



Public perceptions of reliability in examinations

A research study conducted for Ofqual

Final Report
May 14th 2009

Legal notice

© 2009 Ipsos MORI – all rights reserved.

The contents of this report constitute the sole and exclusive property of Ipsos MORI.

Ipsos MORI retains all right, title and interest, including without limitation copyright, in or to any Ipsos MORI trademarks, technologies, methodologies, products, analyses, software and know-how included or arising out of this report or used in connection with the preparation of this report. No license under any copyright is hereby granted or implied.

The contents of this report are of a commercially sensitive and confidential nature and intended solely for the review and consideration of the person or entity to which it is addressed. No other use is permitted and the addressee undertakes not to disclose all or part of this report to any third party (including but not limited, where applicable, pursuant to the Freedom of Information Act 2000) without the prior written consent of the Company Secretary of Ipsos MORI.

Contents

Introduction	5
Aims and objectives	5
Methodology and recruitment.....	6
The workshops and discussion guide	7
Interpretation of the data	7
Publication of the data.....	8
Acknowledgements	8
Summary	10
The concept of “error” in exam grades	10
The different types of error.....	10
Information and transparency.....	11
Consistency and continuity	12
Talking about exam error	12
The concept of ‘error’ in exam grades	15
Importance and consequence of exam error.....	16
Concerns about the level of exam error	17
The different types of error	19
Pupil-related error	19
Examiner-related error	21
Test-related error	23
Information and transparency.....	27
The case for transparency	27
Appeals and special consideration.....	30
Consistency and continuity	33
Awareness and opinions of Ofqual.....	34
Talking about error.....	37
Appendices	39

Discussion guide 39

Introduction

Introduction

Ipsos MORI was commissioned by the exams regulator Ofqual to conduct qualitative research exploring the public perceptions of the causes of error in exam grades and its sources.

According to some sources, approximately 30% of students receive a grade in national tests at Key Stage 2 that is not in line with their level of attainment at the time of the testing.¹ A range of factors can contribute to the awarding of a grade that is either higher or lower than the grade the student should have received.

The causes of error in exam results are varied and complex, and for the layperson, can be difficult to understand. However, Ofqual feels that they have a responsibility to communicate with the public about this issue, and this study was commissioned as the first step in a wider research programme to understand how best to approach discussions on this topic.

Aims and objectives

The overall aim of the research was to explore perceptions of error in exam results and errors in the assessment process among different groups within the population, both in terms of their gut reactions, and more informed perspectives on this topic. Due to the complex nature of what was being discussed, the discussion guides we developed were deliberately not prescriptive, to allow participants the opportunity to express their true views. However, some of the key messages the discussions aimed to extract were:

- Do people understand what Ofqual means when they discuss ‘error in exam results’, and do they share their views?
- What, if different to Ofqual’s view, do people consider error in exam results to be?
- What do people consider to be error and what do they not think of as error?
- How do people feel about different types or causes of error – are they acceptable or not?
- Who, if anyone, should be responsible for the error in exam results or errors in the assessment process?

¹ Source: William, D. (2001). *Level best?* London: ATL.

Methodology and recruitment

The fieldwork for the project involved two, three-hour workshops. The first was held in Birmingham on January 20th 2009, and the second in London on January 21st 2009. Both workshops were held in the evening, and participants were each paid a £50 incentive to thank them for their time.

Each workshop comprised 36 participants, split into six smaller groups as follows:

- Eight teachers (of secondary education students)
- Six students (aged 16-19 and in secondary education)
- Six parents (with children in secondary education, years 7-13)
- Six members of the general public (either with no children, young children under 10, or older children aged 20+)
- Six employers (involved with recruitment either in an HR or line management role)
- Four examiners

All groups were recruited by Ipsos MORI recruiters who were briefed about the project and provided with a screener questionnaire to ensure that only suitable participants were invited to the workshops.

Teachers and examiners were both recruited through schools in the Birmingham and London areas by telephone.² To ensure a good spread of opinion during the workshops, recruiters were instructed to obtain:

- No more than two teachers from any one school
- Teachers/examiners from a range of subject areas
- A mix of independent and maintained schools
- A mix of experience levels/seniority

² Recruiters were able to recruit examiners via schools as examiners often work as teachers too. Prior to coming to the workshops they were asked to attend in their capacity as an examiner, rather than as a teacher.

Employers were also recruited by telephone, and it was ensured that they worked in a variety of industries and were a mixture of human resources staff and line managers involved in recruitment.

The general public, students and parents were recruited face-to-face on the street.

In order to ensure that those attending the workshops had a wide range of experiences and views, all recruiters were asked to ensure a good mix of participants in each group, with regard to:

- Gender
- Age
- Social class; and
- Ethnicity

In some cases, loose quotas were set, and the variation within the groups was successfully achieved.

The workshops and discussion guide

The workshops were facilitated by Ipsos MORI staff: one chairperson, six moderators, six note-takers and a hostess to welcome people to the event.

It was decided that participants would be grouped according to 'type' and remain in this group for the course of the evening, rather than having mixed groups in which, for example, examiners, teachers and students would all be thrown together. It was felt that this approach would enable each group to focus on their own views and experiences, providing clearer evidence of how they felt about the issue of exam error collectively.

The discussion guide was developed in consultation with Ofqual, and it was decided that a varied structure would produce the most interesting findings, and ensure that participants remained alert and engaged with the topic. As a result, each workshop involved a series of smaller discussions in break-out groups, plenary sessions, and group presentations. The full discussion guide is included in the appendix.

Interpretation of the data

While qualitative research was the most appropriate methodological approach for this study, it is important to bear in mind that it utilises smaller samples that are chosen purposively, to ensure representation of a full range of views within the sample. Qualitative research is

designed to be illustrative and does not look to produce statistics; this needs to be taken into account when interpreting the research findings. In addition, it is important to bear in mind that the research deals with perceptions rather than facts (though perceptions *are* facts to those that hold them).

Throughout the report we have made use of verbatim comments to exemplify a particular viewpoint. It is important to be aware that these views do not necessarily represent the views of all participants. Where verbatim comments have been used, the respondent's attributes are given in the following order: respondent 'type', gender and location.

