investigated. Figure 12 depicts the percentages of respondents who selected 'Completely acceptable' or 'Somewhat acceptable' to the question 'Please indicate how acceptable error in exam scores would be if it affected the following people or groups of people.' It is clear that the person the error affected had only very limited influence on the views of the respondents. Only less than 7 per cent of teachers indicated that their degree of acceptance of error in examination scores would be influenced by whom it affected. The endorsement rates for students varied from 13 per cent to 17 per cent, which were generally higher than those for teachers and employers.

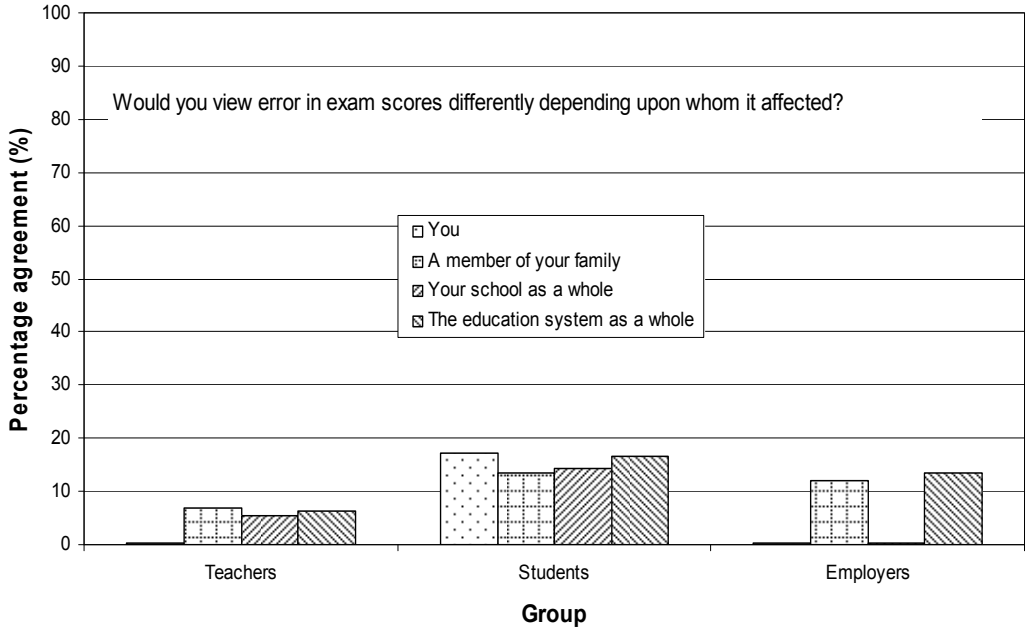

Figure 12: Influence of whom error in examination scores affects on level of acceptance of assessment error.

Figure 13 depicts views of the respondents on the appeals process of the examination system. Nearly 50 per cent of teachers, 55 per cent of students and 35 per cent of employers thought that the appeals process helped the examination system to produce more accurate results. Similar numbers of teachers and students regarded the appealing process as fair. Interestingly, about 30 per cent of teachers thought that 'Teachers sometimes appeal results inappropriately.' Employers were generally slightly less positive about the appeals process than teachers and students.

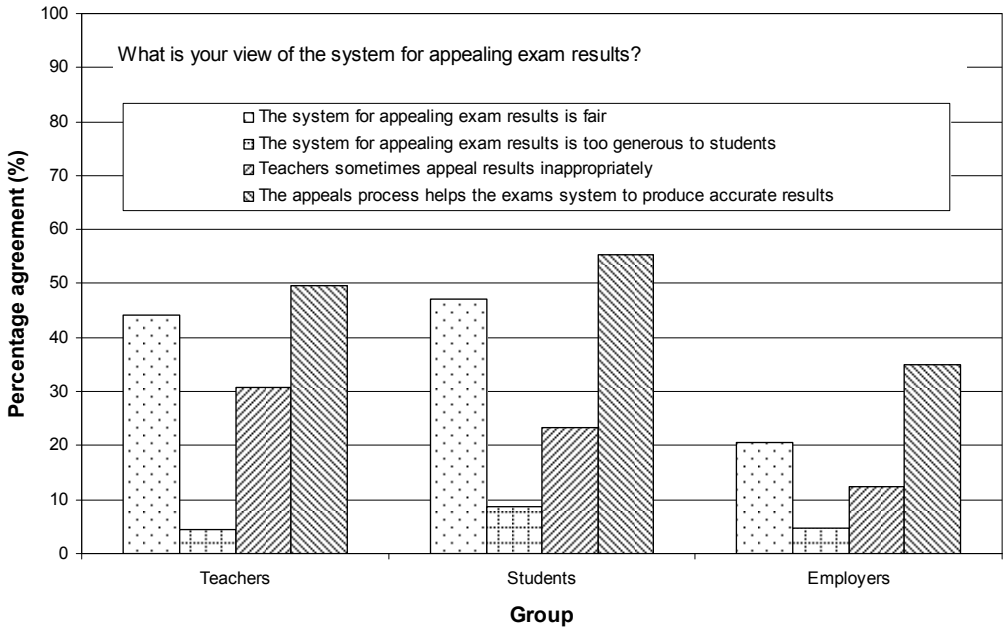

Figure 13: Views on the examination system for appealing examination results.

Approaches for improving reliability

Questions in Topic 4 asked about respondents' opinions about approaches that can be adopted to improve the reliability of examination results. There were nine sub-items in this topic.

Figure 14 shows the percentages of respondents from the three groups who strongly agreed or agreed to three statements about making reliability information available more widely by the government and assessment providers and the consequences. Over 82 per cent of all respondents agreed that 'Government and exam boards need to be open about how much uncertainty there is in exam results.' About 68 per cent of teachers, 73 per cent of students and 53 per cent of employers agreed that 'Publishing information about uncertainty in exam results might reduce public confidence.' Only less than 30 per cent of teachers and over 37 per cent of students and employers thought that 'Exam boards should not be embarrassed by a certain amount of inaccuracy in scores – it is inevitable', suggesting that substantial proportions of the respondents were intolerable of inaccuracies in examination results.

