



Ofqual

Summer
Series
Symposium
2016

START 9:00
END
3hr 12:10
2hr 11:10
1hr 10:35

Sally Collier

Chief Regulator

Summer awarding

Cath Jadhav

Marking

- Assigning marks to a student's work according to the mark scheme

Both writers use emotive language to appeal to the readers sense of pity and also making them laugh, using anecdotes to emphasize their opinions and descriptions and sometimes using short, punchy sentences to impress a point in the readers' minds like 'Only the roads are safe' and 'And they loved it', making the reader really feel as if they were with the characters of the extracts.

Interprets the effects of language. Full and detailed.
Compares. Perceptive comments. Band 4 14/16

Grading

- Setting grade boundaries once marking is (mostly) completed

	A*	A	B	C	D	E	F	G
1380_1F				74	59	45	31	17
1380_2F				75	62	49	36	23
1380_3H	85	68	51	34	20	13		
1380_4H	91	73	54	36	20	12		

	A*	A	B	C	D	E	F	G
1380F				149	121	94	67	40
1380H	173	141	105	70	40	25		

Comparable outcomes (1)

- What is it?



It's all about **grading**

Rule of thumb used by exam boards for decades:

If the cohort is similar (**comparable**) to last year's cohort, then you'd expect results (**outcomes**) to be comparable to last year's.

Comparable outcomes (2)

■ Why do we do it?

- Performance lower in the first year(s) of a new qualification
- Unfair to penalise students sitting new qualifications

And...

- Examiner judgement is not sufficiently precise
- Aligns inter-board standards in a subject

More information on inter-board comparability:

<https://www.gov.uk/government/publications/ofqual-2015-summer-exam-series-report>

Entries – Year 11 students in England

■ Increasing

- Additional science +11%
- Biology +5%
- Chemistry +6%
- Computing +87%
- English literature +2%
- Geography +8%
- History +7%
- Physics +5%
- Science +22%
- Spanish +3%

■ Decreasing

- Citizenship -26%
- Design & technology -10%
- French -8%
- German -8%
- ICT -28%
- Music -5%
- Performing/expressive arts -11%
- Religious studies -6%

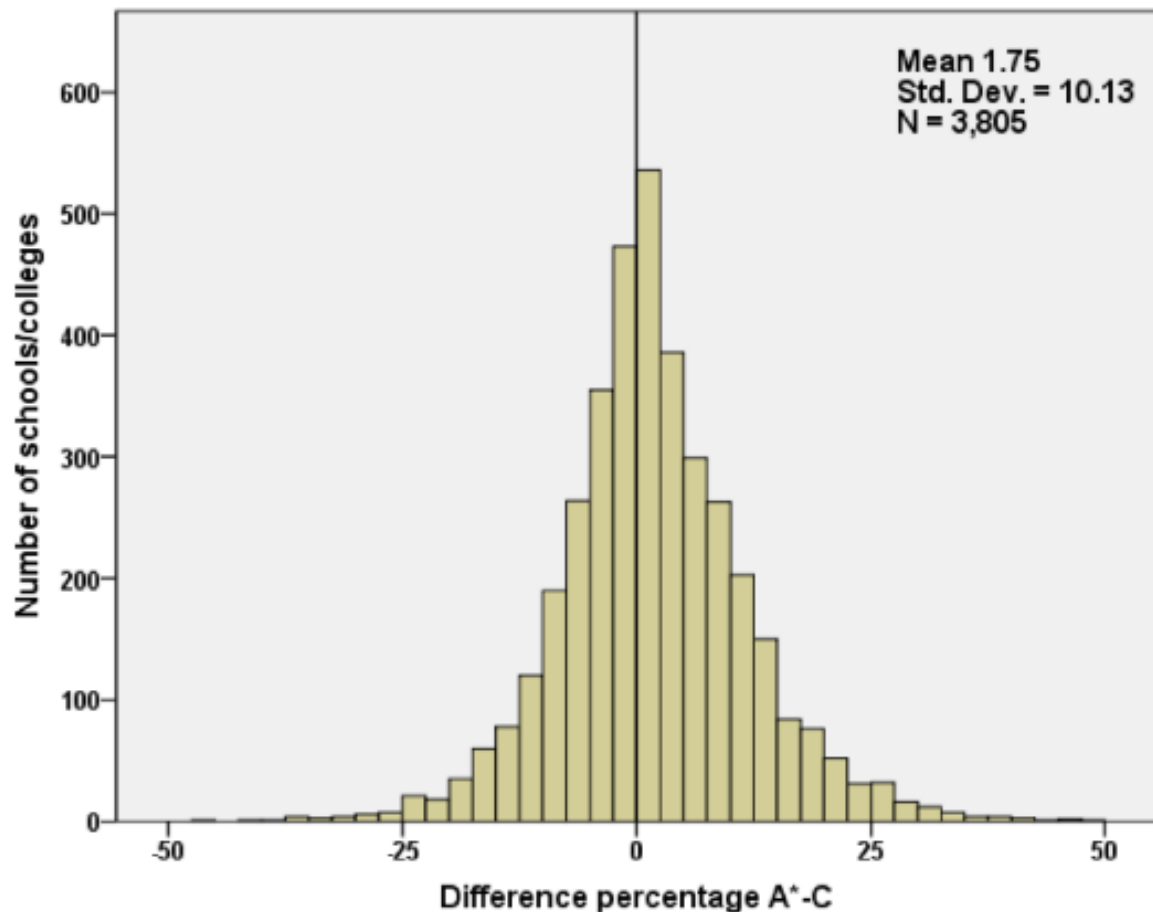
Source: <https://www.gov.uk/government/statistics/summer-2016-exam-entries-gcses-level-1-2-certificates-as-and-a-levels-in-england>

New AS – first awards 2016

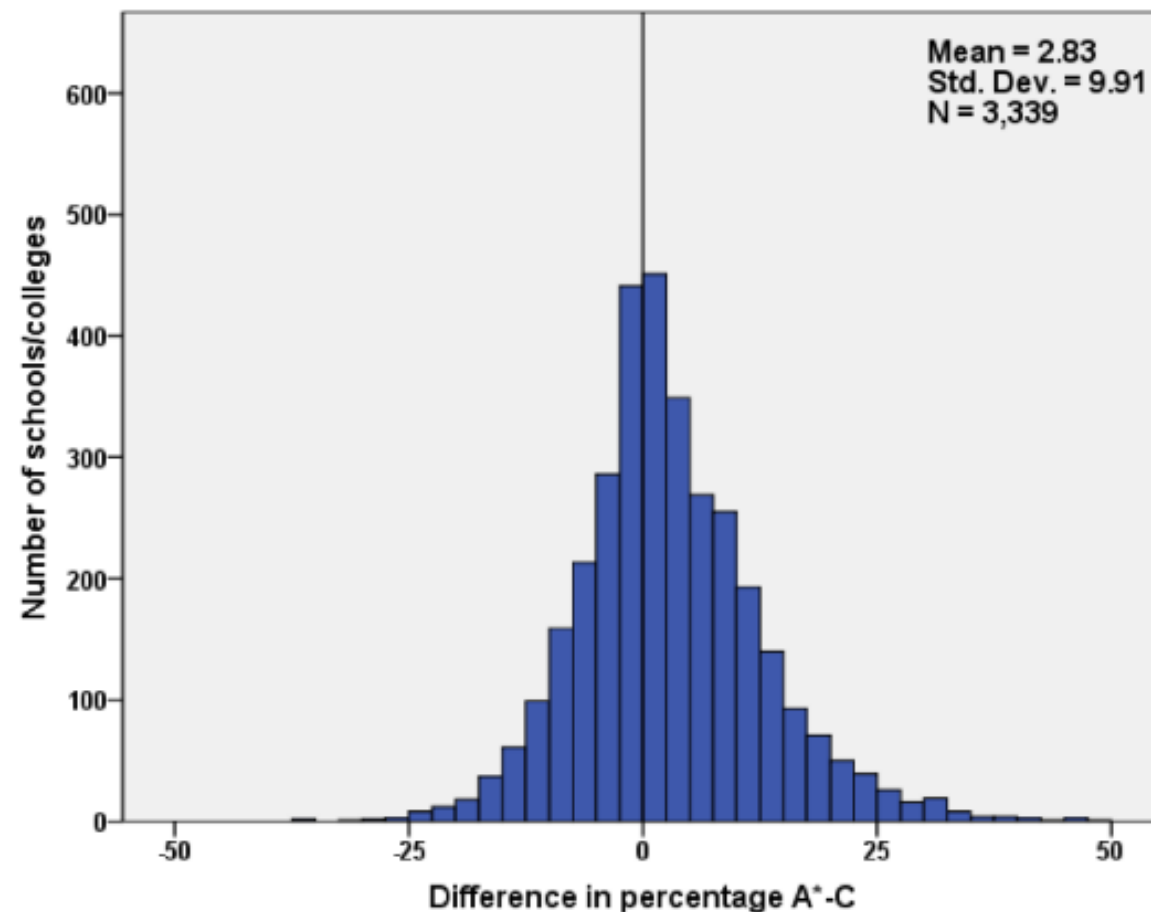
- Art & Design
- Business
- Biology
- Chemistry
- Computing
- Economics
- English language
- English language & literature
- English literature
- History
- Physics
- Psychology
- Sociology