Publication of the data

As with all our studies, these findings are subject to Ipsos MORI's standard Terms and Conditions of Contract. Any press release or publication of the findings of this research requires the advance approval of Ipsos MORI. Such approval will only be refused on the grounds of inaccuracy or misinterpretation of the findings.

Acknowledgements

Ipsos MORI would like to thank Annette Kinsella and Paul Newton at Ofqual for all their assistance with this project. We would also like to thank the participants who gave up their time to participate in the research.

©Ipsos MORI/J34861

Checked & Approved:

Fiona Johnson

Nathan Burrough

Ali Ziff

Benjamin Collins

Summary

Summary

The concept of “error” in exam grades

The Ofqual definition of error in exam grades refers specifically to occasions when a student does not receive a grade in keeping with their level of attainment at the point at which they are examined. Factors which may cause an error of this kind may include student-related error, examiner-related error or test-related error in the assessment process.

Participants tend to be in agreement with Ofqual that these factors can play a part in causing error in exam grades, but, notably, they spontaneously mention the effect of teaching as an error-contributing factor too. This is where participants differ from Ofqual, who considers the effect of teaching to be something that occurs prior to the examination process, as opposed to the other factors which occur at the point of testing or during the marking process.

A key finding from the study is the clear demarcation in the minds of the public between inevitable errors in the assessment process (bad luck or simple mistakes which cannot be eliminated entirely) and preventable errors, where the latter is always unacceptable.

Participants tend to be most preoccupied with preventable errors in the assessment system; a systemic error for example, such as a bad exam paper, is seen as unacceptable, and the exam board should be held responsible.

A key concern that participants have is the consequence of the error in exam results, as to them, an error in the *mark* which does not affect the grade received is not as serious as an error of the same size that affects the grade. Errors at the C/D grade boundary are seen as particularly significant.

People are not overly concerned, however, about the level of error in the system. Some groups offer rough estimates of what they imagine the level of error in exam grades to be (these range between 1-20%), but on the whole they tend to be more concerned about why error occurs, and how it can be controlled and prevented.

The different types of error

Pupil-related error such as being ill on the day of the exam, arriving late and missing the start of the exam or revising the wrong things are not spontaneously thought of as errors by most people. Teachers and examiners are harder on students than other participants in this regard, but even parents and students, the two groups who might be expected to be more

lenient about this are also pragmatic about it and feel that being well prepared and keeping calm are the responsibilities of the student. It is the use of the word 'error' which causes the problem here, as these types of error are seen as an inevitable part of the process (and indeed, of life), as opposed to a mistake that could have been avoided entirely or rectified.

For many participants, examiner-related error tends to be more 'top of mind' than pupil-related error when considering what might go wrong during the examination process. Issues such as examiners misinterpreting what the student is trying to say or making mistakes when adding up marks are considered to be more serious than pupil-related error, but people still remain understanding about the difficulties and pressures examiners are under and accept that it is only natural for people to make mistakes. Students are less accepting of examiner-related error than others, particularly when they learn that examiners are able to mark papers at home, and not in a controlled environment. Most participants tend to agree, however, that although mistakes such as these are unacceptable, they are inevitable to some extent.

Errors relating to the examining method, namely the paper, are seen by all groups as the most concerning and unacceptable type of errors in the assessment process, as they should, in all circumstances, be preventable. Teachers in particular, who feel a sense of loyalty to their students, take a zero tolerance approach to mistakes on exam papers, which they feel are completely avoidable, and they have the support of most other groups on this issue. The examiners, however, are less critical, perhaps due to their knowledge of the current systems in place to prevent errors like these.

Information and transparency

Participants in the research express a desire for information about the exam system, for example, why there are various exam boards, not just one, and where different "system" responsibilities lie (between DCSF, QCA, Ofqual, exam boards and schools). For parents this is particularly important, and some say they would like to attend information sessions at their child's school to learn more. (In fact, many schools do offer sessions such as these, but attendance is often poor).

There is also a suggestion by many participants that more public discussion about error in exam grades and errors in the assessment process would be useful, but examiners and teachers are wary of information about this subject being widely disseminated. The general public and students are undecided about whether publicising the issue of error in exam grades would be a good thing, and feel it would need to be handled carefully if it were to be discussed with a wider audience.

The appeals and special consideration system is something that many people would like to know more about, but there is a strong sense that more awareness about them could lead to their abuse. Several examiners and students are able to give examples of special consideration being awarded in some “unsuitable” circumstances, but not in others where they would be appropriate.

Consistency and continuity

Consistency and continuity are a key concern of participants. Most people question why there are so many different bodies involved in the exam system, for example, several exam boards apparently duplicating the same role, as they feel that this in itself is probably contributing to errors.

Teachers and examiners know more about how the system operates than most, but do not like the competitiveness brought about by having several exam boards. Teachers feel it would be clearer if everything was controlled by one body – the QCA. Employers, too, do not recognise any benefit from having more than one exam board, as in terms of their roles in recruitment, they do not know what a grade from a specific board means compared to a grade from a different board.

With regard to how Ofqual is viewed, there is very little awareness of how Ofqual fits into the current examination system. Only examiners are aware of Ofqual, and question how Ofqual is separate from QCA. There is, however, a positive reaction to its mission statement from most groups, though it is suggested that Ofqual will need to prove that Ofqual is acting on its plans. Teachers and examiners find Ofqual's aims and objectives a little vague and in need of fleshing out. They have their own ideas about what Ofqual should be doing, most notably having a role in monitoring the exam boards to ensure consistency.

Talking about exam error

The original aim of this research was to investigate the views of various population sub-groups on the topic of “exam error” (i.e. error in exam grades and errors in the assessment process). Although some interesting findings on their opinions emerged, the study also provided a previously unforeseen opportunity to consider *how* discussions about the concept of exam error should be handled.

The experience of Ipsos MORI and Ofqual on this project suggests that even a slight variation or imprecision in the terminology used in discussions with the general public can alter the way people think and form their opinions on this topic. For example, we sometimes

referred to pupil illness, or marking inconsistencies between examiners as 'errors' and sometimes as 'causing an error in the mark/grade awarded' (in the sense of the student not getting the grade they deserved). Participants rejected the notion of pupil illness being an error, framing this as 'bad luck', 'tough' or 'just life'. However, they took a softer line on the idea of pupil illness as a cause of error in the grade received. None felt it was as significant as other causes of error (the inevitable versus avoidable demarcation was key here), but they were able then to consider what, if anything, should be done about it.