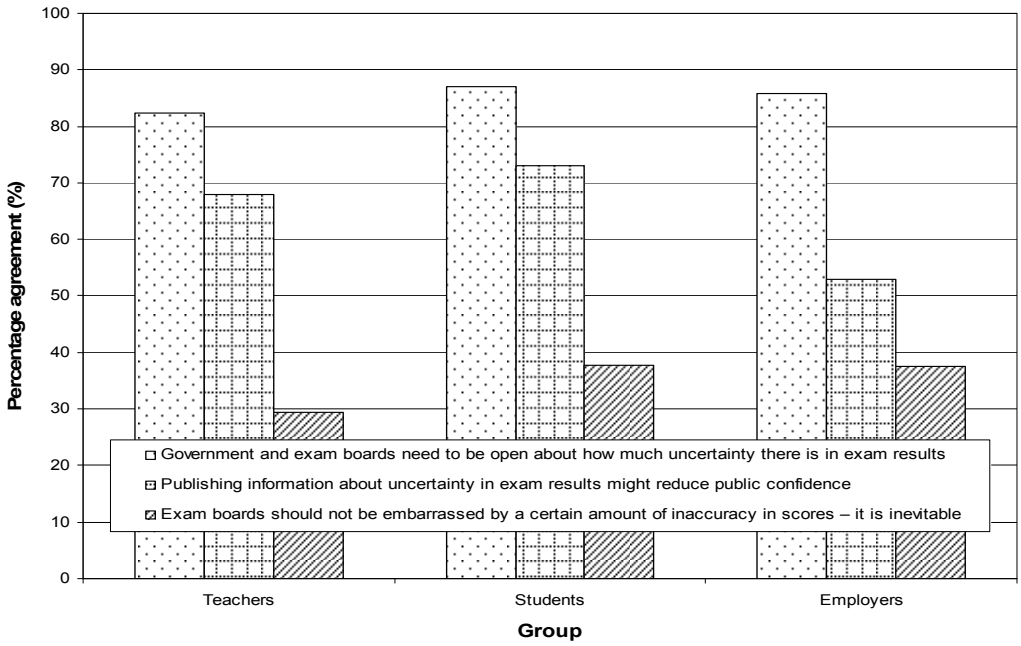

Figure 14: Views on openness of the government and assessment providers about uncertainty in examination results.

Figure 15 shows the level of endorsement for the various approaches that could be adopted to improve reliability by the respondents. Over 92 per cent of the respondents agreed that 'Improve training for markers' was important. Also over 79 per cent of all respondents agreed that 'Have two markers for essays' was important. Interestingly, about 33 per cent of the teachers felt it necessary to 'Use more teacher assessment for awarding qualifications' to improve reliability, while this is over 54 per cent for employers and over 64 per cent for students. This could suggest that teachers did not have a great confidence in teacher assessments, while students and employers did. Only 18 per cent of the teachers and 29 per cent of the employers agreed to 'Use more multiple-choice questions' to improve reliability, while the endorsement rate for students is about 50 per cent. About 22 per cent of the teachers, 17 per cent of the students and 22 per cent of the employers agreed to 'Have longer tests' to improve reliability.

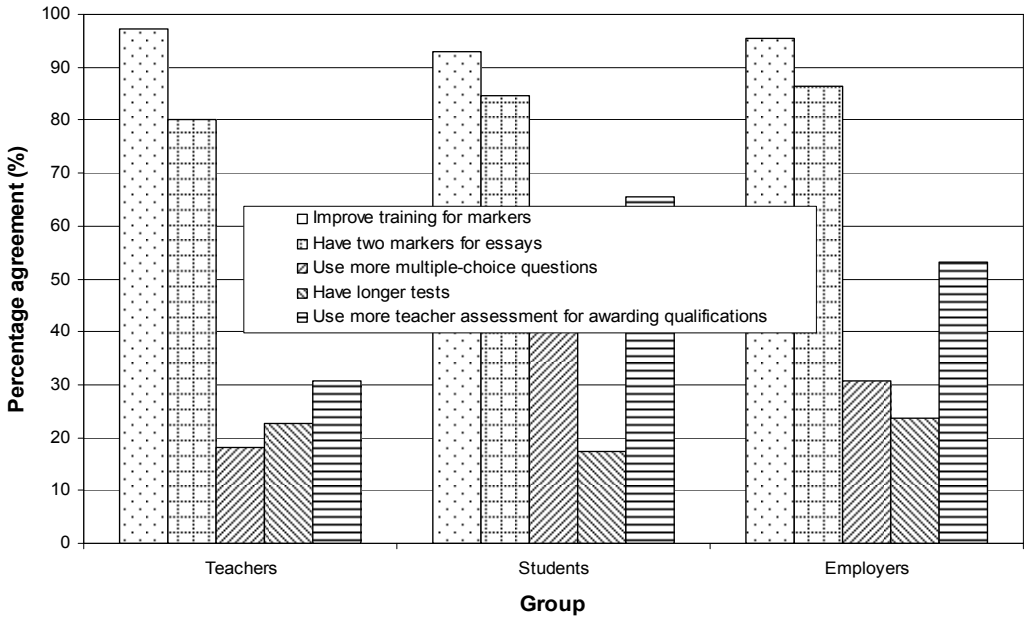

Figure 15: Views on approaches for improving assessment reliability.

