Quality of Marking

Description of the Marking Process Used in External Exams in General Qualifications

Contents

Introduction

This paper provides an overview of the process of marking general qualifications (including GCSEs, A levels and equivalent academic qualifications), focusing on the main activities through which exam boards control the quality of marking of external exams:

1.   Recruiting and retaining examiners

2.   Training examiners

3.   Pre-standardisation and standardisation

4.   Checking traditionally marked scripts/papers

5.   Checking scripts marked on-screen

6.   Carrying out post-marking checks.

This paper does not attempt to provide an exhaustive description of the marking process, nor identify every instance where practices vary between exam boards.

We have collected the data below from visits to and interviews with exam boards between March and October 2013. We interviewed all the exam boards that provide general qualifications in England. These are: AQA, Pearson Edexcel, OCR, Cambridge International Examinations (CIE), the International Baccalaureate (the IB), WJEC CBAC Limited, and the Council for the Curriculum, Examinations and Assessment (CCEA).

1. Recruiting and retaining examiners

Recruitment

Exam boards recruit their examiners through external advertising. Successful applicants go into a pool of examiners from which the examiners required for each exam series are taken. Typically, examiners who have marked a unit successfully in the past will be the first choice to mark in the next series, supplemented by as many new examiners from the pool as are required. Exam boards forecast the number of examiners they require for each series at least six months in advance, using entry data from the previous series and preliminary entry data for the current series. Exam boards generally over-recruit examiners for each series to allow for any examiners who might drop out during the marking, either because the exam board decides to stop the examiner from marking or the examiner chooses to stop marking.

Currently, exam boards are not forecasting their need for examiners beyond the next 12 months. For example, few are systematically investigating any possible risks around an aging examiner population that could result in future examiner shortages.

The minimum requirements for examiners vary by exam board and subject. Minimum requirements for clerical markers and graduate markers are lower than for expert markers, although they should still have an undergraduate degree. Expert markers must have an undergraduate degree in the subject they wish to examine or a related subject. Expert examiners usually must be qualified teachers (most exam boards specify at least one year's teaching experience, often in the subject and at the level they wish to mark). Exam boards do not require their examiners to have experience of teaching the specification they wish to examine, however the majority do.

In almost all cases, the assessment process for new examiners includes them completing an application form, which is then reviewed by the exam board against its examiner criteria. CIE is the only exam board to test prospective examiners' aptitudes for examining as part of the recruitment process, via the Test 2 Assess assessment, which requires applicants to mark scripts from a previous series. OCR is also planning to introduce the Test 2 Assess training in 2014.

Exam boards aim to promote from within, so team leaders are recruited from high-performing examiners. High-performing team leaders with sufficient experience are encouraged to apply for principal examiner vacancies, although these vacancies are also advertised externally. Some exam boards require their principal examiners to go through the recruitment process again, when a new specification is introduced.

Given the importance of their role, the most senior examiners[1] go through a more rigorous recruitment process. Depending on the exam board, this may involve assessment-centre-style testing or an interview. These roles are recruited on an individual basis, although some exam boards require their senior examiners to go through the recruitment process again, when a new specification is introduced.

Retention

All the exam boards believe their retention of examiners from one exam series to the next is good or excellent, although most were not able to provide data to support this assertion. CIE reported its retention of examiners between one series and the next at around 95 per cent. The IB reported an annual attrition of around 1 per cent of its examiners. None of the exam boards reported any instances where they had experienced a mass exit of examiners.

---

[1] Senior examiners include principal examiners, chief examiners and chairs of examiners.

Most exam boards do not systematically collect the reasons given by examiners for choosing not to continue marking for them. CIE, OCR and WJEC are the only exam boards that collate this feedback and use it to identify any patterns that need to be addressed.

2. Training examiners

This section considers the training available to examiners, excluding the training provided as part of the standardisation process (which is the key activity that prepares examiners to mark, see section 3).