Participants commented spontaneously on the complexity of the topic, which reinforces the importance of precision when handling the topic in the future.

Fundamentally, the word 'error', in the technical sense it is used by Ofqual, is one which the general public does not use or recognise.

For Ofqual, it can have a relatively benign "no blame attached" connotation. For the general public, error means mistake, and if a mistake has been made, someone is to blame/can be held responsible. Effective communication between Ofqual and the public on error in the future will be contingent then, on finding a more 'neutral' and in-common way of describing the issues and contributing factors.

The concept of 'error' in exam grades

The concept of 'error' in exam grades

The Ofqual definition of error in exam grades refers specifically to occasions when a student does not receive a grade in keeping with their level of attainment at the point at which they are examined. Put more simply, it is a situation where the student does not get the grade they deserve. A wide variety of factors may cause an error of this nature, which could be related to examiners, the test paper, and the students themselves; more detail on these can be found in the next section.

One point on which participants in the research differ from Ofqual, is the issue of the effect of teaching. Ofqual deliberately omit this, because they feel that although teaching can affect the attainment of a student it does not produce error in the measurement as this happens prior to the examination process, unlike the other factors which occur at the point of testing or during the marking process. However, some groups spontaneously include it in their consideration of the factors which could prevent a student getting the grade they deserve.

"If the teacher hadn't taught the syllabus ... that to me would be an error."

Teacher, Male, London

Aside from this however, participants tend to be in agreement with Ofqual in their understanding that certain factors – for example, student illness – have a part to play in causing error in exam grades.

A key finding though, is that there is a clear demarcation in the minds of the public between inevitable errors in the assessment process (bad luck or simple mistakes which cannot be eliminated entirely) and preventable errors, where the latter is always unacceptable.

Preventable error is what pre-occupies participants. A single mistake by an examiner is widely regarded as unfortunate but inevitable ("nobody's perfect") and tolerable. On the other hand, a systemic error, such as a bad exam paper, is unacceptable, and the exam board should be held responsible. Teachers are particularly concerned about these types of mistakes which they feel should be picked up prior to the exam by proper checking systems and processes.

"Misprints are totally unacceptable."

Teacher, Male, Birmingham

Importance and consequence of exam error

It should be noted that participants' views on error vary depending on the consequences of the error for the student. As well as certain causes of error being classed as either acceptable or unacceptable, the impact of the error influences perceptions of its significance.

Error in the *mark* a student receives which does not affect their grade overall is not a wide cause for concern. Error which results in a student receiving a different grade to the one they deserve is seen as far more serious.

Employers and teachers are particularly concerned about the error at the C/D grade boundary. This is seen as more significant than error at the A/B grade boundary, for example, as grades of C and above are widely recognised as 'good' grades, and it is the proportion of these grades received by students in a school that contribute to the school's overall rating in the league tables.

Employer concerns relate to the fact that many will dismiss candidates who do not achieve a C or above by way of narrowing down the number of applications they look at. They recognise that error at this boundary may mean the difference between someone getting a job or not.

"Employers may disregard people with D or lower. If you're below this, then your experience becomes irrelevant."

Employer, Male, Birmingham

Teachers' concerns relate to the fact that the number of pupils achieving 5 A*-C grades at GCSE is a key measure of their school's and, indeed, their own success. Many feel that an error which causes a script to be 'out' by two or three marks but doesn't influence the grade is less significant than an error of similar size that does, particularly when it means the difference between a C or a D grade. Thus the significance of the consequence for the student correlates to how acceptable the error is perceived to be.

"The difference between a C and D for some schools can be very consequential. If it's not changing the grade it doesn't matter. It's those boundary ones that really matter."

Teacher, Female, London

Concerns about the level of exam error

The majority of participants admit they have no idea how much error is present in the system. Some groups give an indication of what they think it might be, but it is important to note that their estimations are based purely on supposition and gut instinct.

The examiners in Birmingham believe that at least 10% of exam grades are incorrect, but that figure only takes into account those whose grade is lower than they deserve. They feel that an equal number may be receiving grades higher than those they really deserve (something people don't really like to talk about!).

Teachers in London come up with a lower figure, estimating error in exam grades to be between only one and five per cent. Parents in London make the highest estimate, placing overall error in exam grades at 10-20%.

Many of the participants in Birmingham believe the examiners' assessment of the level of error at 10% to be correct, and although some, particularly the students, express initial concern about this, it is not an issue that is dwelled upon.

Instead, showing great pragmatism, participants tend to focus their attention on *why* error exists in the system in the first place, what can be done to prevent or control it and how it can be rectified when it does happen.

The different types of error

The different types of error

Errors can be divided into two distinct categories, to which various factors can contribute. These are known as procedural error and measurement error.

Procedural error (sometimes referred to as human error) relates to the more obvious mistakes that occur during the assessment process. These may occur during the set-up process, when, for example the exam paper does not cover the complete syllabus or contains a spelling mistake which affects how a question is understood. They may also occur during the marking process, if perhaps the examiner skips a page of the script, or incorrectly adds up the total marks. On the whole, participants believe these mistakes are entirely avoidable.

Measurement error, on the other hand, is something that participants feel can never be completely eliminated from the assessment process. Things such as the student being unwell on the day of the exam, or the examiner misinterpreting the candidate's answer, cause a measurement error and can contribute to a student not receiving the grade they deserve.

Due to the technicalities and subtleties of discussing error in terms of procedure and measurement, these topics were split into three more approachable categories; pupil-related error, examiner-related error and test-related error.

Pupil-related error tends to fall under the measurement error umbrella, while test-related error could be either procedural or measurement related. Examiner-related error tends to bridge the two, with more obvious human mistakes such as miscalculating marks being procedural, and more subtle mistakes like misunderstanding what the candidate has written being measurement.