Findings from Chamberlain's work (Chamberlain, 2010, also see Chamberlain's hypotheses discussed previously) indicated that the majority of the participants at the focus groups were not in favour of reporting reliability statistics and believed that doing so would undermine candidates' achievements and create uncertainty, although some participants suggested that the public should be informed of the existence of error in examination results. Further, secondary school teachers felt that teachers and students needed to be better informed. These findings were supported by results from the present investigation. To explore this further, there was also one question in Topic 4 asking whether uncertainties associated with examination grades should be indicated on students' certificate; Figure 16 shows the percentages of respondents from the three groups who selected each of the options 'Yes', 'No' and 'Don't know' for the question 'Do you think the degree of uncertainty associated with an exam grade should be indicated on a student's certificate?'. About 67 per cent of teachers, 33 per cent of students and 52 per cent of employers thought error associated with a grade should not be indicated on the certificate, but over 47 per cent of students thought that error in a grade should be indicated.

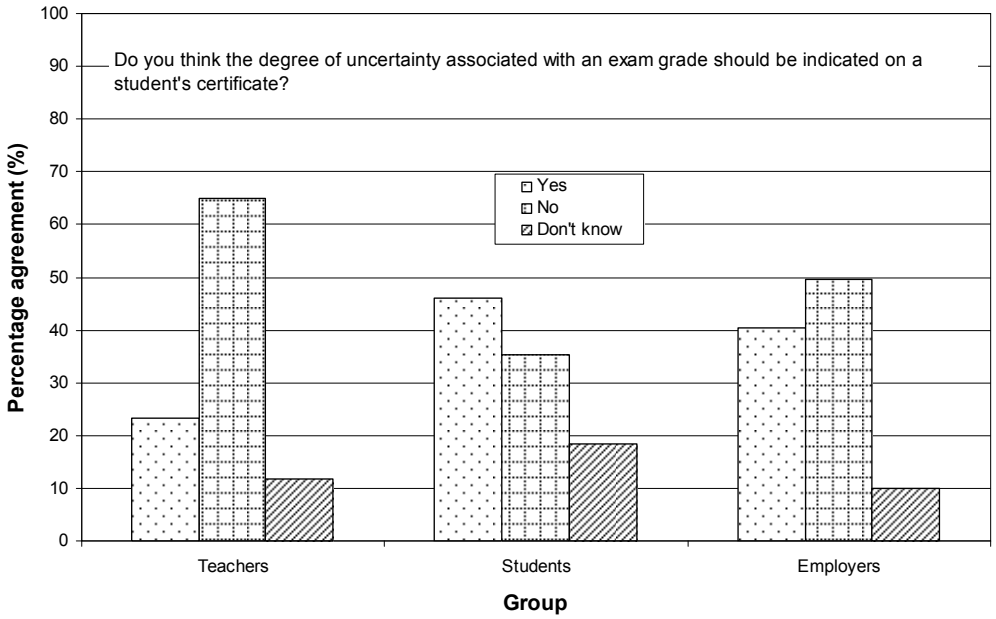

Figure 16: Attitudes towards reporting reliability information.

Approaches to trust

Questions in Topic 5 asked about respondents' approaches to trust. There were 20 sub-items in this topic.

Figure 17 illustrates the percentages of respondents who selected the three options for the statement about trusting people in general. The percentage of students who choose the option 'Most people can be trusted' was 36 per cent which is 5 per cent higher than that from the World Value Survey and is substantially lower than that for teachers (65 per cent) and employers (56 per cent), while the percentage who selected the option 'You need to be very careful in dealing with people' was about 53 per cent, which is substantially higher than that for teachers (31 per cent) and employers (42 per cent), suggesting that students were more cautious in trusting people generally.

Figure 18 shows views of the respondents on trusting people from different groups. Over 94 per cent of the respondents trusted family members and persons they knew personally. Less than 29 per cent of students trusted people they met for the first time, while that figure for teachers is 65 per cent for teachers and 59 per cent for employers. Students were also significantly less trustful of their neighbours than teachers and employers.

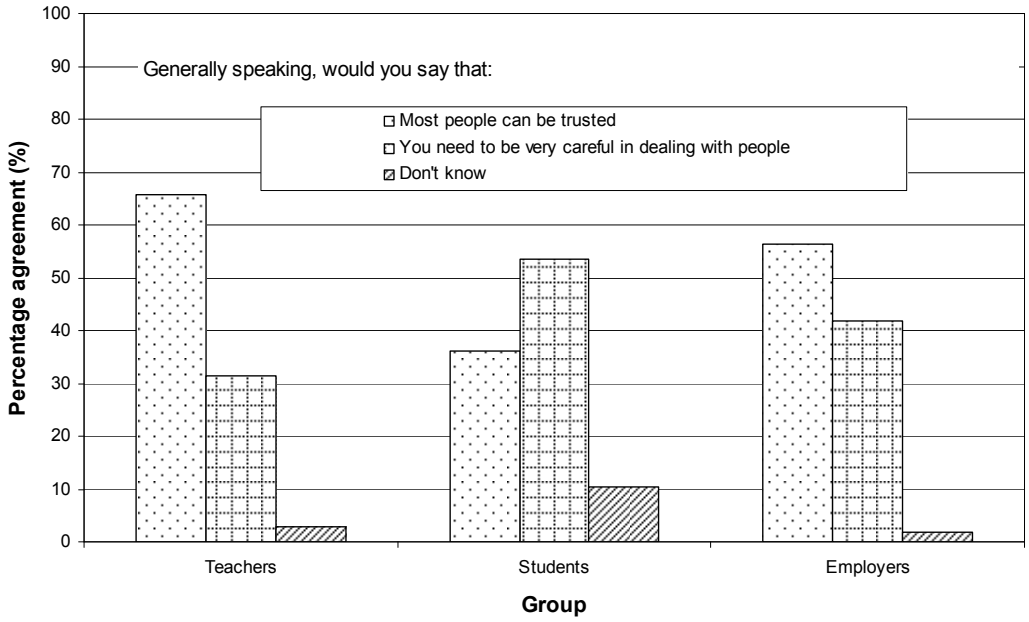

Figure 17: Views on trusting people in general.

