

Research and Analysis

## Online standardisation

Observations, interviews and survey

## **Authorship**

This report was prepared by Caroline Morin, Stephen Holmes and Beth Black in Ofqual's Strategy, Risk and Research Directorate

# Contents

<b>Executive summary</b> .....	<b>44</b>
<b>1 Introduction</b> .....	<b>6</b>
1.1 <i>The online standardisation process</i> .....	6
<b>2 Previous research</b> .....	<b>7</b>
<b>3 Methodology</b> .....	<b>11</b>
<b>4 Analysis</b> .....	<b>12</b>
<b>5 Results</b> .....	<b>12</b>
5.1 <i>Software training</i> .....	15
5.2 <i>Familiarisation phase</i> .....	16
5.3 <i>Practice phase</i> .....	18
5.3.1 Viewing annotations and rationales on answers .....	18
5.3.2 Number of exemplars and rationales in the practice phase .....	21
5.4 <i>Personal feedback</i> .....	24
5.4.1 Personal feedback provided after the practice phase .....	24
5.4.2 Personal feedback provided after the qualification phase .....	26
5.4.3 Content of personal feedback .....	28
5.4.4 Feedback delivery mode .....	29
5.4.5 Timing of feedback .....	31
5.4.6 Training on providing/receiving feedback .....	32
5.5 <i>Qualification phase</i> .....	35
5.5.1 Second set of qualification scripts .....	35
5.6 <i>Other relevant findings</i> .....	36
5.6.1 Importance of practice and qualification phase .....	36
5.6.2 Examiners' satisfaction .....	38
5.6.3 Perceived advantages and disadvantages of online standardisation .....	39
<b>6 Discussion</b> .....	<b>38</b>
<b>7 Conclusion</b> .....	<b>40</b>
<b>References</b> .....	<b>41</b>
<b>Appendix A – Survey</b> .....	<b>44</b>
<i>Survey methodology</i> .....	44
<i>Survey results</i> .....	45
Respondents' characteristics .....	44
Practice phase .....	50
Qualification phase .....	61
Overall experience .....	69
<b>Advantages and disadvantages of online standardisation</b> .....	<b>80</b>

## Executive summary

Ofqual's report on quality of marking (2014) reported that examiners tended to be dissatisfied with their experience of online standardisation and raised a number of issues. To understand these issues further, Ofqual carried out a project on online standardisation aimed at gaining a deeper understanding of the processes involved in online standardisation and identifying practices that could be implemented to improve the experience and performance of examiners carrying out online standardisation.

The project consisted of the direct observation of examiners completing online standardisation followed by semi-structured interviews. A wider survey of examiners was also carried out about their experience of online standardisation (n = 12,681). Based on findings from these 3 methods, a process map was devised capturing good practices that were found across the exam boards.

In this report, we are referring to online standardisation as a multi-staged process consisting of up to 3 discrete phases: familiarisation, practice and qualification. While examiners will be monitored and receive feedback after qualification, this report focuses on the learning process. In the familiarisation phase, examiners look at papers alongside a provisional mark scheme. In the practice phase, examiners mark items/scripts and can have access to the definitive mark, the rationale for awarding the mark and the annotations showing where the marks were given. Examiners also receive personal feedback from their team leader on their performance. Finally, the qualification phase is where examiners mark items/scripts and their performance is used to decide whether they can start live marking. They also can get personal feedback from their team leader after that phase.

The main findings were:

- the majority of examiners reported being happy or very happy with online standardisation. They valued the flexibility and not having to travel. However, they did report that not meeting their team leader or other examiners and having less in-depth discussions around the mark scheme was a disadvantage of the online process
- we identified some good practices: the speed at which team leaders provided personal feedback to examiners, the most frequent medium is by phone which allows for a discussion and the fact that most examiners receive personal feedback after the qualification phase
- however, these are not always implemented or are not implemented consistently

- it is important for examiners to know not only if their marks are the same as the definitive mark<sup>1</sup>, but also if they are awarded on the same basis as was intended. Annotation, written rationales of definitive marks and personalised feedback are key mechanisms and can be helpful to inform future marking. However, less than half of examiners said they could view clear rationales for how marks were awarded. Moreover, the annotation/rationales were not always viewed: examiners were likely only to view these if their marks disagreed with the definitive mark, despite the possibility that their marks had been awarded on the wrong basis
- personalised feedback after the practice phase occurred around half of the time; it was more likely to occur after the qualification phase
- examiners tended to believe that the qualification phase was more important than the practice phase in helping them to understand how to apply the mark scheme correctly. In some ways this is inevitable; whether or not a person passes the qualification phase has a very clear impact on what happens next, it is the determinant of whether they can start to mark live scripts
- two thirds of examiners reported feeling confident at the start of live marking. While confidence is not always associated with competence, a third of examiners reported not feeling confident at the point at which they started live marking. It is not possible from this study to know whether this is higher or lower than levels of confidence where there has been face-to-face or another type of standardisation. Nevertheless, it may indicate some scope for more learning and/or confidence building prior to live marking.