Year-on-year variation

GCSE mathematics summer 2015 vs 2014: all students



GCSE mathematics summer 2015 vs 2014: Yr 11 students only



See: <https://www.gov.uk/government/publications/variability-in-gcse-results-2012-to-2015> and <https://www.gov.uk/government/publications/variability-in-a-level-results-2012-to-2015>

MFL native speaker research

- A level
 - French
 - German
 - Spanish
 - Italian
 - Russian
- 2400 schools/colleges
- 500 returns so far



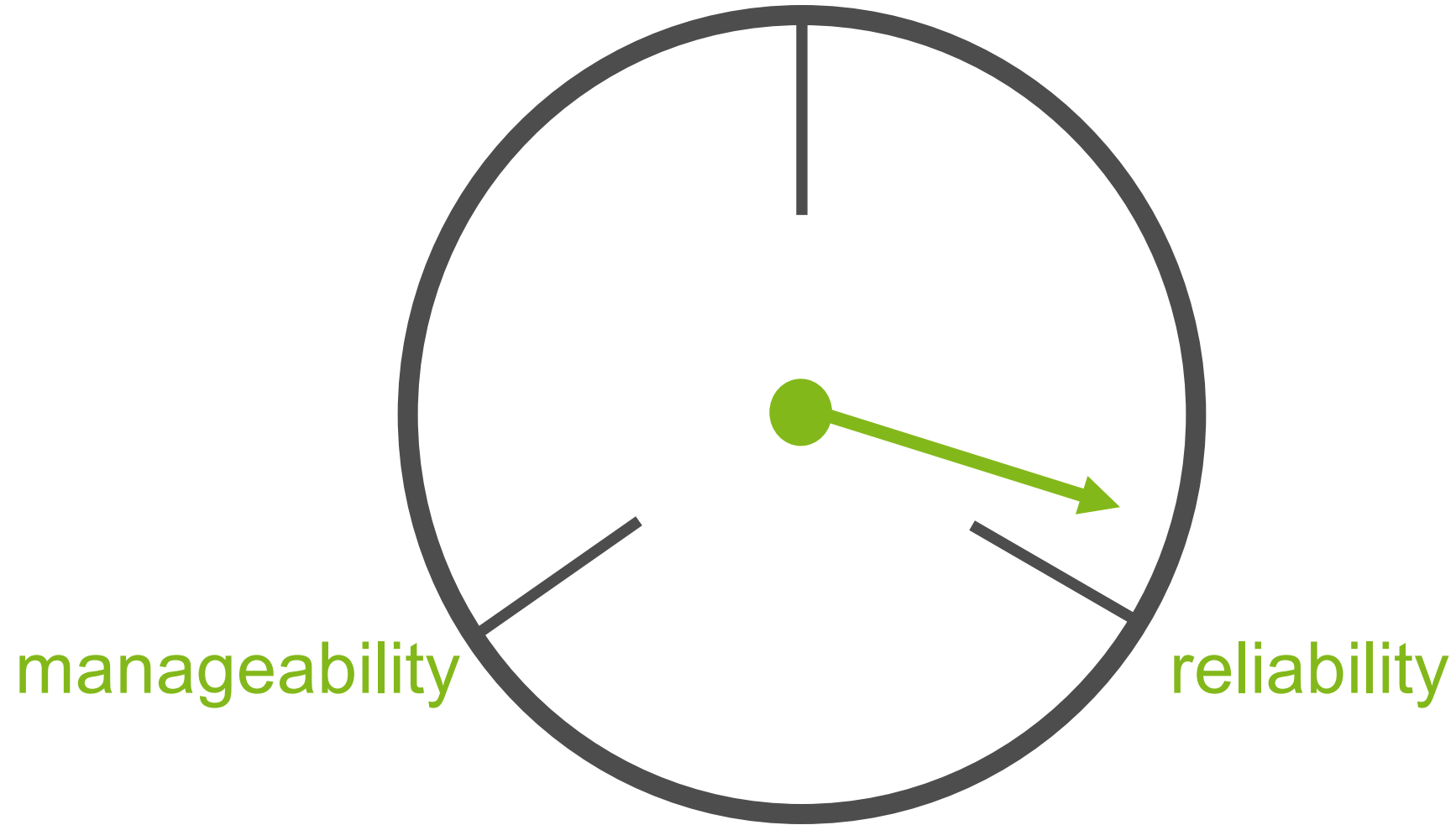
#gcse



Quality of marking

Michelle Meadows and Beth Black

Assessment is a compromise... validity



(i) What is meant by the term *homologous series*?

This is when a series of organic compounds have the same functional group and each successive member differs by CH_2

[2]

100% mark agreement

Creative writing

Write a short piece of travel writing explaining how you coped with a challenging journey

Remember to:

- write a piece of travel writing
- use language to explain

Try to write approximately one page.

AO3 i, ii	Skills
<p>Mark Band 3 'clarity' 'success' 9-10 marks</p>	<p>Communication</p> <ul style="list-style-type: none"> • shows clarity of thought and communicates with success • engages the reader with more detailed explanation • clearly communicates the purpose • writes in a register which is clearly appropriate for audience • uses linguistic features appropriate to purpose • uses vocabulary effectively including discursive markers <p>Organisation of Ideas</p> <ul style="list-style-type: none"> • uses paragraphs effectively in the whole text • begins to use a variety of structural features
<p>Mark Band 2 'some' 'attempts' 5-8 marks</p>	<p>Communication</p> <ul style="list-style-type: none"> • communicates ideas with some success • engages the reader with some explanation • shows some awareness of the purpose • attempts to write in an appropriate register for audience • uses some linguistic features appropriate to purpose • attempts to vary vocabulary and use discursive markers <p>Organisation of Ideas</p> <ul style="list-style-type: none"> • uses paragraphs which may be tabloid and/or at times correctly placed • some evidence of structural features
<p>Mark Band 1 'limited' 1-4 marks</p>	<p>Communication</p> <ul style="list-style-type: none"> • communicates with limited success • reference to one or two ideas linked to task • limited awareness of the purpose • limited awareness of appropriate register for audience • simple use of linguistic feature(s) • uses simple vocabulary <p>Organisation of Ideas</p> <ul style="list-style-type: none"> • random or no paragraphs • limited use of structural features
<p>0 marks</p>	<p>Nothing worthy of credit</p>

I went on a Journey to
sell Bank in my primary
school it was really hard
at first I couldn't cope
with it.

Purpose

Structure

firstly when I got there
we had to take our shoes
off and wear old wellington
Boats which we had to
Borrow from them they
were old, smelly, wet and
dirty. But we were told
By our teachers to so we
had to do so as we were told.

Vocab

Nearly 100% agreement
on 6 marks

secondly we were told to do
sweats ^{for} which we were
already doing in our school
and which we were bored
of we hated the subject
geography but then we just
to the teachers there and did
as we were supposed to do.

Structure

thirdly we were asked to
work in team groups
but mostly everyone was
not coping in these team
group as they were put
in a team of our which
they didn't prefer at all
and where the kids that
were in these teams weren't
getting along as they hated
each other.

Scorching heat, blistering sand and
deadly animals. The desert? The Sahara
desert! This challenging journey started
at the far east in Egypt and the
aim was to reach the far west of
Africa.

The first day was the easiest
because of the endless ^{supply} of water
and shelter in Egypt. Transport was swift
on the backs of the desert pack-horses,
the camel. The only downside was the
increasing stench of rotting meat
from the camel which sends bile racing
up your throat. You try and swallow it

Marks range between 5 &
10 marks; most on
7, 8 and 9 marks

down but as it goes it leaves a
distasteful layer of vile bile behind.

The next couple of days were
the same. The landscape never changed,
it was sand dune after sand dune like
a long mountain range. There is one
sand dune nicknamed "The Daddy" which
we first saw from Egypt but didn't
reach until the fourth day.

By the fourth day we had
to leave the camels behind as
the sand became too hot for them.
As we started walking the sand burned
through the soles of our shoes, like
butter melting in the sun, whilst
giving off a sulfuric smell which brings
tears to your eyes.