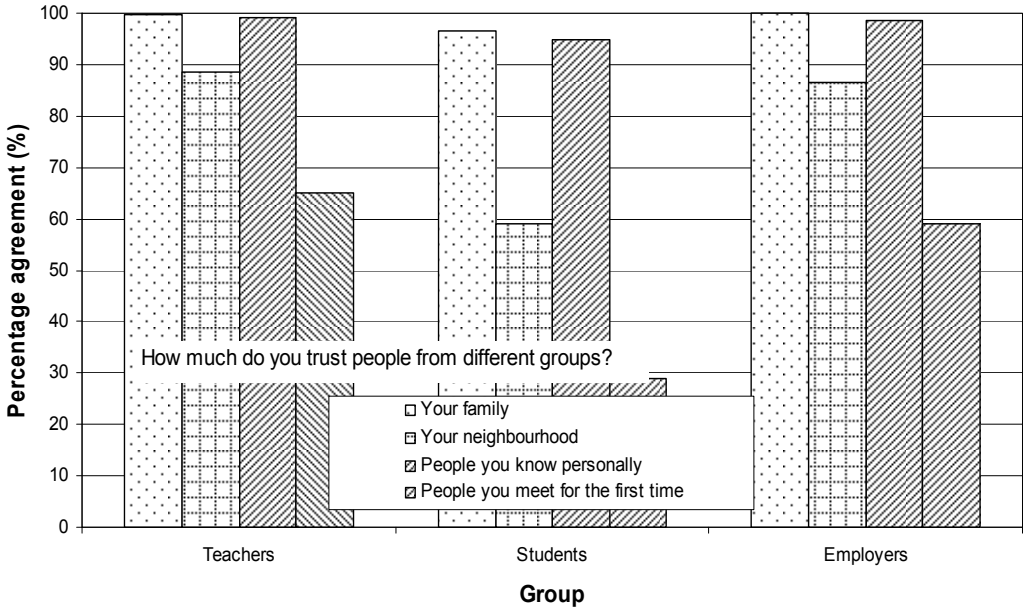

Figure 18: Views on trusting people from different groups.

Figures 19 and 20 show the percentages of respondents' who trusted people from different organisations and in different professions. Over 88 per cent of the respondents from all the three groups selected either 'Agree strongly' or 'Agree somewhat' to the statement 'I trust organisations if I have personal experience of them', and over 84 per cent of the respondents endorsed the statement 'I trust professionals with whom I come into personal contact', while less than 40 per cent of the respondents trusted organisations they heard about through the media. Slightly less than 60 per cent of teachers and employers and 63 per cent of students agreed that 'I trust organisations that have a strong technical focus.' The respondents were much less trustful of professionals working in commercial enterprises and government agencies, with endorsement rates of slightly over 50 per cent from teachers, slightly over 40 per cent from employers and less than 38 per cent from students.

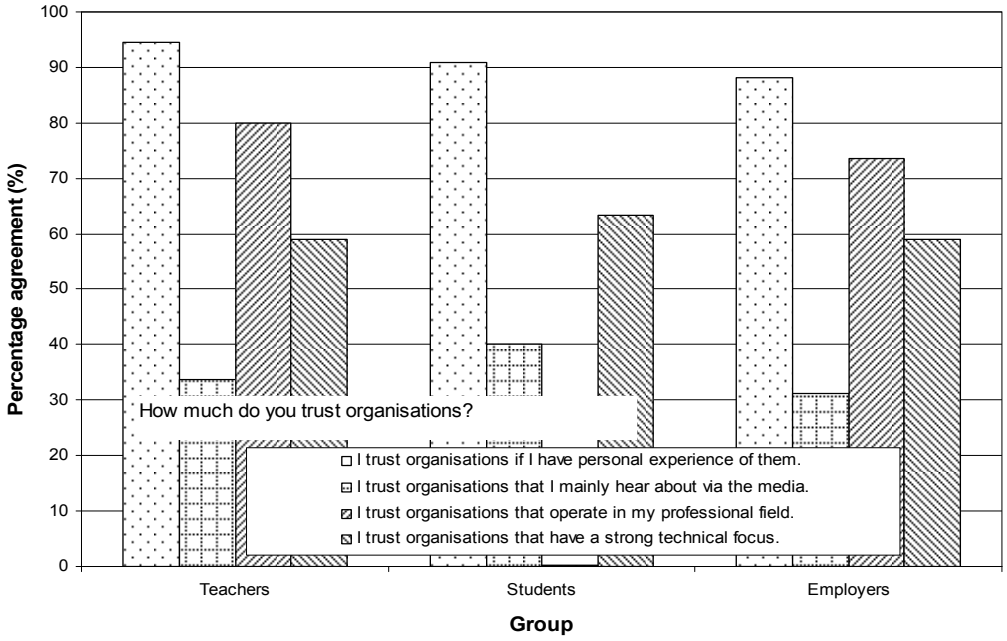

Figure 19: Views on trusting people from different organisations.