This research will serve as a useful basis to engage with awarding organisations around online standardisation in order for them to make improvements where possible.

---

<sup>1</sup> The definitive mark is the mark awarded by the principal examiner and in a hierarchical marking system is usually considered to be the correct mark

# 1 Introduction

In February 2014, Ofqual published a series of reports looking at the quality of marking in A levels, GCSEs and other academic qualifications. One of the reports described the findings from a “Survey of Examiners” carried out in May 2013. A total of 10,204 respondents completed the survey which covered many aspects of marking (motivation, mark schemes, training, etc). One of the main findings was that a large number of examiners mentioned the lack of detailed feedback as a problem but not necessarily in the context of online standardisation. Moreover, examiners were dissatisfied with the move from face-to-face standardisation to online standardisation. Most of the examiners who were dissatisfied made the general comment that they believed face-to-face standardisation to be more effective than online standardisation.

The main aim of the project was to investigate how online standardisation is implemented across the exam boards, how it is experienced and perceived by examiners and where there may be opportunities to increase its effectiveness. The current project stems from this dissatisfaction and aims to outline practices that could be implemented to improve the performance and experience of examiners carrying out online standardisation. The project consisted of 3 separate strands: 1) direct observation of 8 examiners from 4 exam boards carrying out online standardisation 2) semi-structured interviews with the examiners observed and 3) a wider survey of examiners about their experience of online standardisation.

## 1.1 The online standardisation process

Online standardisation was first introduced in the mid-2000s and is now the most frequent method of standardisation for assistant examiners for most general qualifications. Exam boards differ slightly with regard to their processes for online standardisation and the terminology that they apply. For clarity and consistency, we will outline 3 significant phases of online standardisation and will use this terminology throughout this report.

The first phase of online standardisation (see figure 1), the *familiarisation* phase, is not present for all exam boards. During that phase, examiners have access to the provisional mark scheme and can browse live scripts in order to familiarise themselves with both the mark scheme and the different ways candidates have answered the questions. There is no definitive mark (set by the principal examiner or other senior markers) or rationale available at that point. The main purpose of this phase is to gain basic familiarity with the questions and begin to gain a sense of the kinds of ways in which candidates have responded. Sometimes examiners are asked to send their comments or suggestions about the mark scheme to their team leader who can then share their comments with the principal examiner at the standardisation meetings.<sup>2</sup>

---

<sup>2</sup> Typically, for papers with smaller numbers of candidates, team leaders take part in a pre-standardisation meeting where the standardisation scripts for assistant examiners are chosen and, if needed, adjustments are made to the mark scheme to reflect actual responses encountered. For larger units, only the most senior examiners take part in the pre-standardisation meeting and the team



Figure 1. Online standardisation phases leading to live marking.

The second phase of online standardisation, the *practice* phase, is also not present for all exam boards/components (though usually there is at least one of familiarisation and practice phases). The practice phase can be seen as the main opportunity for learning within the online standardisation process. During this phase, examiners are provided with a mark scheme and they are usually expected to complete their practice scripts/items within a window of a few days. The practice scripts/items<sup>3</sup> are usually chosen to exemplify the mark scheme. During that phase examiners log in to their marking system and start marking usually around 5 to 10 scripts or responses for each question. Once they submit their marked scripts/items, they sometimes can see the ‘definitive’ annotations on the scripts as well as a written rationale explaining why the marks were awarded. They also sometimes receive personal feedback from their team leader before they are cleared to start the ‘qualification’ or ‘approval’ phase.

During the qualification phase, the examiners will mark scripts/items as they did in the practice phase. Once they submit their scripts/items, they usually get personal feedback from their team leader. Unlike the other phases, the qualification phase will determine whether an examiner is allowed to start marking live scripts. If the marks awarded are too dissimilar to the marks given by the principal examiner, the examiner will usually be asked to complete another set of scripts/items in order to “qualify” for marking live scripts.