Pupil-related error

Pupil-related error such as being ill on the day of the exam, arriving late and missing the start of the exam or revising the wrong things are not spontaneously thought of as significant error by most people. All of the groups feel that being unwell or missing the bus is simply bad luck, while revising the wrong things, or focusing too much on one topic which then does not come up on the exam, is deemed irresponsible or unlucky. Several groups spontaneously describe error in general as something that happens that could have been rectified, and they do not consider examples such as these to fit that description.

Teachers and examiners feel this particularly strongly, and are harder on students than other participants.

“If someone is sick and they didn’t get the grade they were capable of getting, I would simply call it under-performance on the day.”

Teacher, Male, Birmingham

“I don’t consider a student preparing for an area that isn’t on there as an error ... it’s bad luck on their part.”

Examiner, Female, London

Parents and students, the two groups who might be expected to be more lenient about pupil-related error are just as pragmatic on the issue. Parents are logical and matter-of-fact, and feel that although there should be some consideration for the pressures of exams, being well prepared and keeping calm in a stressful situation is just part of the process. Most admit that should their child receive a lower grade than expected, they would hold them responsible, and assume that they had not prepared properly.

“My daughter got Ds, but it didn’t make me want to question if the examiner did his job right or wrong. I just realised my daughter was a lazy so-and-so.”

Parent, Male, London

Even students themselves do not really consider these things to be significant. They feel that it is their responsibility to revise properly for an exam based on what they have been taught and agree that feeling stressed or unwell on the day is “just one of those things”. Some even feel that it is up to the individual student to take appropriate medication if they are feeling unwell before an exam.

“I don’t know how you can say a student doesn’t get the grade that they **deserve**. You need to be prepared for that hour or 45 minutes; you have something you need to do. You have worked for that whole year and that is your chance and opportunity.”

Male, Student, London

There was some agreement among employers on this point too. They agree that students need to be held accountable for how they perform on the day.

“It’s about accountability – the students need to be responsible for their behaviour.”

Male, Employer, London

Although many participants admit that illness and poor revision may well contribute to a student not getting the grade they deserve, it is clear that they do not regard them with the same concern as Ofqual might. It is the use of the word ‘error’ which causes the problem here, as it is seen as an inevitable part of the process as opposed to a mistake that could have been avoided or rectified.

Examiner-related error

For many participants, examiner-related error in the assessment process tends to be more ‘top of mind’ than pupil-related error when considering what might go wrong during the examination process. Issues such as examiners misinterpreting what the student is trying to say or making mistakes when adding up marks are mentioned spontaneously. People also consider examiner-related error to be more significant than pupil-related error, but still remain understanding about the difficulties and pressures examiners are under and readily admit that it is only natural for people to make mistakes.

Perhaps unsurprisingly, students are less accepting of examiner-related error than others and raise concerns about examiners not using their initiative when marking, especially in subjects like English. They are concerned about the rigidity of marking schemes and question whether they may fail to get a mark for simply not using a specific word, even if their answer clearly demonstrates that they understand the question. Students are unsure about how the marking process works and feel that some subjects should always be double-marked to ensure that the first examiner has graded the paper correctly and fairly.

“I think in English it is definitely down to personal opinion, because people have got different views about what is good or bad. Then you need to get two different people to mark it and get the average of the two.”

Student, Male, Birmingham

Teachers and parents agree with students that examiner-related error should not be occurring, but are more accepting that human error will inevitably occur to some degree. However, parents *do* feel differently when it is their child’s exam grades at stake. They

suggest that there should be a cap on how many papers one examiner can mark in order to minimise the risk of mistakes being made due to fatigue or boredom.

There is a general consensus that the marking of papers needs to be taken more seriously. People are quite surprised to learn that many examiners are actually teachers who mark papers in their spare time. They question whether this set-up in itself contributes to error, and argue that the level of examiner-related error in the system may be easier to control if teachers who examine take time out from teaching to be properly trained, and if there are some professional examiners, for whom examining is their main career.

“There needs to be more training of examiners. Schools should treat it as professional development and allow them time off to be trained.”

Teacher, Male, Birmingham

“There should be a more controlled environment in the marking and examiners should have it as the main focus of their day’s work.”

Parent, Male, London

Students in particular feel strongly about having committed, professional examiners. They appreciate that exam marking must be dull and “robotic”, but feel that teachers doing marking in the evenings, or even (as one participant mentioned witnessing) on the bus, shows no respect for the amount of time and effort they put into their revision.

“My recommendation is that we take this further and I feel strongly about it. I never knew that examiners take marking home to do it. Think about all the distractions you get when you do your homework at home.”

Student, Male, London

However, there is evidence that students are ill-informed about how papers are marked, as many are unaware that double marking and moderating already occur and some even worry that those whose surnames are last alphabetically may be disadvantaged due to examiners being tired by the time they reach their paper.

Examiners themselves readily admit that mistakes are made, but feel that the measures in place to reduce examiner-related error are effective on the whole. If anything, they would like to publicise the systems already in place to reduce marking mistakes, so that people know what is being done. They take a very different view to the majority on the idea of having full-time examiners, suggesting that an examiner with teaching experience brings an invaluable

understanding of the curriculum to their work. They also argue that most of them are very committed to what they do and want the students to do well. They rebuff criticisms that they do not use initiative by insisting that they must stick to their mark schemes in order to be fair.

“The majority of examiners do it because they care about the system. Most want to mark to the benefit of students; they don’t seem to be aware that we are out there with their best interests at heart. You can’t use your initiative – you have to have rigour and be fair to every student.”

Examiner, Female, London

Overall, given that the majority of mistakes directly associated with the examiner are procedural, most participants agree they are serious. There is far greater consensus on this than about the previously discussed pupil-related error. However, while they are regarded as unacceptable, most recognise that this type of error will never be fully eliminated from the system.

Test-related error

Error relating to the examining method, namely the paper, are seen by all groups as the most concerning and unacceptable type of error, because it should be preventable in all circumstances. It is important to note, however, that participants’ unprompted perceptions of test-related error relate to things such as misprints on the paper, typos or students being asked questions on one novel when they have read a different one. Ofqual also includes more complex factors in its definition of error in exam results. These include poor question design that fails to measure the correct skills or knowledge, or exams being biased towards one particular skill. However, the majority of people have not even considered that this type of error in the assessment process occurs in the first place.