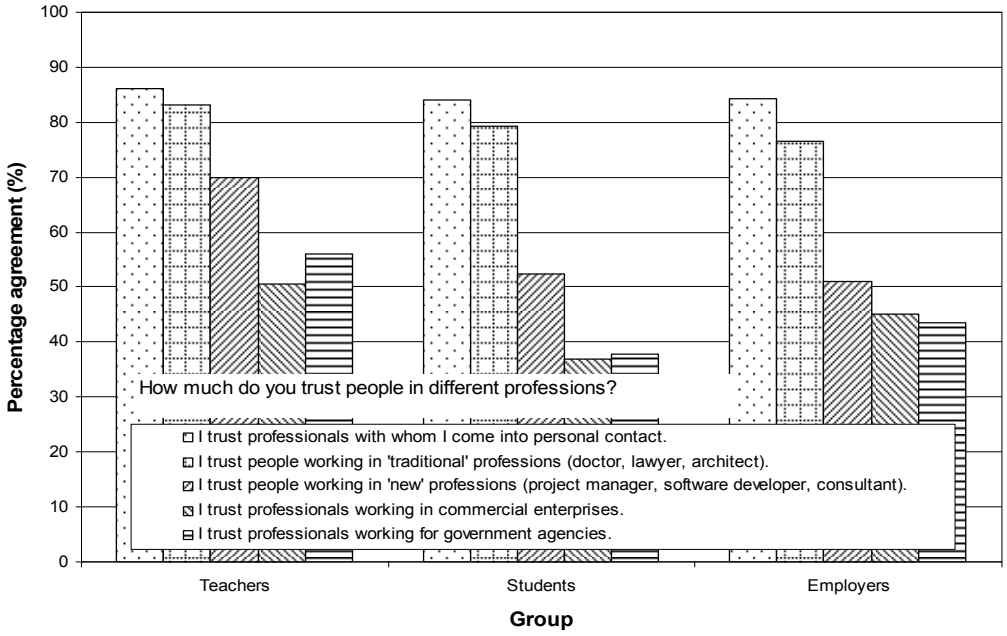

Figure 20: Views on trusting people from different professions.

Figures 21 and 22 depict the percentages of respondents who endorsed the statements about trusting the community and trusting government statistics. In general, teachers and employers were more trustful of the community than students. Only very small proportions of the respondents who had trust in government statistics and their appropriate use, and the way the statistics were produced. These findings are generally consistent with the findings from the qualitative study on public confidence in official statistics conducted by Wilmot et al. (2005).

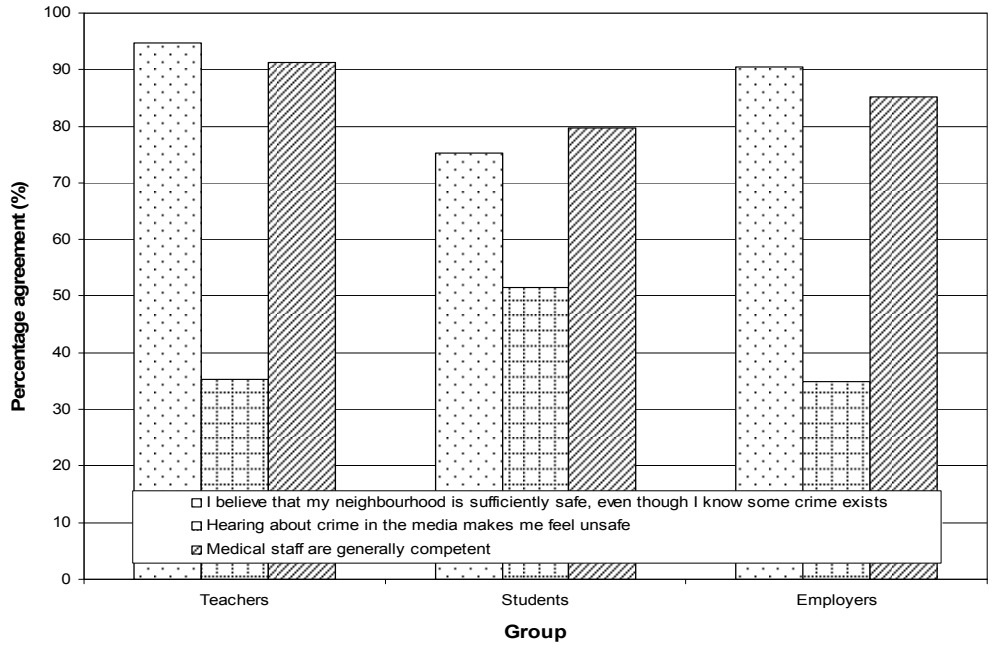

Figure 21 Views on trusting people in the community.

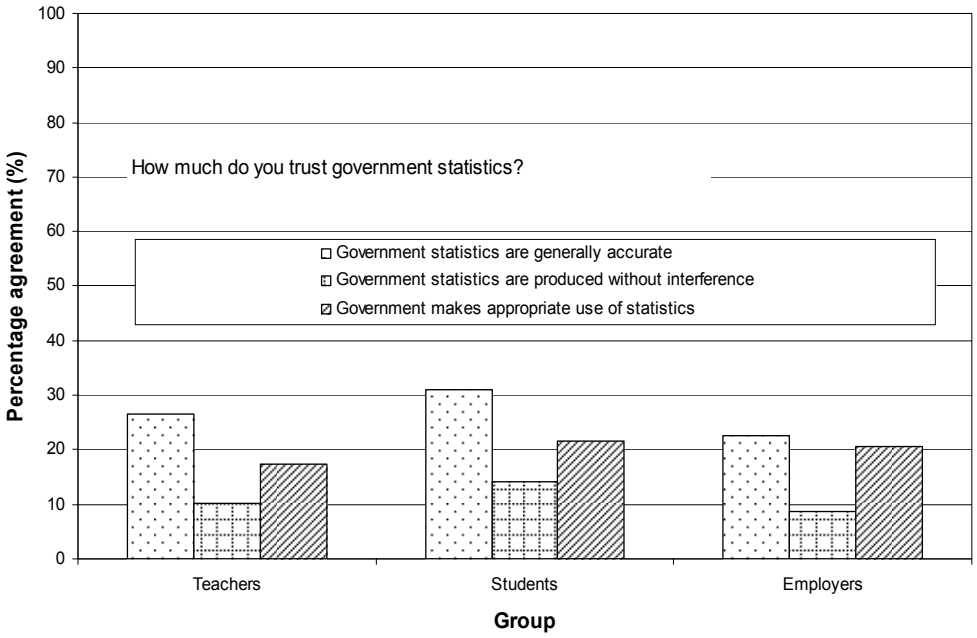

Figure 22: Views on trusting government statistics.