All the exam boards provide some online training materials and guidance documents for their examiners. These materials focus on the technical elements of examining, such as how to use the on-screen marking system and undertake the administrative tasks involved in examining. Some exam boards are able to track the extent to which each examiner accesses these materials, for example which documents an examiner opens or downloads and whether an examiner completes the test questions at the end of a training module. Face-to-face training for examiners (beyond standardisation meetings) is not standard. However, it is carried out by some exam boards for certain priority[2] subjects. For example AQA currently train over 50 per cent of examiners face-to-face due to priority subjects being those with the highest numbers of new examiners.

All examiners have their performance evaluated at the end of each exam series. Where this process identifies a training need for an examiner (usually around the administrative tasks of examining), the exam board may provide this training before inviting the examiner to examine a future series. However, in subjects where there is no shortage of examiners, it is more common for examiners with a training need to be removed from the pool of examiners. Most exam boards tend to assess their examiners' performance on a five-point scale. In the summer 2012 exam series, 28 per cent of examiners[3] received the top rating, 43 per cent received the second highest rating and 23 per cent received the third highest rating. These examiners would usually be invited to continue marking for their exam board. Two per cent of examiners received the lowest rating and would not be invited to mark that unit again. A further 4 per cent of examiners received the second lowest rating and may have

---

[2] Exam boards might define priority subjects differently; this may include complex subjects with a lot of extended response items and include level based mark schemes or subjects with high numbers of new examiners.

[3] Data provided by AQA, CIE, Pearson Edexcel, OCR and WJEC.

been able to mark the unit again upon successful re-training. All examiners may, however, be eligible to mark different units, as performance on one unit is not necessarily a good indicator of future performance on another.

New team leaders receive additional training on how to use the exam board's on-screen marking system (because team leaders use much more of its functionality than examiners do). They are also monitored and mentored by their principal examiner, but, with the exception of AQA, OCR and Pearson Edexcel, team leaders do not receive any formal training in the soft skills required to manage a team of examiners. They have their performance reviewed by their supervising examiners, but not by their team.

The level of training available to principal examiners, chief examiners and chairs of examiners varies by exam board. Across all the exam boards, the subject manager or qualification lead supervises the principal examiner and provides guidance, mentoring and performance reviews which could result in additional training. In addition, CIE and Pearson Edexcel have formal programmes to provide assessment training for their senior examiners:

- CIE runs an annual programme of face-to-face senior examiner training. Online training to supplement the face-to-face training is currently in development.

- Pearson Edexcel has invested in a Chartered Institute of Educational Assessors training programme, accredited by the University of Durham, which covers all stages of the assessment process.

3. Pre-standardisation and standardisation

The purpose of standardisation is to ensure that all examiners are marking consistently to a common standard, set out in the mark scheme.

Pre-standardisation

Pre-standardisation is almost always delivered via a face-to-face meeting, the aim of the meeting is discuss student responses across the papers within the unit and make amendments to the mark scheme if necessary. The meetings are attended by the principal examiner, assistant principal examiners, all or some of the team leaders and representatives from the exam board's subject team. The purpose of the pre-standardisation meeting is to:

- Mark and agree:

  - true scores for a selection of scripts/items that will be used as practice scripts/items for the standardisation meeting;

      □   approval scripts/items to qualify examiners to mark at the end of standardisation;

      □   seeds for use in live marking (on-screen marking only).

      □   the meeting also ensures that this senior team have internalised the standard set by the principal examiner.

These scripts/items are selected to reflect the full range of student performance and the different options that may be available within the question paper.

- Review and finalise the mark scheme prior to standardisation in light of student responses.

In traditional marking it is usually the role of the principal examiner to select practice and approval scripts, often from their own allocation of scripts to mark, in advance of the pre-standardisation meeting. On-screen marking involves selecting more scripts/items (due to the use of common approval scripts/items and seeds), so team leaders are often involved in the selection of appropriate scripts/items prior to the pre-standardisation meeting. Because of the larger volume of scripts/items being prepared, pre-standardisation meetings for on-screen marking usually last two or three days, compared with one day for traditional marking.