Teachers take a zero-tolerance approach to mistakes on exam papers which they feel are completely avoidable, more so than any other type of error. They suggest that current quality procedures are simply not sufficient if these are a factor resulting in exam mark error. Teachers feel a sense of loyalty towards their students. They assert that if exams are designed to test students and pick up on any misunderstanding or mistakes that they make, that they should be entitled to an exam free from mistakes.

"A misprint is unacceptable. People should be checking it over and over again. There should be zero tolerance. Not even small errors are acceptable on the paper."

Teacher, Male, Birmingham

"Don't we penalise the students for making mistakes? It is completely unacceptable to have errors on the paper."

Teacher, Female, Birmingham

Examiners know more than most about the system as a whole and how mistakes are rectified when they do occur. For this reason they are less critical about exam paper mistakes than other participants. They take a more logical approach to the question of how the exam reflects the syllabus, suggesting that sometimes papers *have* to focus on only one or two areas as it is impossible to do more. They also say that the marking schemes are designed to take account of any mistakes that may be on the paper, ensuring that students are not penalised for it.

Students and the general public are particularly concerned about the breadth of the curriculum and say that although they appreciate an exam cannot test everything on the syllabus, the exam paper should place an equal emphasis on all aspects of what has been taught.

"There should be a fair probability of what should come up and it needs to be taught fairly too. They should vary what comes up ... if it doesn't come up and you have spent ages on it then it's a waste of time."

Student, Female, Birmingham

"If you work to a curriculum then the exam should reflect that and you can get the grade you deserve. If you have a year that covers ABCDEFG then the exam should cover all of it, not just one thing or another. That would help."

General Public, Male, Birmingham

The discussion around problems with exam papers leads several of the groups to question whether the exam is actually a successful method of assessment. Many participants suggest that modular assessments and more coursework represents a more accurate measure of a students' ability than how they perform in one exam on one particular day. Employers feel most strongly on this issue, arguing that exams are limited in what they can prove about how

capable a person will be in their job. On the whole, employers agree that exam marks are rarely a reflection of a person's ability and skills.

"It shows that you can memorise it, but it doesn't necessarily show the knowledge you're going to carry around with you."

Employer, Male, London

However, the two groups of employers differ on how important they consider exam grades to be. Employers in Birmingham, who primarily worked for larger firms, confirm that applicants with lower grades may not even be considered, despite agreeing that exams are not a strong measure of ability. Grades are simply used as a sifting mechanism to narrow down the number of applications they need to consider. The London group, however, representing smaller businesses, say that they don't even look at grades on a CV, because experience is far more important to them.

"You never even look at grades, full stop. A grade is not going to tell you if the candidate is good or not."

Employer, Male, London

Information and Transparency

Information and transparency

A desire for information about the exam system, particularly regarding the reasons for having several exam boards and information about where responsibilities lie, is a key theme in both Birmingham and London. Some groups, most notably parents, feel as though they currently have very little information about the exam system, while perhaps unsurprisingly the teachers and examiners feel more informed. Parents express a desire to be provided with more information about the examination system. For example, they want to know whether exam boards are meeting targets and which boards their child's school will be using.

The examiners, as those closest to the system, are wary of information about error being widely disseminated. While this is something that many people say they would like to know more about, examiners feel the issue should be kept between themselves and Ofqual, as they are the groups that are accountable.

It is very common, across a wide variety of subject areas, for the general public to say that they desire more information on any given topic. They tend not to be specific about the additional information they require and in practice they do not always utilise the information they receive. Having information and acting on it are two entirely separate things. It is important therefore to attempt to distinguish between legitimate areas of confusion about the exam system that require clarification and those which are merely a general, placating request for more information.

The case for transparency

Parents, perhaps more than any other group, highlight a lack of confidence in the exam system, stemming from limited knowledge and what they perceive to be little or no transparency. It should be noted, however, that this lack of confidence emerged after they were questioned on the topic, and it is unlikely that is an issue of great concern to them in their everyday lives. Their lack of confidence is not because they are aware of or even worried about error in the examination system. To the contrary, it is because they lack knowledge and are uncertain about how the system works. There is concern that information about error in the system is being kept quiet and even guilt among parents that they haven't given it enough thought before.

"You shouldn't cover up the errors."

Parent, Female, London

"I feel guilty that I've under-reacted to error. But it is one of those things in life ... it happens."

Parent, Female, Birmingham

Parents also express a desire for greater involvement in the system, with some parents suggesting that schools should run examination open days where they could receive examples of well-answered and poorly-answered exam papers, as well as information about the standards required of their children. However, their suggestions were made on the spot and were thus probably not well thought-through - it is questionable how many parents would use this facility and what they would gain from it if they did attend (many schools already *do* offer sessions of this nature, which are often poorly attended). Other parents feel that if they were more involved in the process they would be able to identify an error in their child's exam grade more readily themselves, thus taking some of the responsibility for identifying error away from the school or exam boards. A more practical idea that parents consistently raise is whether the examination system could be audited.

"I'd want to know there was a system in place open to audit."

Parent, Female, Birmingham

General public participants are undecided as to whether or not full transparency about error is a positive thing. Some indicate that it is important to know about the issue, both in terms of being aware of the level of error in the system and understanding how and why it occurs, while others feel that an open debate about the issue would not be constructive.

"[It's] important to hear the truth."

General Public, Male, Birmingham

"If you really knew [about exam error], if you opened everything up to scrutiny then you wouldn't be able to sleep at night ... some boxes shouldn't be left open."

General Public, Male, Birmingham

Students in Birmingham have a pragmatic view of how much transparency there should be around the issue of error. While they agree that it is important to have the information in the public domain, they feel it should be done in a considered manner.

"I think you should be aware of it but I don't think it should be over publicised."

Student, Female, Birmingham

Of particular importance to students, however, is information about how their papers are marked. There is a sense that once exam papers are sent away for marking, they lose any knowledge of the process or their individual performance. They obviously receive a grade, but would like to know about the process that leads to that particular grade being awarded. Some say they would like to know specific details about their performance in terms of what they got right and what they got wrong, even if the exams are their last and they will not be sitting similar ones again.