Relationships between belief, knowledge and approaches to trust and attitude to unreliability

The data collected was also used to explore the relationship between respondents' attitudes to unreliability in examination results and their attributes like belief about the examination system, knowledge about reliability concepts and approaches to trust. The internal consistency reliability of a topic, represented by Cronbach's alpha, to a certain degree reflects the uni-dimensionality of the topic in measuring the underlying construct, and values of Cronbach's alpha listed in Table 3 suggest that all the topics for the three groups had reasonably adequate internal reliabilities except for Topic D (attitudes to approaches for improving reliability) for teachers. An attempt was made to investigate the relationships between the topics, and Tables 5–7 list the correlation coefficients between topic scores for the three groups. Significant correlation exists between Topics C (attitudes to unreliability) and the other topics, indicating the influence of knowledge and beliefs and approaches to trust on attitudes to unreliability. The magnitudes of the correlations reflected the degree of the effect of the various attributes on attitudes to reliability. In view of the relatively low level of reliability of the topics, the values of the unattenuated correlation coefficients for the individual topics would be substantially higher than those listed in Tables 5–7.

Table 5: Correlations between scores on different topics for teachers.

	Topic A	Topic B	Topic C	Topic D	Topic E
Topic A	1				
Topic B	0.227**	1			
Topic C	0.063	0.152**	1		
Topic D	0.056	0.153**	0.143*	1	
Topic E	0.192**	0.016	0.111	–0.017	1

*Significant at p<0.05, ** significant at p<0.01

Table 6: Correlations between scores on different topics for students.

	Topic A	Topic B	Topic C	Topic D	Topic E
Topic A	1				
Topic B	0.352**	1			
Topic C	0.125*	0.219**	1		
Topic D	0.222**	0.317**	0.337**	1	
Topic E	0.271**	0.296**	0.296**	0.328**	1

*Significant at p<0.05, ** significant at p<0.01

Table 7: Correlations between scores on different topics for employers.

	Topic A	Topic B	Topic C	Topic D	Topic E
Topic A	1				
Topic B	0.433**	1			
Topic C	0.378**	0.406**	1		
Topic D	0.288**	0.378**	0.341**	1	
Topic E	0.368**	0.233**	0.259**	0.194**	1

*Significant at p<0.05, ** significant at p<0.01

Variation of knowledge about and attitudes towards unreliability between sub-groups

For each group, the respondents were further divided into four sub-groups according to the subjects they were teaching (for teachers), or studying (for students) or studied (for employers) (see Table 2), and variability in responses to questions in the five topic areas between the sub-groups within each group was examined. There was no substantial difference in the knowledge about and attitudes towards unreliability between sub-groups within each of the three groups. There was also no substantial variation between genders within each group, although considerable variability in the responses to some individual questions existed.

Concluding remarks

There has been little large-scale research to monitor the reliability of results from national tests and public examinations in England and limited understanding of the public's knowledge of and attitudes towards unreliability in examination results. The Ofqual Reliability Programme was designed to address these issues, which is important for improving the quality of the examination system further. It is essential to understand the public's attitudes towards uncertainty in examination results when developing regulatory policy on reliability in order to improve the examination system further and to increase the public's confidence in it. Results from this study indicated that knowledge about and attitudes to unreliability in examination results vary between respondents for the three stakeholder groups investigated. The majority of the respondents from the three groups appeared to understand the assessment process and the factors that affect students' performances in examinations. The respondents to a degree 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.

Results from this study indicated that respondent's attitudes to unreliability were generally positively correlated to their knowledge about aspects of reliability, beliefs about the examination system and approaches to trust. A substantial proportion of respondents from the three groups lacked awareness of some aspects of reliability. There a need to educate the public to understand reliability concepts and the existence of uncertainty in examination results. This was also recognised by many assessment experts (see Ofqual, 2009; Boyle et al., 2009). Further study in this area would involve conducting experiments to investigate how attitudes to unreliability could be affected by the increase in understanding of aspects of reliability. It is also important to explore effective ways of educating the public to understand reliability concepts.

Although the findings from this study generally supported the findings from the qualitative investigations by Ipsos MORI (2009) and Chamberlain (2010), the difference in the implications of the findings from the two different approaches has to be recognised. While the views expressed at the workshops or focus group discussions were under controlled environment, the self-reported attitudes through the responses to a questionnaire exclude external influences. It is very likely that the use of workshops or focus groups would have helped the participant to develop knowledge and views about reliability. However, since the attitudes of the participants were not measured before and after the workshops/focus group process, it was impossible to assess the impact of the increase in knowledge about reliability on the change in their attitudes. As discussed previously, the present study to a certain degree was able to establish the relationship between attitudes to unreliability and knowledge and other attributes of the respondents.

It is also realised that the present study was restricted to only three groups of the public: teachers, students and employers to whom the reliability of examination results would probably be more important than to other groups. Further research would involve studying the perceptions of reliability from other stakeholder groups such as parents and the general public.

References

American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) (1999) *Standards for educational and psychological testing*. Washington, DC: AERA.

Black, P. and Wiliam, D. (2005) Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The curriculum journal*, 16(2), 249–261.

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: http://www.ofqual.gov.uk/files/2009-09-iaea-no-news-is-good-news.pdf.

Bradberry, T. (2007) *The personality code*. New York: Putnam.

Bradshaw, J. and Wheater, R. (2009) *International survey of results reporting*. Ofqual, Coventry, UK Available online at: 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*. Available online at: http://www.ofqual.gov.uk/files/Ofqual_10_4708_public_perceptions_reliability_report_08_03_10.pdf.

Coren, S., Ward, L.M. and Enns, J.T. (1999) *Sensation and perception*. New York, NY: Harcourt Brace.

Fukuyama, F. (1995) *Trust: social virtues and the creation of prosperity*. New York, NY: Free Press.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2004) *Questionnaire methodology*. Hoboken, NJ: John Wiley & Sons, Inc..