## 2 Previous research

Previous research has not looked specifically at the different processes within online standardisation, but a few studies have been carried out looking at the effectiveness of online standardisation compared to face-to-face standardisation in the context of general qualifications in England.

Billington and Davenport (2008) ran a pilot to compare online standardisation and face-to-face standardisation and found that although small differences existed, they were most probably driven by different procedures in the selection of scripts. In the online standardisation condition, the definitive marks were compared to those

---

leaders are standardised face-to-face in a standardisation meeting a few days after the pre-standardisation meeting was held.

<sup>3</sup> Some exam boards ask examiners to mark complete scripts whilst others use “clips” which are individual responses.

awarded by the examiner. In the face-to-face standardisation condition, examiners chose which of the scripts they had marked to send back to their team leader for back-reading<sup>4</sup>; the team leader would therefore see the marks awarded. There is evidence that this leads to smaller mark differences as the team leader might be influenced by the marks awarded by the examiner (Fearnley, 2005; Vidal Rodeiro, 2007). Moreover, Baird, Meadows, Leckie & Caro (2015) also found that back-reading led to an underestimation of overall rater inaccuracy compared to systems using an expert (principal examiner) generated definitive mark.

Billington and Davenport (2008) also asked the examiners to complete a survey on their views of the system. The most beneficial aspects of online standardisation identified by examiners was the fact that they could complete online standardisation at their own pace. Other practical aspects were also cited as being beneficial: completing online standardisation at home or another place of their choice, at a convenient time and without the need to take days off work. However, two thirds of examiners completing online standardisation for the first time reported a decrease in their depth of understanding of the mark scheme. Also, more than 97% of examiners found it isolating and more than a quarter of examiners who had carried out standardisation online for the first time the previous year said they felt less enthusiastic at using the system.

Chamberlain and Taylor (2011) compared the marking accuracy of 2 groups of experienced examiners marking a paper for the first time. One group took part in face-to-face standardisation while the other did online standardisation. In this study, both groups marked the same scripts which had been previously marked by the principal examiner (to give the definitive mark). All examiners had completed face-to-face standardisation in the past but had never done online standardisation. Examiners marked 60 scripts each, 30 before training and 30 after training. When looking at the absolute mark difference from the definitive mark at paper level, the face-to-face group was less accurate than the online standardisation group before training but both groups improved after training. Analyses showed a significant effect of training but no effect of the mode of standardisation. Similar effects were found for the absolute mark difference at item level. Finally, the rank order correlation to the definitive mark for both groups was high before training and went up slightly after training. This can be seen as demonstrating that the mark scheme alone had a strong standardising effect. However, the increase was significant and showed that training increased the ability of examiners to rank candidates in the same way as the principal examiner but that the type of training did not have a significant effect.

Under research scrutiny, online standardisation seems to be as effective as face-to-face standardisation. However, these were small studies which only included experienced examiners and this may have impacted on the results.

Other research has been carried out which is relevant to specific aspects of the online standardisation process.

---

<sup>4</sup> Back-reading refers to the reviewing of a script by the team leader where the marks awarded by the examiners are visible.



## Personal feedback

Feedback is a key feature of learning through practice. Johnson (2016) carried out a literature review on the effectiveness of feedback in professional assessment contexts. The aim was to identify characteristics of the feedback messages that are effective in a remote and hierarchical environment similar to the one in which assistant examiners and team leaders interact when carrying out online standardisation. The 4 core factors are language, content, timing and form which include the mode of delivery and those will be discussed in more detail below.

According to Johnson (2016), the aim of feedback is “to bring the less senior examiner’s interpretation of the mark scheme in line with those of more senior examiners”. Feedback is hence considered to be effective when it supports the alignment of interpretations.

Feedback is most effective when it is individual to the recipient as opposed to being generic (Ahmed et al, 2012). It should relate to specific aspects of the examiners performance instead of simply restating the mark scheme. Knoch, Read and von Randow (2007) suggested that when provided with generic feedback, the examiner needs to decipher the comments and link them back to their own performance. This reduces the feedback effectiveness as it relies on the examiner being able to relate the generic feedback to their own performance so as to inform and improve future performance.

The mode of communication used to deliver feedback becomes important when considering the number of communication dimensions it supports, referred to as ‘media richness’ by Daft and Lengel (1984). They propose that media differ in their ability to transmit the meaning of information on 4 information richness dimensions: language variety, synchronicity, personal focus and multiple cues.