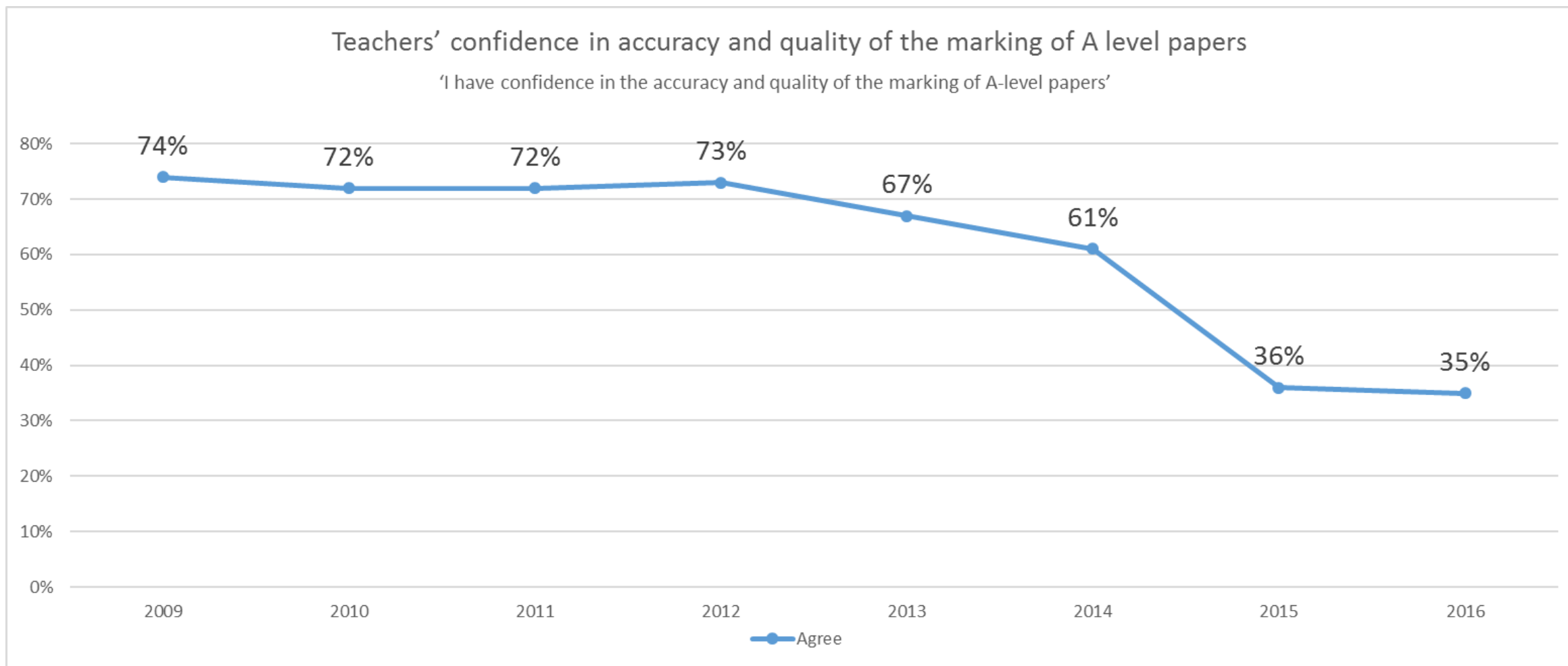
By the last day we were
pretty much dead! With sand clogging
up my gullet, stopping the oxygen getting
to my lungs; the sight of dune after
dune, venomous snakes back after back,
apparitions after apparitions were driving us

Ofqual's *Review of the Quality of Marking in General Qualifications* (2014)

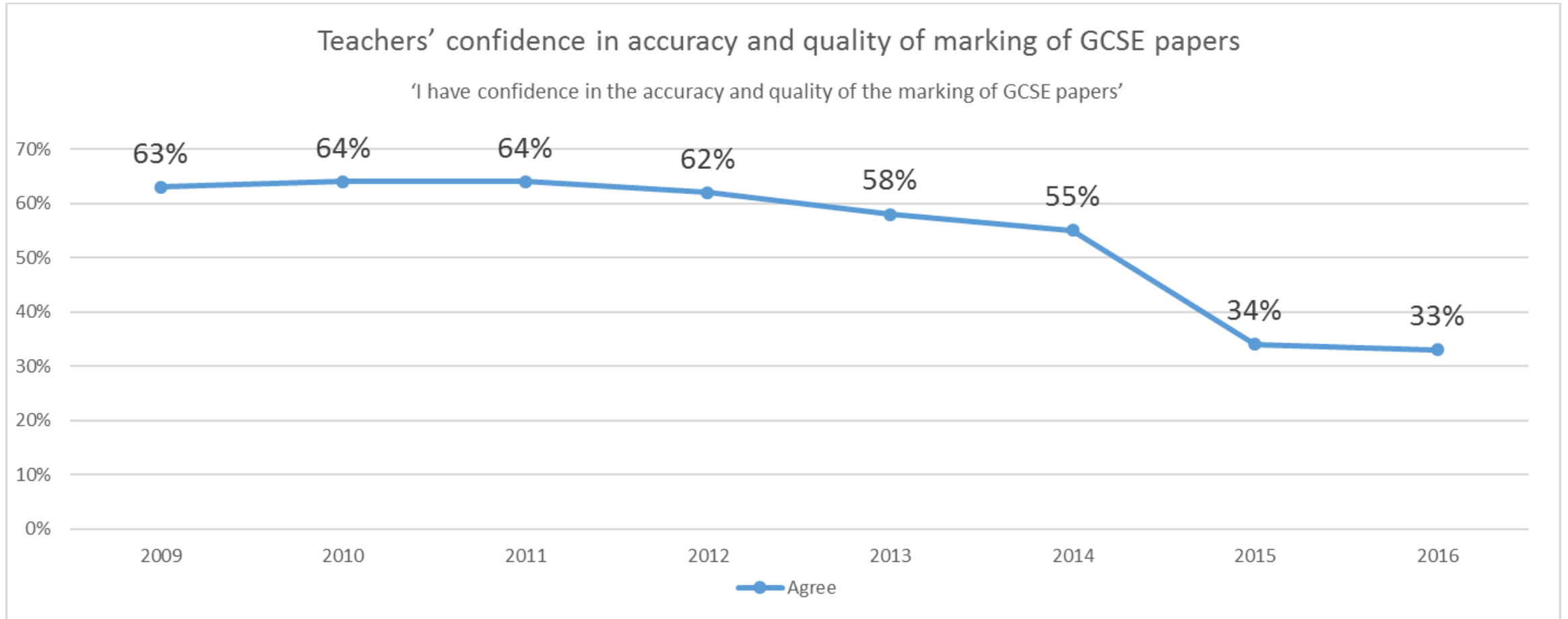
- 4 group discussions and 46 interviews
- 82 individuals surveyed: 32 teachers or head teachers who were not also examiners, and 50 examiners of whom 32 were current teachers

Source: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/393960/2014-02-14-ofqual-quality-of-marking-qualitative-research-study.pdf

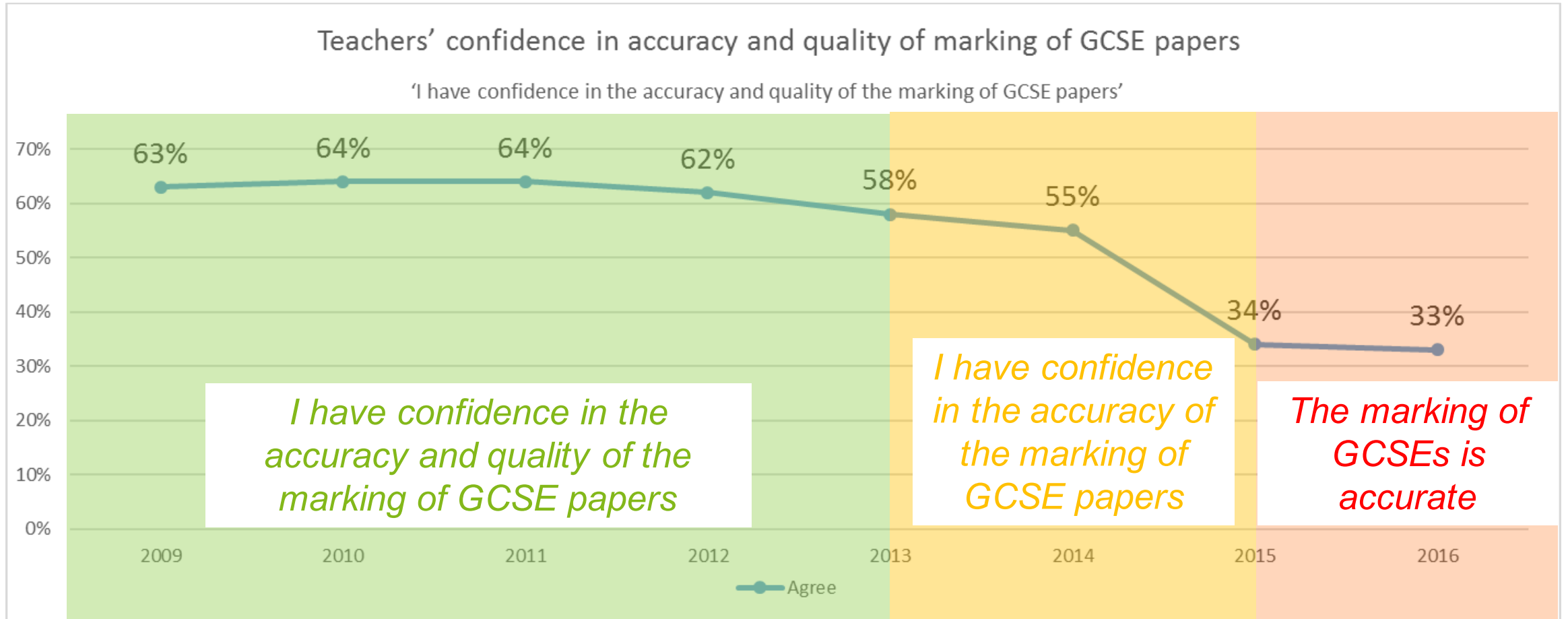
The percentage of teachers who agree they have confidence in accuracy and quality of the marking of A level papers