“I would care [about getting back a paper]. When I do mocks, I look back and know what I did wrong. They are guidelines, and it’s nice to know where you went wrong and where you went right for personal reasons.”

Student, Male, London

As mentioned previously, most examiners feel that Ofqual should not discuss the issue of error in the public domain.

“What would be the purpose in that?”

Examiner, Female, London

They take the perhaps patronising view that the public cannot affect the system and make a difference, and therefore they shouldn’t be part of the decision-making process; any publicity on error could, they feel, cause panic. A few examiners disagree with this, feeling that the public absolutely do have the right to know about error, but think that due to the importance and potentially emotive nature of the subject, if it is to be publicised it needs to be done in a very careful way.

“[Error is] such a big subject, it stirs up unnecessary disquiet ... it needs to be very carefully handled.”

Examiner, Female, London

Teachers tend to take a similar view to the examiners on the subject and are uneasy with the prospect of error being made a public issue. They identify the issue of scaremongering that already surrounds exams, singling out the media as culpable for this. Teachers feel that the issue should be kept ‘in house’ among educationalists and that if it is discussed with students and others, it needs to be dealt with pragmatically.

"There's a certain amount of paranoia in the teaching profession, but just because you're paranoid it doesn't mean they [the media] are not out to get you. When the GCSE results appear every August, the Daily Mail runs a story saying standards are slipping."

Teacher, Male, London

Appeals and special consideration

While discussing information and transparency, the issue of appeals and special consideration in marking is consistently raised. Teachers and examiners raise concerns about the fairness of the appeals and special consideration process, while the other groups express a general lack of knowledge about how these systems work.

There is a feeling among examiners that the appeals process is driven by parents and that as such, there is an in-built inequality as some parents will inevitably be pushier than others. A lack of knowledge about how the system works may disadvantage one parent, and in turn a student, compared to another. While some examiners are aware that marks can go down as well as up following an appeal, others get the impression that grade changes are applied too liberally, and that marks are automatically moved up when an appeal is made.

"At our school we sent papers back for remarks ... out of a sample of twenty, three-quarters went up one or two grades."

Examiner, Female, Birmingham

Examiners also raise concerns about the fairness of the special consideration system which they feel is not always applied correctly. If a student suffers a bereavement during the exam period, examiners say they should receive special consideration on *all* the exams they sit, not just one or two of them, as is sometimes the case. They also state that there need to be rules in place to be sure that special consideration is used appropriately.

"We had a candidate ... whose dad died ... and she's not looking for special consideration. And then you have another one who wanted special consideration because her hamster is ill. There is an awful lot of abuse of the special consideration system."

Examiner, Female, Birmingham

Teachers disagree with examiners that the appeals system is driven by parents, stating that it is largely the schools that push for appeals. They do concede, however, that different

schools have different attitudes and this often leads to an uneven application of the remarking and special consideration system.

The views of both parents and the general public confirm the suspicions of teachers, that is, that is that they know very little about the appeals system and how it works. Once explained to them, they both say they would like more information about it and some are surprised that such a system even exists, saying that when they were younger nobody would ever have questioned their grades.

“Certainly in our day we never sent a paper back to be remarked ... you would just never know about error.”

General Public, Male, Birmingham

Students themselves know more about the appeals system than their parents, including the fact that there is a fee to have a paper returned. This is largely because most know someone who has appealed against a grade or have done so themselves. They also know about the special considerations system and agree that it is hard to monitor who deserves it and who does not.

“People might try to take advantage of it.”

Student, Female, Birmingham

Consistency and continuity

Consistency and continuity

Consistency and continuity, and how a lack of these might be contributing to error in the system, is a key concern of participants. Most people are unaware of how the exams system operates and who does what, and upon having this explained, question why there are so many different bodies involved in the system and several examination boards duplicating the same role.

Parents, students and the general public feel very similarly on this issue. All raise concerns about the number of examination boards and question why, for example, every child in England doesn't sit the same maths paper? There is bemusement over why one body sets the curriculum, only for several separate bodies to then design their own questions based on that curriculum. No wonder, they say, there is room for error to creep in.

"Wouldn't it make sense for the same people who set the curriculum to then set the papers? There should just be one authority."

Parent, Female, London

"There should be only one exam board, so it is fair for all the colleges."

Student, Male, Birmingham

The general public take the idea of consistency one step further by suggesting that the environment in which the papers are marked should be monitored and kept consistent too. In the extreme, some members of the general public think that examiners should all sit down and mark papers in a controlled environment, just as the students did when they took the exam.

"Improvements need to be made to the system; the structure and pattern. They should keep it consistent. If examiners have to mark at a set time, they should plan for it."

General Public, Male, London

Teachers' proximity to the exam system means that they are more aware of the reasons for having different exam boards than most other participants. They discuss the competitive nature of the organisations and the reasons why schools will choose one board over another. Despite this, they agree with the parents, students and general public that the design of the system lends itself to errors and feel that it would be clearer if everything was controlled by one body – the QCA.

“Why are there multiple exam boards? Why doesn’t QCA just control everything? Having multiple boards creates an element of competition as schools want to do well in the league tables.”

Teacher, Male, Birmingham

Examiners delve deeper into the rivalry between the different exam boards, suggesting that they need to be rigorously checked to ensure there is consistency across the board. There is a consensus among examiners that schools will choose the board they perceive will give out higher marks to ensure they do well in the league tables, and that the boards play up to this by adapting their products accordingly.

“The business side of these exam boards should be questioned.”

Examiner, Female, Birmingham

Employers also know about the different exam boards, but do not recognise any benefit in having more than one. They think about things in terms of their roles in recruitment, in the sense that employers will not know what a grade from a specific board means compared to a grade from a different board.

“You cannot say, ‘This person got an A through OCR and this person got an A through Edexcel’ because that doesn’t mean anything to an employer.”

Employer, Male, London

Awareness and opinions of Ofqual

There is very little awareness of how Ofqual fits into the current picture, which is perhaps unsurprising given that it is less than a year since Ofqual was formed. With the exception of examiners, very few participants have even heard of Ofqual and therefore feel unable to pass comment. Those examiners who are aware of Ofqual are unclear as to how Ofqual is separate from QCA.