Standardisation

Standardisation of examiners usually takes place at a face-to-face standardisation meeting or via online standardisation. Exam boards differ in the proportions of online and face-to-face standardisation they use:

Methods of standardisation used by exam boards in 2013

Exam board	Percentage of each exam board's units/components standardised through:			
	Face-to-face standardisation	Online standardisation	Webinar (online meeting)	Other method/no standardisation[4]
AQA	34%	55%	0%	11%
CCEA	100%	0%	0%	0%
CIE	44%	35%	0%	21%
The IB	0%	21%	12%	67%
OCR	17%	83%	0%	0%
Pearson Edexcel	25%	28%	35%	13%
WJEC	97%	3%	0%	0%

Pearson Edexcel and the IB use webinars – online, virtual meetings – as a third way of standardising their examiners. A webinar standardisation meeting follows the same process as a face-to-face one.

Face-to-face standardisation

Face-to-face standardisation meetings are attended by all the examiners who will be marking a unit in the question paper. These meetings follow the process described in section 4 of our *GCSE, GCE, Principal Learning and Project Code of Practice.*[5] Prior to a face-to-face standardisation meeting, examiners must carry out some provisional marking to familiarise themselves with the mark scheme. The principal examiner begins the meeting with an introductory briefing and a walk-through of some exemplar scripts. Examiners then usually break into their teams (led by each team leader) to discuss and mark the practice scripts prepared at the pre-standardisation meeting. Typically, examiners mark a script and then stop to discuss the marks they gave to each question. The team leader records the marks that each examiner gives to each script: this record should show the team's marks converge with each additional script they mark until they are all marking to the same standard.

---

[4] The remainder of units either required no standardisation (the principal examiner completed all the marking) or used 'other' techniques.

[5] www.ofqual.gov.uk/files/2011-05-27-code-of-practice-2011.pdf

Following the standardisation meeting, the examiners are cleared to mark. If the examiners are marking traditionally, they will be asked to mark a selection of scripts (usually a minimum of ten) from their own allocation of scripts to mark. These are reviewed by the examiner's team leader, who gives the examiner feedback and either clears the examiner to mark or asks for an additional sample of marked scripts if he or she is not satisfied with the quality of marking shown in the first sample. Following the additional sample, examiners are either cleared to mark or stopped from marking. Common pre-marked scripts are used by OCR (where three of the ten scripts are common) and CIE (where in exceptional circumstances some common scripts may be used).

If examiners are marking on-screen, the process of clearing them to mark is similar to that for traditional marking, except examiners log onto the exam board's on-screen marking system and mark a number of approval scripts/items that are common to all examiners marking that unit. Examiners are either standardised online or face-to-face and the mode of marking does not have to be the same method.

Online standardisation

Online standardisation follows a three-stage process:

1. Examiners log onto the on-screen marking system to familiarise themselves with the mark scheme and review some live scripts. At some exam boards the team leader will contact each examiner to give a personal briefing.

2. Examiners mark practice scripts. The number of practice scripts provided varies from ten (CIE and OCR) to three to five (the IB). Examiners mark approval scripts (this is sometimes referred to as the first phase sample). CIE and OCR require examiners to mark a minimum of ten approval scripts, WJEC six approval scripts, AQA between five and ten approval scripts and the IB five approval scripts.[6] Pearson Edexcel and WJEC require examiners to qualify for marking at item level.

The approval scripts/items are reviewed by the team leader, who either clears the examiner to mark or asks the examiner to complete a second set of approval scripts/items if the quality of the examiner's marking does not meet the required standard.

---

[6] The IB requires examiners marking on-screen to complete five approval scripts. Examiners marking traditionally do not complete approval scripts. Instead, their marking is sampled when the marking process is 30 to 40 per cent of the way through (see section 4).

Clearing team leaders to mark

Where common scripts/items are used to clear examiners to mark, CIE, OCR and WJEC give any team leaders who were present at the pre-standardisation meeting (where the common scripts/items were selected) a separate allocation of scripts/items they have not seen before. AQA require team leaders to pass the common scripts during the pre-standardisation period. Pearson Edexcel has a number of alternative methods for clearing team leaders to mark if they are already familiar with the common scripts/items. At the IB, the team leaders who are involved in common scripts/items selection are not required to pass approval scripts/items and pass directly to live marking.