The percentage of teachers who agree they have confidence in accuracy and quality of the marking of GCSE papers



The percentage of teachers who agree they have confidence in accuracy and quality of the marking of GCSE papers



Complaints from schools and colleges

■ Quality of marking

2013 53

2014 75

2015 43

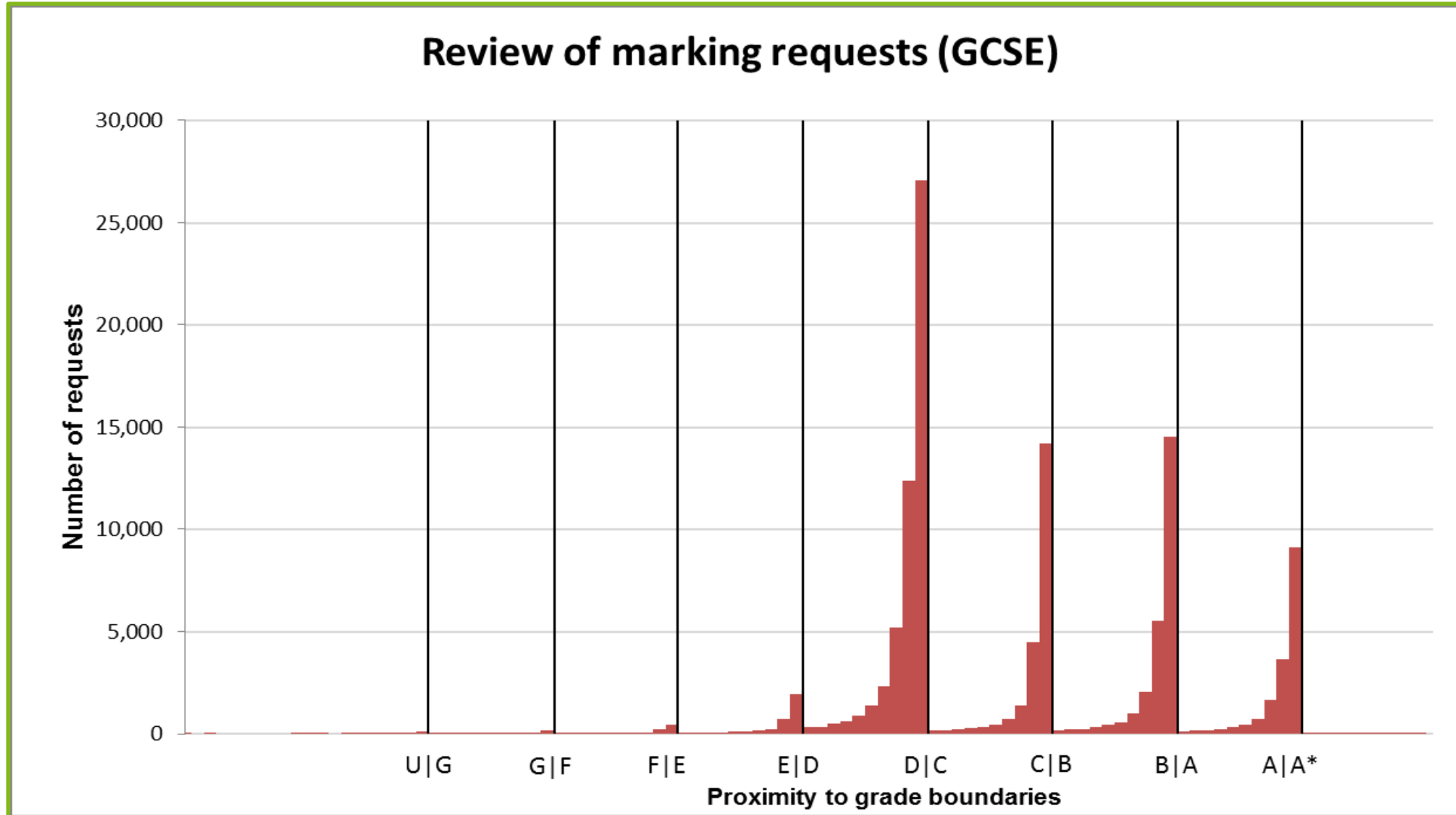
■ Enquiries and appeals process

2013 14

2014 12

2015 35

Pattern of GCSE enquiries about results



Grade changes of two grades or more

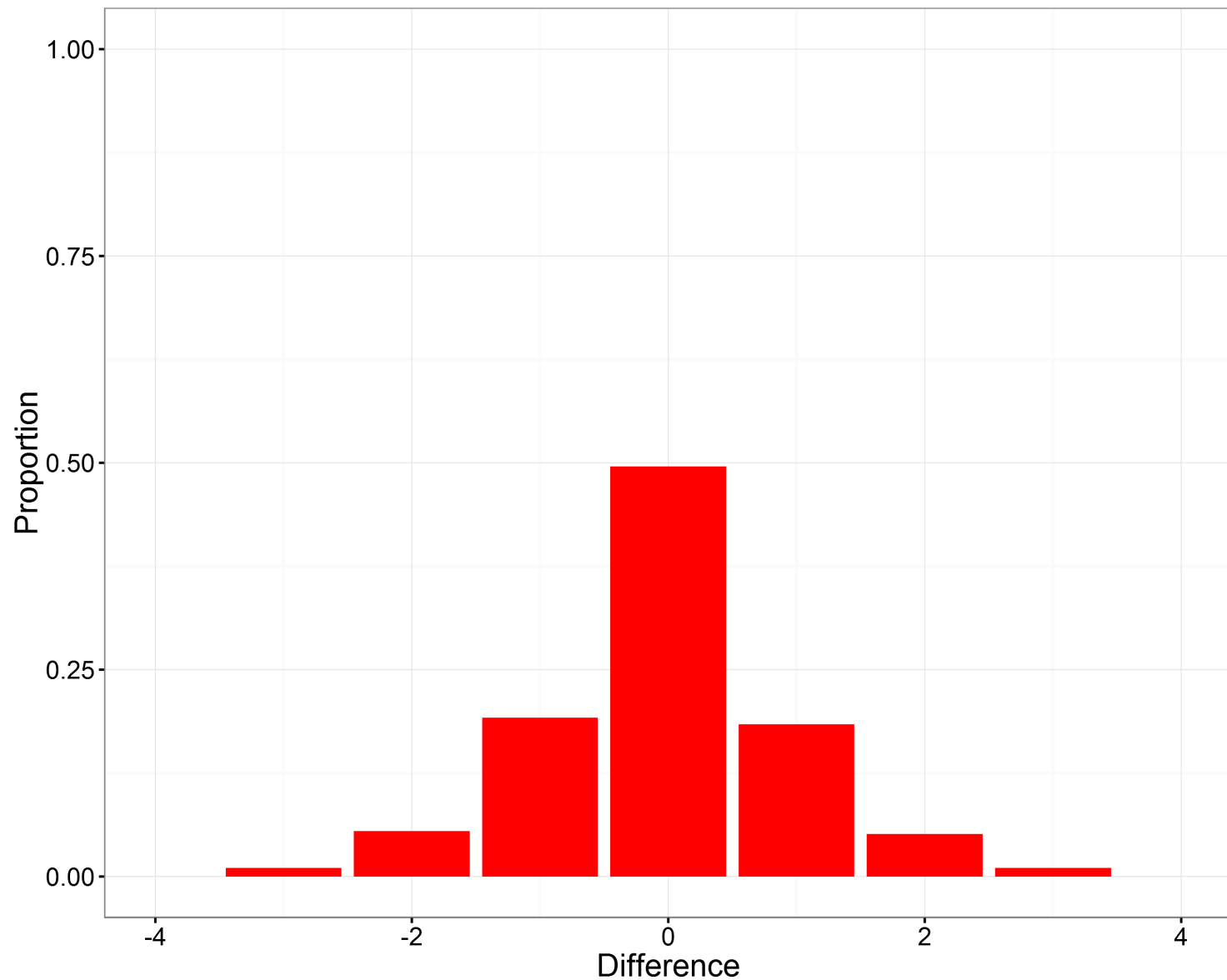
Grade change	Applied double	A level	AS level	GCSE full course	GCSE short course
Decrease two grades	0	0	5	0	1
Increase two grades	2	89	247	218	8
Increase three grades	0	3	15	39	0
Increase four grades	0	0	2	8	0
Increase five grades	0	0	0	2	0
Total	2	92	269	267	9

Consistency of marking metrics

- To calculate any quality of marking metric you need some form of double marking (2 or more views of the mark worthiness of a student's response)
- Exam boards' live online marking quality assurance processes (e.g. seeding) give us double marking data.
 - Vast evidence base (approx. 25 million marking instances)
- Caveats:
 - Metrics tell us about consistency of marking rather than *just* marking errors
 - Seeds that function well as QA might not function well to give us metrics

Geography

39,091 items



Difference	Percentage (%)
-3	1.04
-2	5.46
-1	19.15
0	49.52
1	18.35
2	5.09
3	1.04

How do we compare internationally?