Types of communication media that score low on the 4 dimensions are poor, for example those that exhibit no synchronicity and offer minimal cues. Any reduction in cues offered by a mode of communication restricts the information available to an addressee to decipher the intent of a message.

Email as a type of text-based communication affords quick transmission, however they convey no non-verbal cues which can often signal important information regarding the speaker’s intent. Non-verbal cues can refer to speaker rate, intonation, stress and loudness, and in face-to-face or video-mediated communication, facial expression and hand gestures are examples of non-verbal cues (Bull, 2001). Non-verbal cues are especially important when the literal meaning of a message is ambiguous (Kruger, Epley, Parker & Ng, 2005). In comparison to voice only and face-to-face communication, the amount of information regarding meaning in an email is thus restricted. In addition, the use of email leaves little opportunity for potential misunderstandings to be first noticed and subsequently resolved given the lack of synchronicity. These limitations render email communication more susceptible to misunderstanding.

The use of synchronous modes of communication can help examiners have the confidence to ask questions. Asking questions is a key personal resource as “standards” cannot be explained but they can be understood via discussion and questioning (Price, 2005). According to Handley, den Outer and Price (2013) it is the

discussion around the exemplars that makes them so powerful. Discussion provides an opportunity to “identify, and then convey as fully as possible, the essence of the ‘knowledge’ that makes up the standards.” (Sadler, 2005, p.193).

There is some evidence suggesting that as well as allowing for clarification, video communication carries as much as 4 times more information than text (Stannard, 2008). It also promotes a sense of community and can improve connection (Mathieson, 2012).

### **Providing feedback**

Johnson and Black (2012) conducted a small-scale study looking at the monitoring of assistant examiners by team leaders, mainly centred on feedback. They identified 2 types of team leaders: evaluators and comparators. Evaluator team leaders give feedback that is more personalised, tailored and explanatory. Comparator team leaders give less personalised, more generic, feedback and are less likely to be concerned by the assistant examiner’s confidence. The comparator team leaders also give less praise. Johnson (2016) suggested that team leaders could benefit from training on how to deliver feedback to the examiners they supervise. In a different context, Grant (2007) showed that a 2-day coaching skills training programme could improve coaching skills and hence feedback delivery. The distinction between the 2 types of team leaders could help define the training required by team leaders in order to provide the most effective feedback.

Johnson and Black (2012) also found that feedback is perceived to be more powerful when it is received soon after the marking has occurred and early intervention made it easier for examiners to change their marking. Similarly, Rovai (2000) suggests that immediate feedback reduces isolation, increases social presence and increases learner motivation in a distance learning environment. When feedback is received a long time after submission, examiners can feel isolated or find it difficult to relate to their own performance (Johnson & Black, 2012). Also, feedback should not only be provided when things go wrong (Ahmed et al, 2012).

Stephenson and Cadwallader (2014) compared the effect of additional feedback during live marking. One group of examiners had the usual support of their team leader during live marking (providing feedback when problems arise) whilst the second group received additional feedback (team leaders were trained on how to deliver feedback and contacted examiners 3 times across the marking window). They found no statistically significant difference in the reliability of marking for the two groups but examiners who had been given additional feedback felt more confident and enjoyed marking more than those who didn’t get the additional feedback. This is consistent with the finding from Meadows (2004) suggesting that the clarity of marking procedure and the competency of senior examiners were aspects that were strongly associated with examiners’ intention to continue marking. There is also evidence that the longer an examiner has been marking the more reliable they become (Pinot de Moira, 2005, as cited in Stephenson & Cadwallader, 2014). Given the importance of reliability in marking, it might be important to consider these variables in order to retain as many examiners as possible. This could also lead to a reduction in costs associated with appointing and training new examiners (Stephenson & Cadwallader, 2014).

## **Feedback, annotations and rationales**

Handley, den Outer and Price (2013) suggested that examiners should read the rationale and look at the annotations for each question, even when their marks agree with the definitive marks. This is in part to make sure that they have awarded the marks for the right reasons as agreeing a mark is not enough to confirm that examiners share the same underlying standard.

For the items where the examiner's mark was different from the definitive mark, there is some evidence that the availability of the definitive answer (in this case the rationale, annotations and definitive mark) leads to an increased retention compared to when the only information available is whether the answer was correct or incorrect (Pashler, Cepeda, Wixted & Rohrer, 2005).