4. Checking traditionally marked scripts

During live marking, the team leader samples examiners' marking to check they are continuing to mark accurately. Team leaders have their marking sampled by the principal examiner or, in subjects with larger entries, by an assistant principal examiner.

Three exam boards (CIE, OCR and Pearson Edexcel) take two samples of each examiner's marking during live marking, and three exam boards (AQA, the IB and WJEC) take one sample. This is in addition to the approval sample that qualifies an examiner for live marking. During live marking, examiners are told how many scripts to provide in their sample and given guidance about the types of scripts they should include. Exam boards vary in how much discretion they give examiners to select their own samples: some exam boards require samples to be taken from specific schools or colleges; others (such as the IB) request specific scripts. As a minimum across all exam boards, examiners' samples must reflect a broad range of student performance and, where applicable, reflect a range of optional questions within the question paper. In almost all cases, examiners provide a larger number of scripts than the team leader reviews. This allows the team leader to judge which scripts from within each sample to review.

The following table shows, for each exam board, the approximate point in the marking process the samples are taken, and how many scripts are sampled:

	First sample			Second sample		
Exam board	How far through marking?	Minimum scripts provided	Minimum scripts re-marked	How far through marking?	Minimum scripts provided	Minimum scripts re-marked
AQA	50%	50	15	NA[7]	NA	NA
CIE	40%	50	15	100%	50	5
CCEA	10%	15	8	50%	25	8
The IB	30 to 40%	20	20	NA	NA	NA
OCR	40%	50-60	10	100%	40	10
Pearson Edexcel	30%	20	10	95%	40	10
WJEC	30-40%	80	25	NA	NA	NA

Across all exam boards, the number of scripts reviewed is a minimum: team leaders have discretion to request larger/additional samples and review a larger proportion of scripts if they have any doubts about an examiner's performance.

The total number of scripts marked by an examiner in an exam series is likely to vary both across examiners, subjects and the different exam boards. Therefore, the proportion of each examiner's scripts that is re-marked by his or her supervising examiner will also vary across examiners and exam boards.

Although the team leader is subject to the same checks on his or her marking as other examiners, the team leader's sampling of examiners' marking is not checked as a matter of course. It is assumed that if team leaders are marking accurately, they are also sampling examiners' marking accurately. CIE and OCR are the exception to this rule: at either the first or second sample stage the principal examiner will ask the team leader to provide all the marking he or she received at the first sample stage from one examiner in his or her team.

Most exam boards apply a marking tolerance when reviewing each examiner's sample. In most cases, this tolerance recognises that there can be legitimate differences in the professional judgement of examiners. Many exam boards apply a 6 per cent marking tolerance, agreed through the Joint Council for Qualifications as a maximum. For many subjects the marking tolerance is lower: in the most objective subjects, such as maths, the marking tolerance may be zero. The IB takes a slightly

---

[7] In summer 2013 AQA piloted the use of a second sample during live marking.

different approach to marking tolerances and uses linear regression to identify marking that may need further review. Pearson Edexcel uses adjacency values, which are not a recognition of any acceptable level of variation but a flag to identify examiners who may need closer scrutiny.

For those exam boards taking two samples of an examiner's marking, there are three possible outcomes as a result of the first marking sample:

■   the examiner's marking is accurate (within tolerance): no action required;

■   the examiner's marking is aberrant: the examiner is stopped from marking and all of their scripts are re-marked;

■   there is some doubt about the examiner's marking: the examiner is allowed to continue marking and a final decision will be taken when the second sample has been reviewed (or at marking review).