When given details about Ofqual’s mission statement, the general public, employers, students and parents react positively. The idea of Ofqual acting as an independent regulator instils confidence and they hope that Ofqual may be able to provide the consistency the system currently lacks. They do feel, however, that Ofqual will need to prove it is acting on its plans.

“The mission statement sounds spot on, but it needs to be set in stone.”

Parent, Male, London

Teachers and examiners are somewhat more critical and find Ofqual’s aims and objectives a little vague and in need of fleshing out. They have their own ideas about what Ofqual should be doing, most notably having a role in monitoring the exam boards to ensure consistency. Examiners in particular want Ofqual to ensure that grades from different boards mean the same thing and that a grade is worth the same now as it will be worth in the years to come (and in the years before).

There are various suggestions as to what Ofqual should be doing to ensure that error is reduced to a minimum. These include having sight of exam papers prior to their release for checking, setting up a framework of good practice to put exam boards into competition with each other with regard to standards and making checks on the chief examiner for each board to ensure they are being rigorous and fair.

“Ofqual should be an umbrella organisation who act as an independent regulator to oversee a single exam board.”

Teacher, Male, Birmingham

“Ofqual need to have the power to slap the hands of exam boards when things go wrong, and make people accountable. They should also be keeping records of any error.”

Examiner, Male, London

Talking about exam error

Talking about error

The original aim of this research was to investigate the views of various sub-groups on the topic of “exam error” (i.e. error in exam results and errors in the assessment process). Although some interesting findings on their opinions emerged, the study also provided a previously unforeseen opportunity to consider *how* discussions about the concept of exam error should be handled.

Prior to this study, no attempts had been made by Ofqual to talk to non-experts about the issue of exam error, and some key messages have emerged from work described here regarding the best way of approaching this topic in the future.

The experience of Ipsos MORI and Ofqual on this project suggests that even a slight variation or imprecision in the terminology used in discussions with the general public can alter the way people think and form their opinions on this topic. The way in which the piece of research was framed (i.e. in describing it as being about “exam error” as opposed to “factors which might cause a student to get a grade other than the one they deserved”) did not account for the subtleties or complexities involved. As a result, how participants were set up to think about the research topic at the recruitment stage, through to how we designed the research materials, all affected the way in which respondents thought about error, and this was not necessarily in the way intended.

For example, we sometimes referred to pupil illness, or marking inconsistencies between examiners as ‘errors’ and sometimes as ‘causing an error in the mark/grade awarded’ (in the sense of the student not getting the grade they deserved). Participants rejected the notion of pupil illness being an error, framing this as ‘bad luck’, ‘tough’ or ‘just life’. However, they took a softer line on the idea of pupil illness as a cause of error in the grade received. None felt it was as significant as other causes of error in the grades (the inevitable versus avoidable demarcation was key here), but they were able then to consider what, if anything, should be done about it.

This is not to say that respondents were led down the wrong path, as it is clear from their responses that they had grasped the issue at hand. However, it may mean that their views and opinions were not as focused as they might have been, had a stricter use of terminology been applied.

It should be noted that participants in both London and Birmingham commented spontaneously on the complexity of the topic, which reinforces the importance of precision when handling the topic in the future.

Fundamentally, the word 'error', in the technical sense it is used by Ofqual, is one which the general public does not use or recognise.

For Ofqual, it can have a relatively benign "no blame attached" connotation. For the general public, error means mistake, and if a mistake has been made, someone is to blame/can be held responsible. As such, the public have a much narrower definition of error than Ofqual. Effective communication between Ofqual and the public on error in the future will be contingent then, on finding a more 'neutral' and in-common way of describing the issues and contributing factors.

Appendices

Discussion guide

Ofqual Error Research Programme		
Final Discussion Guide (19.01.2009)		
Activity	Indicative timing	Aims and outcomes
Arrival/ Refreshments Registration and refreshments.	18.00 for 18.30	
Plenary Session 1 Introduction A few words from Ipsos MORI's Chairperson: "What we're here to talk about and why it's important".	18.30 – 18.40	<i>Set the agenda for the evening.</i> <i>Explain the purpose of the event.</i>
Discussion Group 1 Moderator introduction: Confidentiality/ MRS Permission to record? Ground rules: Respect others opinions, disagreement is encouraged, speak one at a time – polite and recording, lots to covers – moderator may interrupt and move on. Warm-up exercise – paired introductions <i>I want us to start this evening by thinking about a common situation. Imagine that a person goes to their doctor, experiencing an issue with their health. I would like you to think about the things that might lead to that person not receiving the right treatment for them.</i> WORK THROUGH EXAMPLE: Misdiagnosis in the medical field. NOTE TO MODERATORS: Example is split into 3 key sections, with possible issues listed in order of severity. Please start at the top of the list and work down – never jump straight in with one of the more severe possibilities.	18.40 – 19.00	<i>Unclouded, uninfluenced, top of mind reactions to the issue of error.</i>

FLIPCHART

What could go wrong when a medical diagnosis is made? **PROMPT** Think about the patient? And the doctor? What about the instruments the doctor uses?

Pull out issues that relate to the different contributions of:

THE PATIENT:

- What if they describe a lot of co-incidental symptoms unrelated to the real problem?
- What if they do not describe their symptoms accurately enough?
- What if their symptoms are not present on the day of their appointment?
- What if their symptoms are not severe enough to make the doctor aware of how unwell they really are?

THE DOCTOR:

- What if they are pre-occupied with the previous patient, or under pressure to finish their clinic and fail to diagnose the problem?
- What if they are not up-to-date in their training on this particular condition?
- What if they simply misinterpret the symptoms in front of them and make an incorrect judgement?

THE INSTRUMENTS

- What if the sample of blood or cells taken for diagnosis becomes corrupted?
- What if the thermometer is broken or the sensitivity of the heart monitor is set incorrectly, even if the doctor is using the instruments correctly?
- What if the instruments themselves are poorly designed, or broken?

In each case;

Who is responsible for the problem?

What are the implications of the problems?