- **USA** (Longford, 1996) Test of Written English (part of TOEFL) essay scored from 1-6
 - Exact agreement varied between 65-75%
- **USA** (LeMahieu et al, 1995) portfolio assessment scored from 1-6
 - Exact agreement varied between 46-57%, adjacent agreement 87-98%
- **New Zealand** (Brown et al, 2004) essay scored from 1-11
 - Adjacent agreement 66-92%
- **USA** (Penny et al, 2000) essay scored from 1-6 with augmented scoring up or down
 - Exact agreement varied between 59-63%
- **USA** (Supovitz et al, 2010) reading and writing scored from 1-9
 - Reading exact agreement varied between 55-71%
 - Writing exact agreement varied between 63-72%

How do our questions of similar maximum mark fare?

- Exact agreement from international literature for 6 mark items:
 - lowest is 46% highest 75%
- Adjacent agreement (one study)
 - lowest is 87% highest 98%



Not quite comparing like with like.

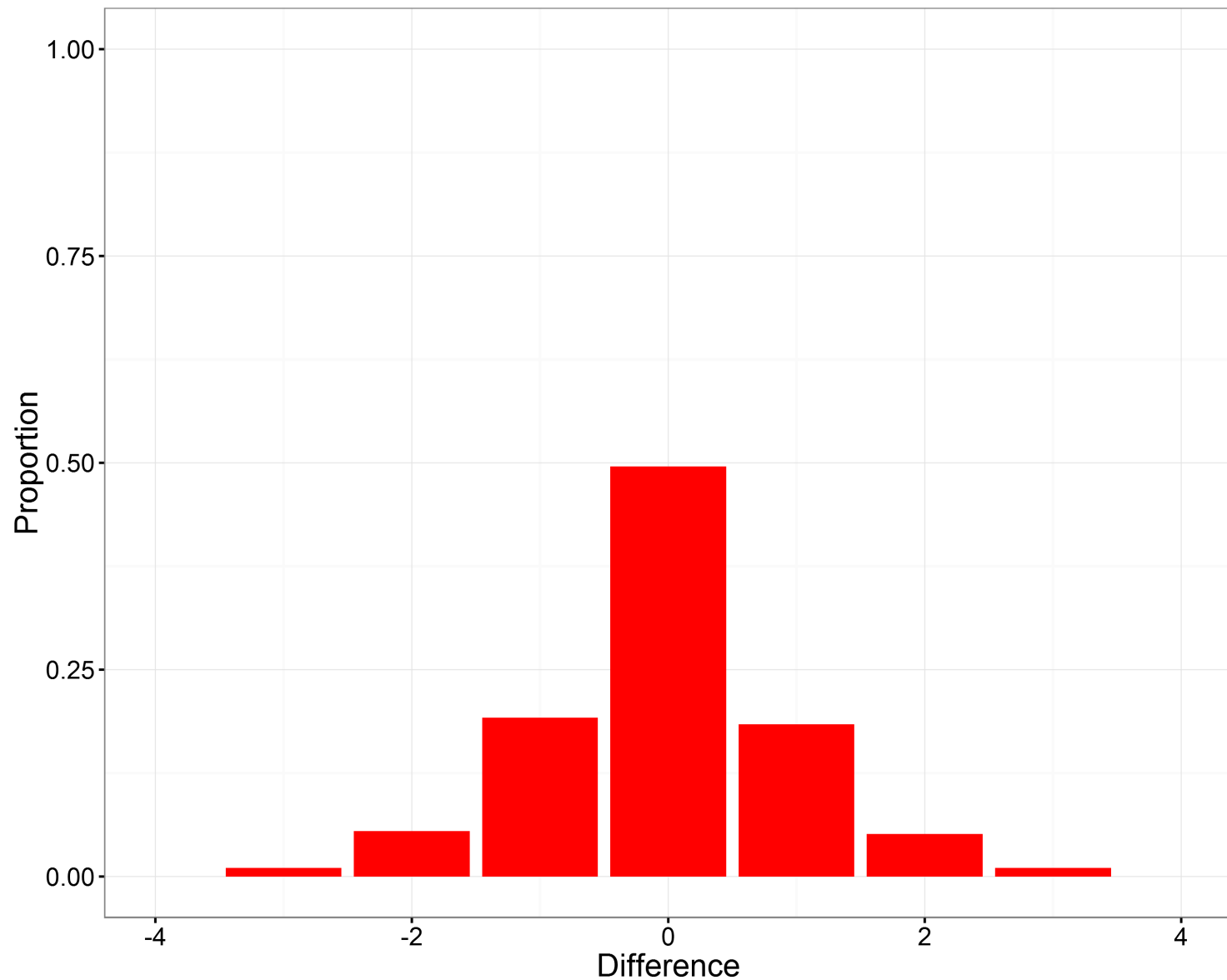
Marking study data vs. seed item data – similar...

- ... marker engagement and authenticity?
- ... types of responses selected for each purpose?