Examiners usually do not receive feedback on their first marking sample. This is because the exam board may decide, following the second sample, to apply scaling to an examiner's marks. Scaling is applied if an examiner's marking is within tolerance but consistently lenient or severe, and is generally applied to all the scripts marked by that examiner. For example, if an examiner marks all of his or her scripts too leniently by two marks, all of his or her scripts would have two marks deducted. If an examiner received feedback at the first sample stage, he or she would be likely to alter his or her marking behaviour, and this would make it impossible to apply scaling across all of the scripts they marked.

The decision to stop an examiner marking is a judgement made by the exam board based upon the available evidence and the recommendation of the team leader and principal examiner.

At the final sample of scripts, a final decision is taken about each examiner's marking:

■   marking is accurate and no action is required;

■   marking is lenient or severe, and consistent: scaling applied;

■   there is lingering doubt about the examiner: the scripts he or she has marked are passed to marking review (see section 6).

Classifying an examiner as "lingering doubt" is a judgement made by the team leader and/or exam board representatives. Examples of when a team leader may decide to pass an examiner's marking to marking review include:

- The examiner has been recommended for scaling, but there is some doubt about whether the examiner's marking is sufficiently consistent for scaling.

- The examiner's marking is broadly within tolerance, but there are some inconsistencies present.

5. Checking scripts marked on-screen

Exam boards use three main methods to monitor the quality of on-screen marking:

1. Seed scripts/items[8]
These are scripts/items that have already been marked and given a true score at pre-standardisation or at the pre-standardisation phase. Seeds appear at random in an examiner's marking allocation and must be marked by the examiner within tolerance. Seeds typically appear at a rate of about 5 per cent, 10 per cent in the case of the IB. That means, out of 20 scripts/items marked by an examiner, one is likely to be a seed. Exam boards differ in their approach to varying the rate of seeds: some, such as OCR and WJEC vary the rate of seeds by the unit that is being marked. CIE and the IB vary the rate of seeds by individual examiner, if necessary, and for Pearson Edexcel seed rates for clerical markers are different than for other mark types, therefore they vary their seed rates by marker type. CIE also varies seeding rates during the marking period with a greater proportion of scripts being seeds for the initial portion of the examiners marking. Two exam boards (AQA and WJEC) also use daily approval seeds: the examiner must pass these seeds to qualify for live marking that day.

Three exam boards (AQA, the IB and WJEC) will automatically suspend an examiner from marking if he or she fails too many seeds. The examiner is locked out of the on-screen marking system and cannot carry out any more marking until the team leader restores access. The other exam boards use seeds as a performance flag: failure at seeds by an examiner will trigger further investigation by the team leader, but the examiner is not automatically suspended from marking. As with traditional marking, across all the exam boards, any decision to stop permanently an examiner from marking is a judgement made by the exam board and informed by the recommendation of the team leader and/or principal examiner.

---

[8] AQA, Pearson Edexcel and WJEC examiners mark on-screen at item level and, therefore, seed items are used. CIE, the IB and OCR examiners mark whole scripts on-screen and, therefore, seed scripts are used. Exam boards using item-level marking are able to stop examiners from marking items they appear to be struggling with, but examiners can continue marking other items.

The IB is the only exam board that gives examiners access to immediate feedback on all the seeds they have marked (via comparison with the marks given by the principal examiner or team leader), irrespective of whether they have passed or failed those seeds. All the other exam boards only give feedback on seeds to examiners if they fall below the minimum standard, although AQA and Pearson Edexcel provide their examiners with a summary report on their performance on seed items at the end of the marking process.

Seeds are the main way in which exam boards monitor the accuracy of examiners' marking, and can be supplemented by two other methods:

2. Double marking
This method is used by AQA and WJEC in some of the more subjective subjects, and runs concurrently with, or instead of seeding. Around 5 to 10 per cent of items in these scripts (usually the higher tariff items) are double marked. Two examiners mark the same item and, if their marks are out of tolerance with each other, a senior examiner adjudicates and decides the correct mark. Examiners found to be marking incorrectly receive penalties, and usually examiners are stopped from marking an item if they have received two penalties for that item.