<p>Discussion Group 2</p> <p>Top of Mind Reactions to Exam Error</p> <p><i>Moving along, I now want us to think more specifically about exams and the examination system. As with the medical example, I would like you to think about the things that may go wrong in exams and the examination system more generally.</i></p> <p>FLIPCHART</p> <p>PROMPT Think about the student. What about the examiner? And what about the exam paper?</p> <p>NOTE TO MODERATORS: Again, work through the 3 key sections – you do NOT need to order them from less severe to more severe: get your group to decide on which issues they feel are the most serious.</p> <p>Pull out issues that relate to:</p> <p>THE STUDENT</p> <ul style="list-style-type: none"> – What if the student has revised, but on the day of the exam they fall apart, and are unable to get their act together? – What if they are ill on exam day, and are unable to perform to the best of their ability? – What if they have revised, but have focused their revision on topics that do not appear on the paper? <p>THE EXAMINER</p> <ul style="list-style-type: none"> – What if they fail to understand what the student is trying to express in their answer? – What if they miss out a question when they are marking the paper? – What if they miss out a whole page or several pages when they are marking the paper? <p>THE EXAM PAPER/ METHOD OF EXAMINATION</p> <ul style="list-style-type: none"> – What if the questions are poorly worded and students interpret them in different ways? – What if the questions are not pitched at the appropriate level? 	<p>19.00 – 19.20</p>	<p><i>Views on the main topic – examinations</i></p>
---	----------------------	--

<p>– What if the paper is simply not representative of the syllabus (i.e. too biased towards one topic, or skills set.)</p> <p>In each case:</p> <p>Who is responsible?</p> <p>Would you consider it to be an ‘error’?</p> <p>Is it an acceptable mistake/ error?</p>		
<p>Plenary Session 2</p> <p>Introduction to Exam Error</p> <p>A few words from Ipsos MORI’s Chairperson on exam error.</p>	19.20 – 19.25	<i>Explain the wide variety of errors that exist</i>
<p>Discussion Group 3</p> <p>What are your reactions to what you’ve just heard? Do you agree or disagree with that explanation? PROBE Why do you say that?</p> <p>What other things could be considered an error within this explanation?</p> <p>Do you agree that these things are errors? PROBE Which do you agree or disagree with the most?</p> <p>What do you think are the implications of all these different possible errors occurring?</p> <p>What do you think that exam results mean/ enable/ achieve? Has this changed given what you’ve heard?</p> <p>Given this description of error, how much error do you think currently exists in the examination system? PROBE Is that an acceptable level? PROMPT Why? Why not?</p> <p>PROBE What could or should be done</p> <p>PROBE What if an error affected your child/ friend/ relative?</p> <p>Compile a very short (5 minutes maximum) presentation for the other stakeholders.</p> <p>Include:</p> <p>What you consider error to be (briefly!)?</p>	19.25 – 19.55	<i>Discussion about possible breadth of exam error and reaction to that.</i>

<p>Why exams are important to you as a group?</p> <p>What do you think exams achieve? What is there value?</p> <p>What level of error there is and what is acceptable?</p>		<p><i>Key feelings of this stakeholder group to be presented to other stakeholder groups.</i></p>
<p>Coffee break</p> <p>Encourage participants to bring refreshments back to tables for next session.</p>	<p>19.55 – 20.10</p>	
<p>Plenary Session 3</p> <p>Stakeholder presentations, aided by the moderators if necessary. 5 minutes each maximum</p> <p>Moderators to take notes on other stakeholders main points for later discussion</p>	<p>20.10 – 20.40</p>	<p><i>To increase awareness of other stakeholders group's opinions</i></p>
<p>Discussion Group 4</p> <p>What do you think about the other people's viewpoints?</p> <p>What do you think about Examiners</p> <p>PROMPT What do you think about their concerns? What are the potential consequences/ implications on this group?</p> <p>What do you think about Teachers</p> <p>PROMPT What do you think about their concerns? What are the potential consequences/ implications on this group?</p> <p>What do you think about Students</p> <p>PROMPT What do you think about their concerns? What are the potential consequences/ implications on this group?</p> <p>What do you think about Parents</p> <p>PROMPT What do you think about their concerns?</p>	<p>20.40 – 21.00</p>	<p><i>Revisiting the issues in light of the views of other stakeholders</i></p>

<p>What are the potential consequences/ implications on this group?</p> <p>What do you think about Employers</p> <p>PROMPT What do you think about their concerns? What are the potential consequences/ implications on this group?</p> <p>What do you think about the General Public</p> <p>PROMPT What do you think about their concerns? What are the potential consequences/ implications on this group?</p> <p>Moving on, I want to talk about the role of Ofqual in this issue. PROMPT (if necessary) Ofqual are the new regulator of qualifications, exams and tests in England. They claim to be ensuring that children, young people and adult learners get the results their work deserves, that standards are maintained and the qualifications learners receive count now and in the future.</p> <p>What do you know about Ofqual? PROBE What are their responsibilities? Or what should they be?</p> <p>Does their statement (above) match what you'd expect from them? PROBE What else should they be doing?</p>		
<p>Discussion Group 5</p> <p>Please use this session to explore any areas of interest that have arisen so far.</p> <p>You may like to think about the following questions:</p> <p>Should this issue be discussed/ addressed? What are the implications?</p> <p>Is greater publicity of this issue a good thing? Is there potential that this could damage trust in the examination system? Is that ok?</p> <p>If there is error, is it not better to be open about it?</p> <p>How can this situation be improved? Should we ignore/ engage?</p> <p>What changes would you propose introducing in to the system?</p> <p>What are the implications for the different groups of people here tonight? Are the consequences greater for different groups? And within those groups (GCSE, A Level, Professional Qualifications)?</p>	<p>21.00 – 21.20</p>	<p><i>Wider issues discussed and explored.</i></p>

<p>Plenary Session 4</p> <p>Close</p> <p>A few words of thanks from Ipsos MORI's Chairperson.</p>	<p>21.20 – 21-25</p>	<p><i>Round off the evening. Thank people for giving up their time and the effort they have given.</i></p>
<p>Discussion Group 6</p> <p>Admin and incentives.</p>	<p>21.25 – 21.30</p>	<p><i>To ensure that this is done before when respondents are expecting to go home. Should help hold their interest in the penultimate session.</i></p>