Hardin, R. (2002) *Trust and trustworthiness*. New York, NY: Russell Sage Foundation.

Ipsos MORI (2009) *Public perceptions of reliability in examinations*. Available online at: http://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: www.ofqual.gov.uk/files/2009-02-26-ofqual-perceptions-of-alevels-gcses.pdf.

Lycan, W.G. (ed.) (1999) *Mind and cognition: an anthology*, 2nd Edition. Malden, Mass: Blackwell Publishers.

McLeod. C. (2002). *Self-trust and reproductive autonomy*. Cambridge, MA: MIT Press.

Newton, P.E. (2005a) The public understanding of measurement error. *British Education Research Journal*, 31, 419–442.

Newton, P.E. (2005b) Threats to professional understanding of assessment error. *Journal of Education Policy*, 20(4), 457–483.

Ofqual (2009) *The reliability programme – technical seminar report*. Ofqual: Coventry, UK. Available online at: http://www.ofqual.gov.uk/research-and-statistics/research-reports/92-articles/20-reliability

O'Neill, O. (2002) *A question of trust: the BBC Reith Lectures 2002*. Cambridge: Cambridge University Press.

Pearson, S.D. and Raeke, L.H. (2000) Patients' trust in physicians: many theories, few measures, and little data. *Journal of General Internal Medicine*, 15(7), 509–513.

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: http://www.ofqual.gov.uk/research-and-statistics/research-reports/.

Putnam, R.D. (1995) Bowling alone: America's declining social capital. *Journal of Democracy*, 6(1), 65–78.

Reiss, S. (2000) *Who am I?: the 16 basic desires that motivate our actions and define our personalities*. New York: Tarcher/Putnam.

Tschannen-Moran, M., and Hoy, W.K. (2000) A multidisciplinary analysis of the nature, meaning, and measurement of trust. *Review of Educational Research*, 70, 547–593.

Wilkinson, R. and Pickett, K. (2009) *The spirit level*. London: Allen Lane.

Wilmot, A., Jones, J., Dewar, A., Betts, P., Harper, R. and Simmons, E. (2005) *Public confidence in official statistics. A qualitative study on behalf of the Office for National Statistics and the Statistics Commission*. Available online: http://www.ons.gov.uk/about/who-we-are/our-services/data-collection-methodology/reports-and-publications.

World Values Survey (WVS) (undated) *World values survey: the world's most comprehensive investigation of political and sociocultural change*. Available online at: http://www.worldvaluessurvey.org/.

Appendix A Questionnaire for attitudes to unreliability and inconsistency in examination results

About you

Please answer the following questions about yourself:

a. Please indicate your primary occupation. Are you a:

Please select one option

Student
Teacher
Employer

b. Gender

Please select one option

Male ☐ ☐ Female

c. Age [DISPLAY AS DROP DOWN LIST]

Please select from the drop down list

Display full list of years ranging from 16 years old to 65 years old

Are you currently in: [STUDENTS ONLY]

Please select one option

Year 12 or 1st year of college
Year 13 or 2nd year of college

d1. Which subjects do you teach? [TEACHERS ONLY]

Please select all that apply

English	
Maths	
Science	
Languages (e.g. French, German)	
Technologies (e.g. design technology, graphics)	
Art	
Drama	
Music	
Geography	
History	
PE	
ICT	
Religious studies	
Other	

d2. Which subjects do you study? [STUDENTS ONLY]

Please select all that apply

English	
Maths	
Science	
Languages (e.g. French, German)	
Technologies (e.g. design technology, graphics)	
Art	
Drama	
Music	
Geography	
History	
PE	
ICT	
Religious studies	
Other	

d3. What is the highest level of education you have attained? [EMPLOYERS ONLY]

Please select one option

Finished secondary school (at age 15/16)
Studied at college or school sixth form (after age 16)
Studied at university or polytechnic for a degree

d4. What genre is your qualification in? [EMPLOYERS ONLY]

Please select one option

Arts
Mathematics
Sciences
Social sciences
Other

Part A: This section asks about your experiences, knowledge and beliefs about the examinations system

1a. Students

Please indicate your level of agreement with each statement.

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
Doing exams is unreasonably stressful						
In general, I get the grades I deserve from my exams						
In general, my experience of the exams system is acceptable						

1b. Teachers

Please indicate your level of agreement with each statement.

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
In general, my students get the grades they deserve in exams						
Most exams are fair to most students						
My students are generally well prepared for exams						

1c. Employers

Please indicate your level of agreement with each statement.

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
In general, students get the grades they deserve in exams						
Grades generally reflect students' actual ability						
We select candidates for interview primarily based on their exam results						
We sometimes use our own tests to assess candidates' skills						

2. What is your view of the national examination system?

Please choose the statement that fits your view most closely

The national exam system is doing a very good job	
The national exam system is doing a good job	
The national exam system is doing a good job but can be improved further	
The national exam system is not doing a good job and should be reformed	
I don't know	

3. To what extent do you agree or disagree with the following statement?

Please indicate your level of agreement

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
I have confidence in the national examination system.						

4. Please indicate the extent to which different features of the examination system are important in your opinion.

Please select one option in each row

It is important that…	Very important	Fairly important	Fairly unimportant	Not important at all	Don't know
Exam papers have demanding questions					
Exam papers are free of typos and mistakes					
If people take the same exam from different exam boards, they will get comparable results					
Exam standards stay the same over time					
Exams are fair to all people in society					
Exam results are consistent; most people would get the same result if they sat the exam again					
Most people get the exam results that they deserve					

5. How far do you agree that the following creates trust?

Please indicate your level of agreement for each statement

This creates trust in the exam system...	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
Subject experts making sure that exams measure the right things and that they are at the right level						
Statistical procedures making sure that exams give the right results						
The exams system is a national system						
Exam boards have the necessary expertise and experience						

Part B: This section is about your awareness of reliability issues in relation to exams

6. How far do you agree that the following affects a student's exam score on an exam?

Please indicate your level of agreement for each statement

A student's score is affected by…	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
How much the student knows about the subject						
How well the student is prepared						
How difficult the question paper is						
How well the student feels on the day						
Who marks the question paper						

7. How far do you agree that the following introduces inconsistency into exam results if the exam procedure was repeated?