3. Spot-checking (or back reading)
Team leaders are able to review each examiner's completed marking. Some exam boards include spot-checking as a formal part of their quality assurance processes for on-screen marking, and they specify a minimum number of scripts that team leaders should review for each examiner. Other exam boards leave spot-checking entirely to their team leader's discretion. Spot-checking is typically used to investigate a possible issue, for example to review the marking of an examiner whose performance on seeds has caused concern.

6. Carrying out post-marking checks

Exam boards carry out three different types of checks after all the scripts have been marked:

a) Marking review
The marking review takes place only for scripts that have been traditionally marked. Any examiners about whom there is lingering doubt (see section 3) are put forward to the marking review, where a sample of their marked scripts are re-marked by a principal examiner or team leader. This process of re-marking a sample of scripts typically takes place as a meeting at the exam board's head office, or by post if there are a very small number of scripts to review. Depending on the findings from the re-marking of an examiner's scripts, the reviewing examiner may:

- confirm that the examiner's marking is accurate, so no further action is required;

- decide that the examiner's marking is aberrant, in which case all the scripts marked by the examiner will be allocated for re-marking;

- take a further sample of scripts from that examiner to reach a final decision.

Exam boards vary in their approach to the marking review. AQA, CIE and the IB focus the initial stages of their marking review on those students whose marks place them just below a grade boundary. These are the students who are most at risk of receiving the wrong grade due to errant marking. However, across all the exam boards, if an examiner is found to be aberrant, all of his or her scripts are re-marked.

WJEC's marking review process considers all traditionally marked scripts. The findings from the samples of each examiner's marking are reviewed alongside statistical information to identify examiners whose marking needs to be re-marked. This process effectively combines marking review and marking data analysis (see below) into a single activity.

The marking review process is heavily dependent on senior examiner judgement. There are generally no formal guidelines that specify how many scripts from each examiner there is doubt over should be sampled as part of the marking review, or to inform the decision about whether it is necessary to re-mark all of an examiner's scripts. However OCR always considers 20 scripts at marking review for each lingering doubt examiner.

b) Marking data analysis

Exam boards use marking data to identify any patterns that may indicate errant marking. Typically, this process includes both scripts marked traditionally and on-screen, but, for most exam boards, usually there is more detailed data available for on-screen marking. Exam boards vary in the analyses they choose to carry out. These analyses include:

- comparing a school or college's performance in a unit on a given year with its performance in the same unit the previous year: a school or college's performance may vary from one year to the next for legitimate reasons, but a significant difference may indicate errant marking, particularly if all of a school or college's scripts have been marked by one examiner;

- comparing the mean and distribution of marks given by each examiner for a unit to the average mean and distribution of marks for that unit: this identifies any examiners whose marking varies significantly from the average;

- comparing students' predicted grades with their actual grades: very significant differences may indicate errant marking.

Across all these examples there could be valid reasons for the differences that have nothing to do with the accuracy of examiners' marking. Therefore, these analyses are only ever used as flags to identify possible issues that require investigation through a review of the scripts.

Exam boards vary in the extent to which they use marking data analysis. Some, such as CIE and WJEC, find it a very useful technique and make significant use of it. OCR does not currently use marking data analysis at the end of the marking process. However, OCR is piloting the use of marking data to monitor examiner performance during marking and is also investigating the use of regression modelling at the end of the marking process to provide a statistical flag for errant marking.

c) Clerical checks

The most common clerical checks are carried out for traditionally marked scripts, to double check that all the marks have been included, transcribed and added up correctly. These checks are generally carried out by exam board staff.

For on-screen marking, Pearson Edexcel and OCR carry out additional clerical checks, including checking scripts where a student scored zero in any items. These scripts receive a clerical check followed by an examiner check to look for any responses in the script the original examiner may have missed (for example, if the student failed to provide their response in the space provided for that question).

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

First published by the Office of Qualifications and Examinations Regulation in 2014

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation

Spring Place	2nd Floor
Coventry Business Park	Glendinning House
Herald Avenue	6 Murray Street
Coventry CV5 6UB	Belfast BT1 6DN

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