NB Neither disentangle legitimate differences of opinion from errors

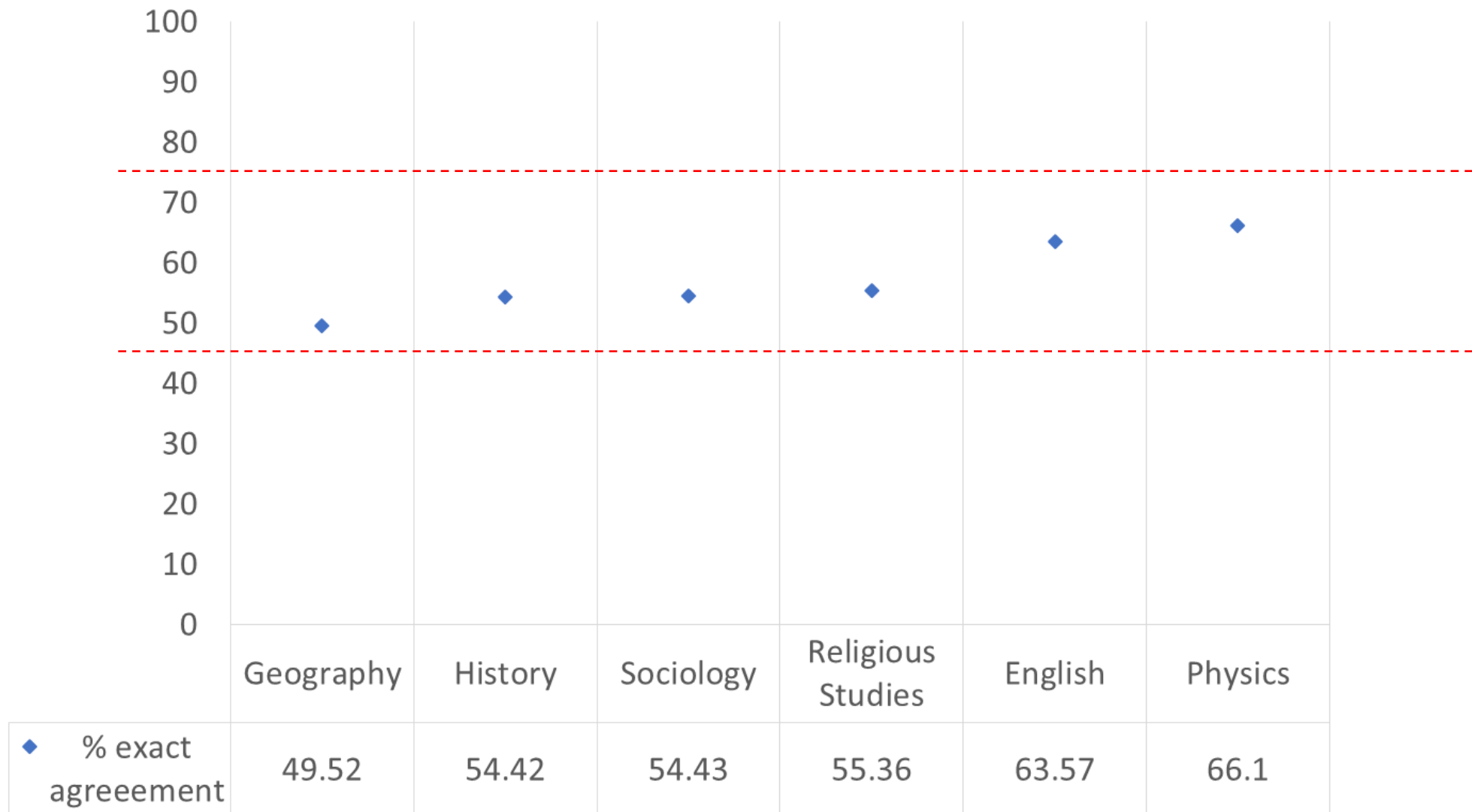
Geography

39,091 items

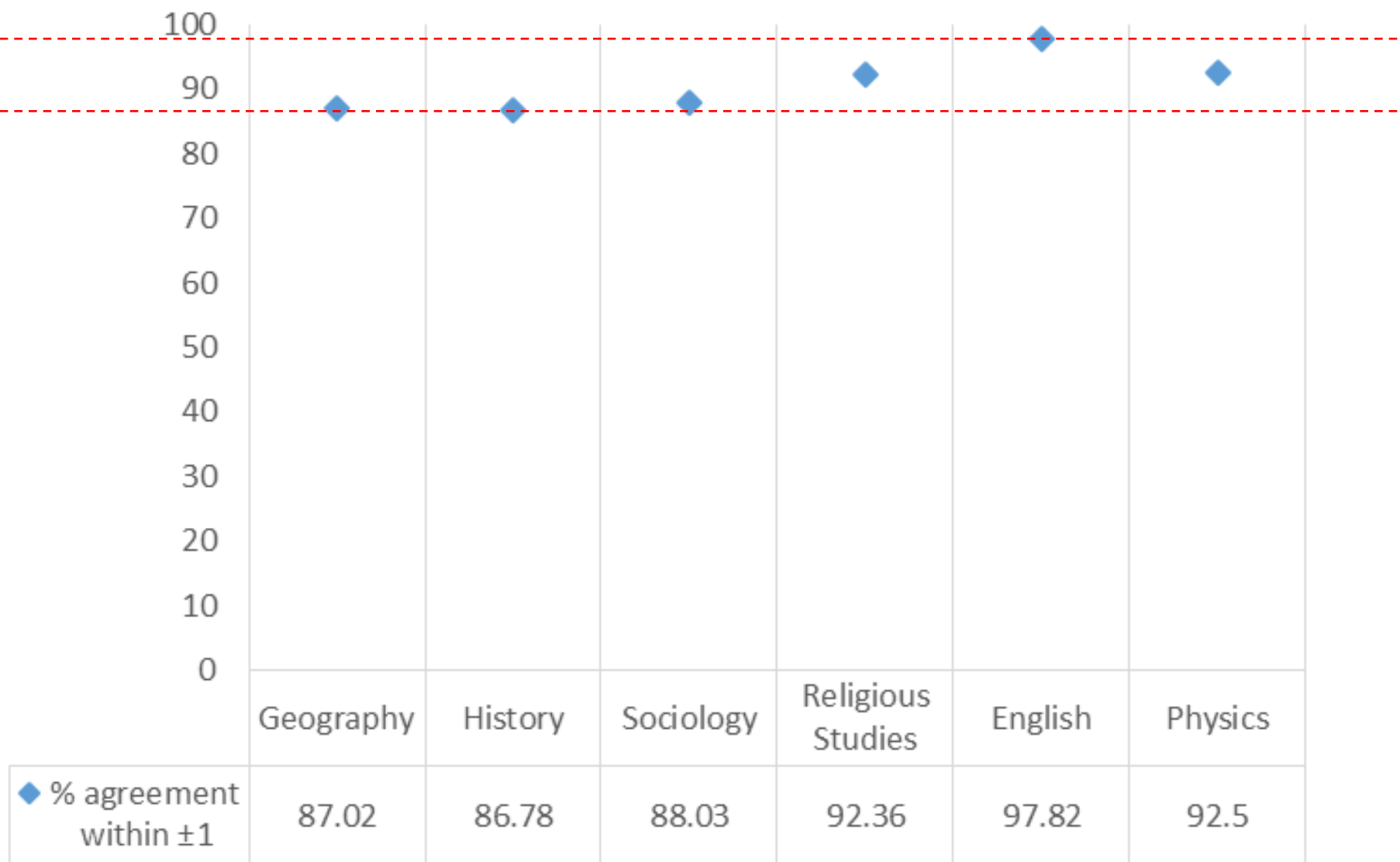


Difference	Percentage (%)
-3	1.04
-2	5.46
-1	19.15
0	49.52
1	18.35
2	5.09
3	1.04

6 mark items in different subjects – exact agreement



6 mark items in different subjects – within +/- 1 mark



Rolling programme of designed marking reliability studies

- Based on intelligence from the metrics and teachers' reports
- Qualifications not yet decided but likely to include over time GCSE English, A level English Literature, A level History, A level Religious Studies
- Re-marking of cleaned scripts by senior examiners
- Will tease out errors from legitimate differences of opinion
- Will explore how marking might improve e.g. by highlighting the characteristics of hard to mark answers
- Over time will provide a benchmark against which we can better evaluate reliability and take regulatory action

Reviews of marking and appeals

Julie Swan

Context

- Exam outcomes are increasingly high stakes – for students, teachers and schools and colleges
- Growing number of reviews of marking being requested – typically where the mark is just below the grade boundary
- If assessments are to be valid, marker judgement must often be used

Our aims for the review and appeal arrangements

- Marking errors are found and corrected
- Review and appeal arrangements are more transparent and consistent
- The system is as fair as it can be for all

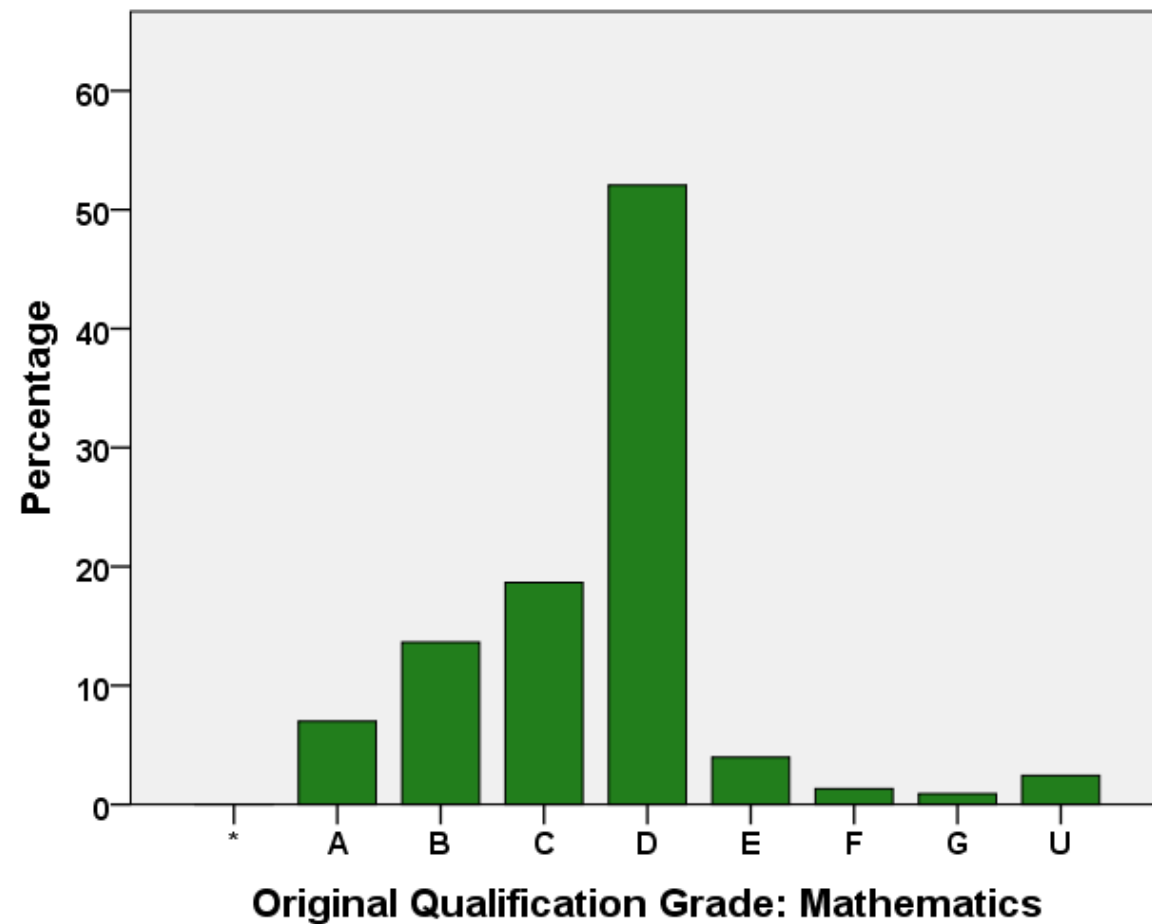
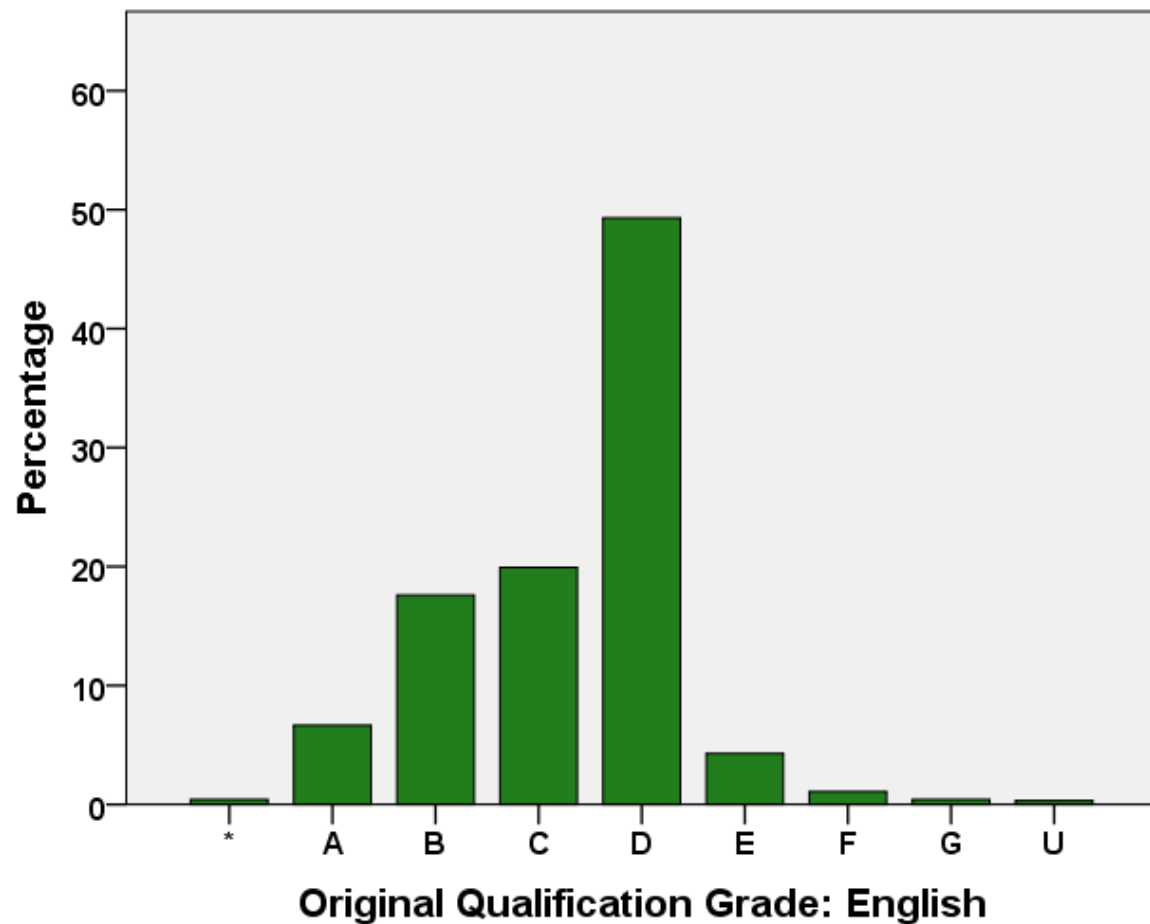
Concerns with current arrangements