Please indicate your level of agreement for each statement

The consistency of exam results is affected by…	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
The occasion (e.g. if assessed on another day, the student might have performed differently						
Test questions (e.g. if a different test had been set, the student might not have been disadvantaged by the wording of an essay question)						
Marking inconsistency (e.g. if a different marker had been assigned, the student might have achieved a different result)						
The awarding process (e.g. if a different team of people had been involved in the process of awarding, different grade boundary marks might have been set)						

8. The amount of misclassification

In your view, what is the proportion of students who have a GCSE grade that does not reflect his/her actual ability?

Please select one option in each row

	Less than 10%	10% – 20%	21% – 30%	31% 50%	More than 50%	Don't know
In mathematics						
In science						
In English						

Part C: This section asks about your attitudes to exam accuracy and reliability

9. Attitudes to accuracy in the exams system

Please choose the statement that fits your view most closely.

Any level of error has to be unacceptable – even just one candidate getting the wrong grade is entirely unacceptable.	
Exam results are essentially an estimate – a certain amount of error is inevitable.	
Don't know	

10. Inherent variability versus avoidable mistakes

Please choose the statement that fits your view most closely.

All inaccuracy has to be removed from the system, there's no such thing as 'inevitable and acceptable variation'	
There's a difference between an avoidable mistake – like a typo on a paper – and something inevitable like inconsistency between two markers.	
Don't know	

11. To what extent do you agree or disagree with the following statements?

Please indicate your level of agreement for each statement

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
The performance on the day of an exam can be affected by feeling stressed or unwell, but it is "just one of those things"						
Students need to be held accountable for how they perform on the day of an exam						
Exam boards should do everything they can to minimise inconsistency from their processes						

12. To what extent do you agree or disagree with the following statements?

Please indicate your level of agreement for each statement

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
Error in the mark a student receives which does not affect a grade overall is not a cause for concern						
Error which results in a student receiving a different grade to the one they deserve is serious						
Error that changes a grade C to a grade D in a GCSE exam is particularly important						

13. Would you view error in exam scores differently depending upon whom it affected?

Please indicate how acceptable error in exam scores would be if it affected the following people or groups of people.

	Completely acceptable	Somewhat acceptable	Somewhat unacceptable	Completely unacceptable	Don't know
You [STUDENTS ONLY]					
A member of your family					
Your school as a whole [STUDENTS AND TEACHERS ONLY]					
The education system as a whole [ALL]					

14. What is your view of the system for appealing exam results?

Please indicate your level of agreement with each statement

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
The system for appealing exam results is fair						
The system for appealing exam results is too generous to students						
Teachers sometimes appeal results inappropriately						
The appeals process helps the exams system to produce accurate results						

Part D: This section is about what could be done to improve reliability

15. To what extent do you agree or disagree with the following statements?

Please indicate your level of agreement with each option.

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
Government and exam boards need to be open about how much uncertainty there is in exam results						
Publishing information about uncertainty in exam results might reduce public confidence						
Exam boards should not be embarrassed by a certain amount of inaccuracy in scores – it is inevitable						

16. Do you think the degree of uncertainty associated with an exam grade should be indicated on a student's certificate?

Please select one option

Yes	
No	
Don't know	

17. All the options below could increase the certainty of exam grades.

Please indicate your level of agreement with each option.

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
Improve training for markers						
Have two markers for essays						
Use more multiple-choice questions						
Have longer tests						
Use more teacher assessment for awarding qualifications						

Part E: This section is about your approach to trust.

We would like to understand and know more about your attitudes towards trust in general situations to give some comparison to your responses in previous questions. Answers will not be used to identify any individuals.

18. Generally speaking, would you say that:

Please select one option

Most people can be trusted	
You need to be very careful in dealing with people	
Don't know	

19. How much do you trust people from different groups?

Please indicate how much you trust people from each group.

	Trust completely	Trust somewhat	Do not trust very much	Do not trust at all
Your family				
Your neighbourhood				
People you know personally				
People you meet for the first time				

20. How much do you trust organisations?

Please indicate your level of agreement with each statement.

	Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly
I trust organisations if I have personal experience of them				
I trust organisations that I mainly hear about via the media				
I trust organisations that operate in my professional field [ASK TEACHERS AND EMPLOYERS ONLY]				
I trust organisations that have a strong technical focus				

21. How much do you trust people in different professions?

Please indicate your level of agreement with each statement.

	Agree strongly	Agree somewhat	Disagree somewhat	Disagree strongly	Don't know
I trust professionals with whom I come into personal contact					
I trust people working in 'traditional' professions (doctor, lawyer, architect)					
I trust people working in 'new' professions (project manager, software developer, consultant)					
I trust professionals working in commercial enterprises					
I trust professionals working for government agencies					

22. Please indicate your level of agreement with each statement about trust in the community.

	Agree strongly	Agree somewhat	Disagree somewhat	Strongly disagree	Don't know
I believe that my neighbourhood is sufficiently safe, even though I know some crime exists					
Hearing about crime in the media makes me feel unsafe					
Medical staff are generally competent					

23. How much do you trust government statistics?

Please indicate your level of agreement with each statement.

	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Don't know
Government statistics are generally accurate						
Government statistics are produced without interference						
Government makes appropriate use of statistics						

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

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

© Crown copyright 2010

Office of Qualifications and Examinations Regulation

Spring Place
Coventry Business Park
Herald Avenue
Coventry CV5 6UB

2nd Floor
Glendinning House
6 Murray Street
Belfast BT1 6DN

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

www.ofqual.gov.uk