- Our research found that
 - reviewers will often 'find' an extra mark or two
 - this can disrupt the rank order of students
- Schools and colleges feel under pressure to request reviews
- Schools and colleges find appeal arrangements opaque and frustrating
- The system can be unfair

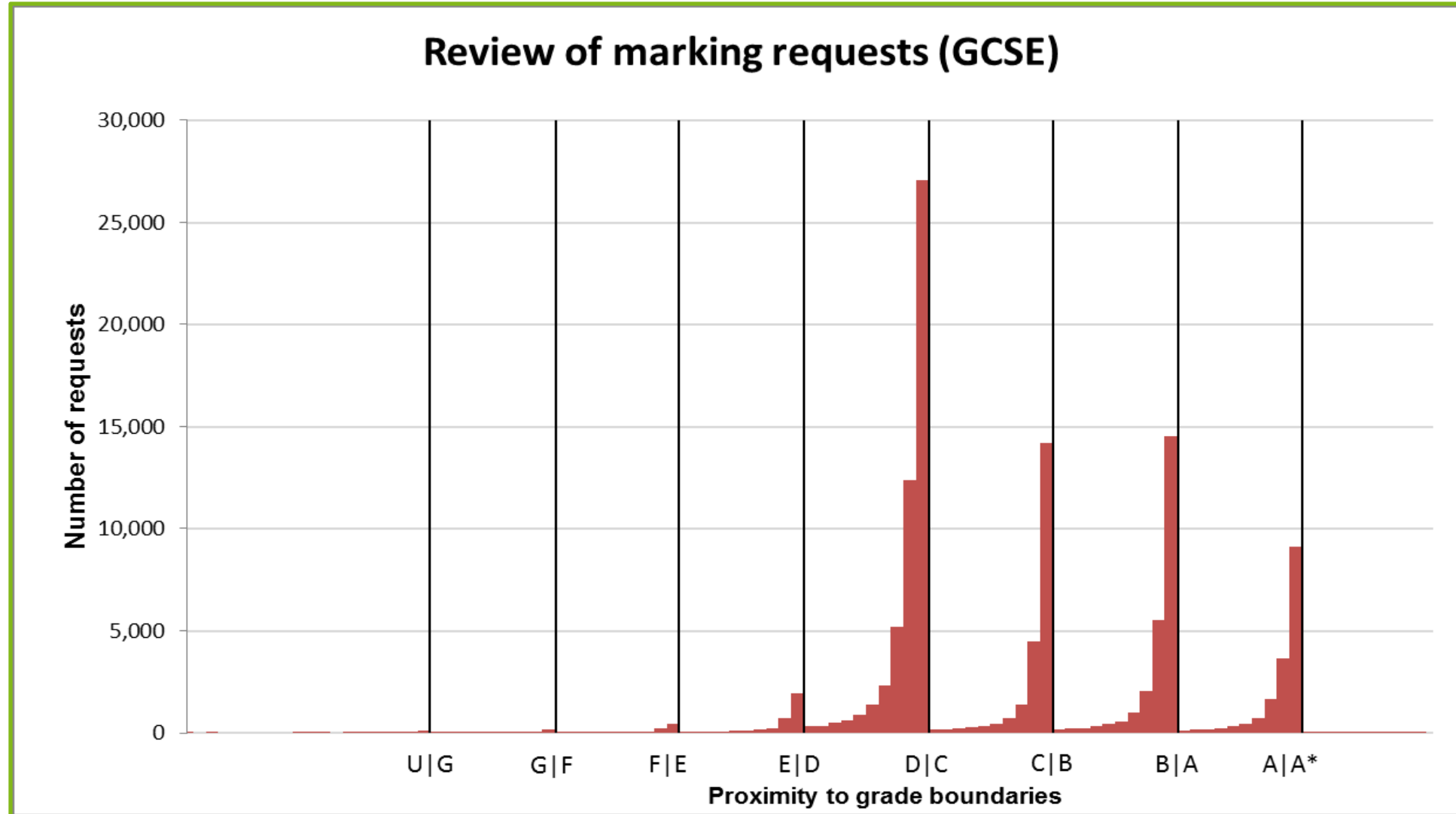
Progress

- Research into current practice
- Policy consultations Dec 2015 – March 2016 (the what)
- Technical consultation May 2016 – June 2016 (the how)
- We have announced our key policy decisions
- Some detail is yet to be decided and announced – our technical consultation closed last week

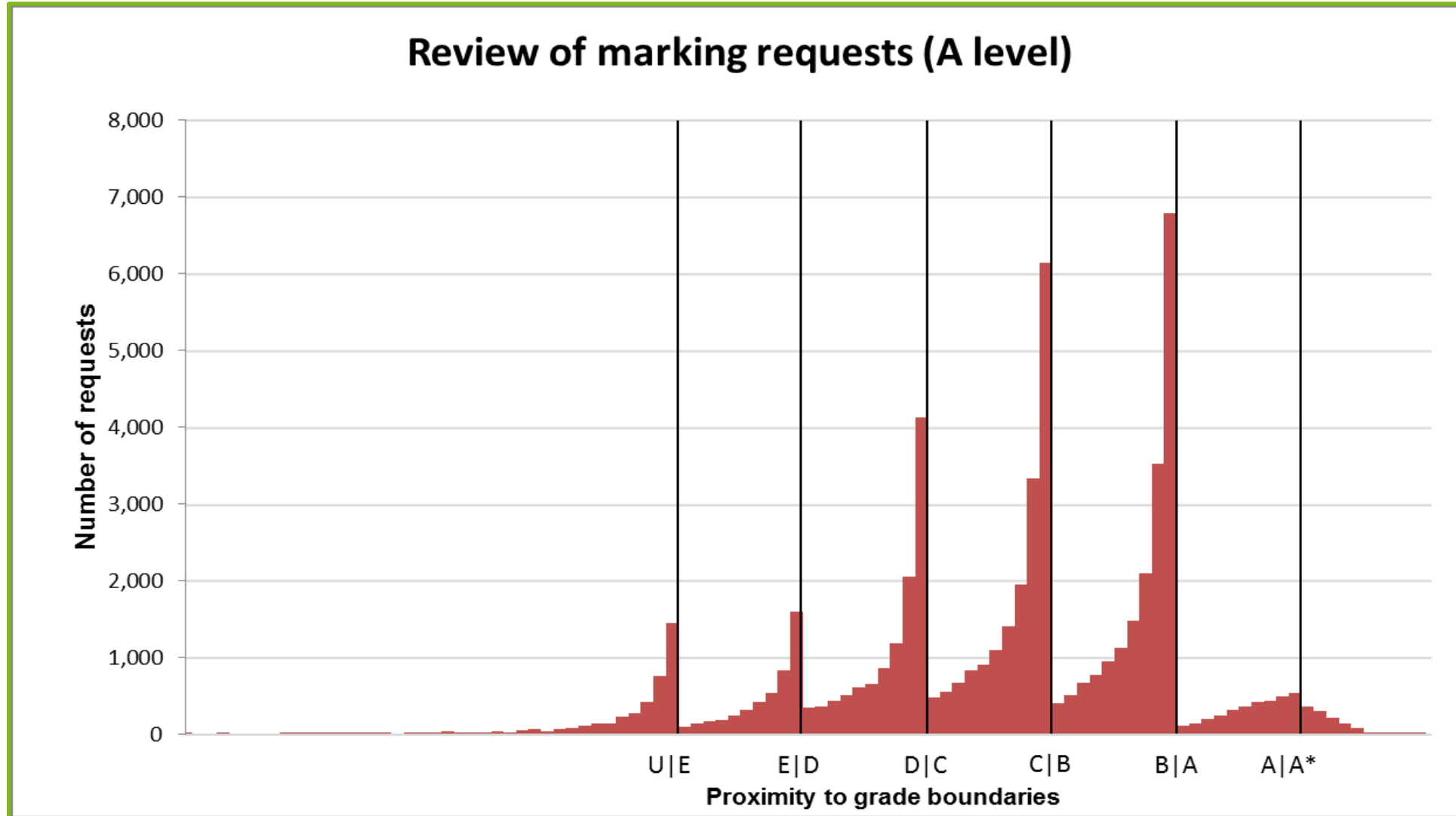
Pattern of GCSE enquiries about results in English and maths



Requests for reviews are more likely to be made when a mark is just below a key grade boundary




Requests for reviews are more likely to be made when a mark is just below a key grade boundary



ORIGINAL RESULT

3 students



mark 53

Students

Implications of existing system on rank order

The rank order is disrupted and those whose mark wasn't reviewed are potentially disadvantaged


REVIEW

2 reviews requested



POST REVIEW RESULT

1 student



mark 53

Student

This student didn't request a review and keeps the mark (and the grade) they were originally given

1 student




mark 53

Student

This student keeps the mark (and the grade) they were originally given

1 student



mark 54

Student

This student is given 1 extra mark on review; although the reviewer didn't find a marking error had originally been made, they get the higher grade

53

Grade boundary

54

Our key decision:

- marks must only be changed where there was a marking error

To bring this about

- Trained reviewers will check the original marking
- Reviewers will be monitored to promote consistency
- If they find an error they must correct it - however large or small
- Otherwise they must leave the mark unchanged
- Errors might arise from administrative error, incorrect application of the mark scheme or unreasonable exercise of judgement
- We have consulted on guidance on how reviewers should find and deal with errors

The definition of a marking error on which we consulted:

- The awarding of a mark for a task which could not reasonably have been awarded given the evidence generated by the learner, the criteria against which learners' performance is differentiated and any procedures of the awarding organisation in relation to marking, including in particular where the awarding of a mark is based on:
 - an administrative error
 - failure to apply such criteria and procedures to the evidence generated by the learner where that failure did not involve the exercise of academic judgement, or
 - an unreasonable exercise of academic judgement

Examples of unreasonable exercise of academic judgement included in our draft guidance

- Marking is unduly lenient or harsh, beyond what might be expected
- Part of a student's answer wasn't credited with a mark, although any marker, who had been properly trained and who had the appropriate subject knowledge, would have given a mark
- There was no rationale for the mark given

Complaints from schools and colleges

■ Quality of marking

2013 53

2014 75

2015 43

■ Enquiries and appeals process

2013 14

2014 12

2015 35

Key proposal on appeals

- Currently schools/colleges must make a case for an appeal based on a procedural failing by the exam board
- We have proposed exam boards should pilot in a small number of qualifications this year the opportunity for a school or college to appeal on the grounds of a marking error that was not corrected at review

Other issues

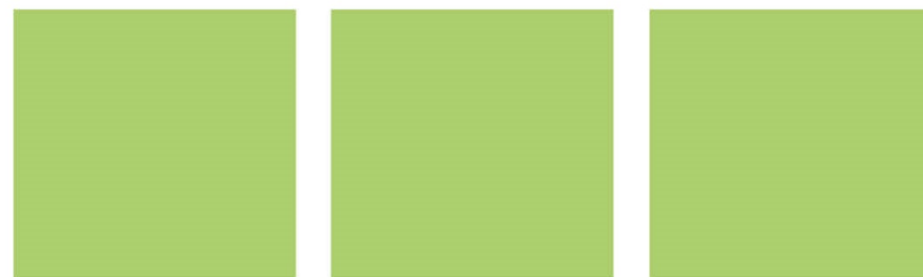
- Access to marked GCSE scripts
- Provision of reasons for review outcomes
- Reviews of teacher marking
- Reviews of the moderation of teacher marking

Further update on reviews and appeals

- Next week
- Including information for schools and colleges

A solid green square is located in the top-left corner of the page.

Accreditation



Accreditation

Accreditation comes at the end of a lengthy process of qualification reform during which:

- the Government develops and consults on subject content
- Ofqual develops and consults on assessment requirements – eg tiering, balance of exam and non-exam assessment and assessment objectives
- the exam boards develop their qualifications

Our test on accreditation is:

Has the exam board demonstrated to us that it's capable of meeting all our rules for the qualification, on an **on-going basis?**

Exam boards' submissions to us include:

- The specification
- Sample assessment materials
- Their assessment strategy for the qualification

Accreditation panels

- Each submission considered in detail by a panel
- Ofqual staff chair, but most panel members are subject experts
- 2016 – 107 subject experts
- Focus on quality of assessments and mark schemes – linked to quality of marking

Subject content for 2017 subjects – already published by the Government

<u>GCSE ancient history</u>	4 th February 2016
<u>GCSE astronomy</u>	17 th December 2015
<u>GCSE business</u>	17 th December 2015
<u>GCSE classical civilisation</u>	4 th February 2016
<u>GCSE design and technology</u>	4 th November 2015
<u>GCSE economics</u>	17 th December 2015
<u>GCSE electronics</u>	4 th February 2016
<u>GCSE engineering</u>	17 th December 2015
<u>GCSE film studies</u>	4 th February 2016
<u>GCSE geology</u>	17 th December 2015
<u>GCSE media studies</u>	4 th February 2016
<u>GCSE physical education short course</u>	1 st March 2016
<u>GCSE psychology</u>	17 th December 2015
<u>GCSE sociology</u>	1 st March 2016
<u>GCSE statistics</u>	4 th February 2016

Subject content for 2017 subjects – already published by the Government

<u>GCE AS and A level accounting</u>	4 th February 2016
<u>GCE AS and A level ancient history</u>	4 th February 2016
<u>GCE AS and A level archaeology</u>	4 th February 2016
<u>GCE AS and A level classical civilisation</u>	4 th February 2016
<u>GCE AS and A level design and technology</u>	17 th December 2015
<u>GCE AS and A level electronics</u>	4 th February 2016
<u>GCE AS and A level environmental science</u>	17 th December 2015
<u>GCE AS and A level film studies</u>	4 th February 2016
<u>AS and A level further maths</u>	18 th December 2014 (but some formulas added April 2016)
<u>GCE AS and A level geology</u>	1 st March 2016
<u>GCE AS and A level history of art</u>	21 st January 2016
<u>GCE AS and A level law</u>	4 th February 2016
<u>AS and A level maths</u>	18 th December 2014 (but some formulas added April 2016)
<u>GCE AS and A level media studies</u>	4 th February 2016
<u>GCE AS and A level music technology</u>	17 th December 2015
<u>GCE AS and A level philosophy</u>	17 th December 2015
<u>GCE AS and A level politics</u>	1 st March 2016
<u>GCE AS and A level statistics</u>	1 st March 2016

Timetable for qualifications to be taught from September 2017

- Expecting 120 submissions
- Decisions on first submissions being communicated from June
- We can only accredit when an exam board satisfies the accreditation test

Follow progress

- There is a detailed schedule on our website which we update at least weekly:

<https://www.gov.uk/government/publications/accreditation-of-gcses-as-a-levels-for-teaching-from-2017>