The rationale provided should follow a set format and should bridge how each aspect of the answer is rewardable to the relevant area in the mark scheme (Johnson & Black, 2014). There is some evidence that the use of standardised terms in the feedback could help reduce the examiners' cognitive load (Coninx, Kreijns & Jochems, 2013).

## **Exemplars**

Several research studies have considered the number and type of exemplars that may be optimal to learn the standard. Sadler (1987, 2005) suggested that a small number of exemplars alone can't adequately define a standard when a number of criteria are used. Baird, Greatorex and Bell (2004), after manipulating the type of exemplar presented in standardisation, concluded that the ideal approach would be to give examiners exemplars illustrating the range of achievements associated with each mark band (where the item was marked using a levels of response mark scheme), instead of only exemplars at the top and bottom of the band.

In a study looking at the efficacy of training and the complexity of the cognitive marking strategies, Suto, Greatorex and Nadas (2009) found that training was more beneficial for "apparently more complex strategy questions" and that inexperienced examiners benefitted more than experts or experienced examiners (also see Weigle, 1998, 1999 and Elder, Knoch, Barkhuizen & Von Randow, 2005). It might be that some questions only need a small number of exemplars while others need many more.

The selection of scripts/items is a very time consuming process. Recent research has looked at whether the use of artificial exemplars might be helpful in speeding up the process. Vijayathanan, Pinot de Moira and Stringer (2016) found that it took less time to create artificial exemplars than to find the exemplars from the live scripts. Another advantage is that this could be completed before the examination is sat when the principal examiner may have more time. However, the aim of the study was to verify whether artificial exemplars could masquerade as live responses and more research would be needed to verify whether artificial exemplars lead to at least a similar level of marking accuracy compared to live scripts.

### 3 Methodology

This research included 3 strands: observations of assistant examiners completing their practice or qualification scripts, semi-structured interviews with the examiners observed and finally a survey sent to all assistant examiners of four exam boards.

The observations and interviews were carried out in June and November 2016 with examiners from 4 different exam boards. All the examiners involved were assistant examiners although those observed in November were also team leaders during the summer series. The observations and interviews were carried out either in a quiet room in the Ofqual offices or at the exam board's offices. Where possible, examiners were observed through remote video recording using TechSmith Morae 3.3.4 usability software<sup>5</sup>.

Following the completion of their practice or qualification scripts, we interviewed the assistant examiners. The questions looked at the online standardisation process used by the exam board, their satisfaction with the process and whether it could be improved. The median interview duration was 70 minutes.

Finally, a survey was sent to all assistant examiners through the exam boards between July and September 2016. The survey covered subjects such as the number and quality of exemplars used, the feedback they received, their general satisfaction with the process, the advantages and disadvantages of the online standardisation process, how confident they were after completing online standardisation, whether they had any suggestion on how to improve the process and whether they had any other general comments. We received a total of 12,681 survey responses<sup>6</sup>.

### 4 Analysis

During the observation, notes were taken about specific events happening during the standardisation. These were used during the interview to ask more specific questions relating to what was highlighted during the observation.

The semi-structured interview comprised of a series of open questions. Depending on the answers given, the interviewer was then able to ask follow-up questions specific to the examiner's answer. The semi-structured interviews we analysed using thematic analysis. Additionally, the recordings of online standardisation were also analysed using thematic analysis.

Finally, the survey was analysed. All the closed questions were analysed taking into account the number of examiners who answered the questions after 'routing'. The open responses were analysed thematically.

---

<sup>5</sup> This software captures the material that is on the examiner's screen as well as a video feed of the examiner in order to monitor whether they are looking at the screen or at the mark scheme.

<sup>6</sup> The survey responses are presented in the appendix.

## 5 Results

From all 3 strands: the observation of the online standardisation process, interviews and survey responses, a process map was devised (see Figure 2). The process map shows all the possible stages of online standardisation (green boxes) together with aspects of some of the stages which can promote good practice (blue boxes). Online standardisation does not always include every process stage and the aspects in the blue boxes were sometimes implemented inconsistently within and between boards.

In the following sections, each feature identified in the process map will be discussed by pulling together evidence from the observations, the interviews and the results from the survey.

When quotes are used, they will either be from the interviews with examiners (labelled 'examiner' in the text) or from open questions in the survey asking about ways to improve the online standardisation process or asking for any other comments on the process (labelled 'respondent' in the text).

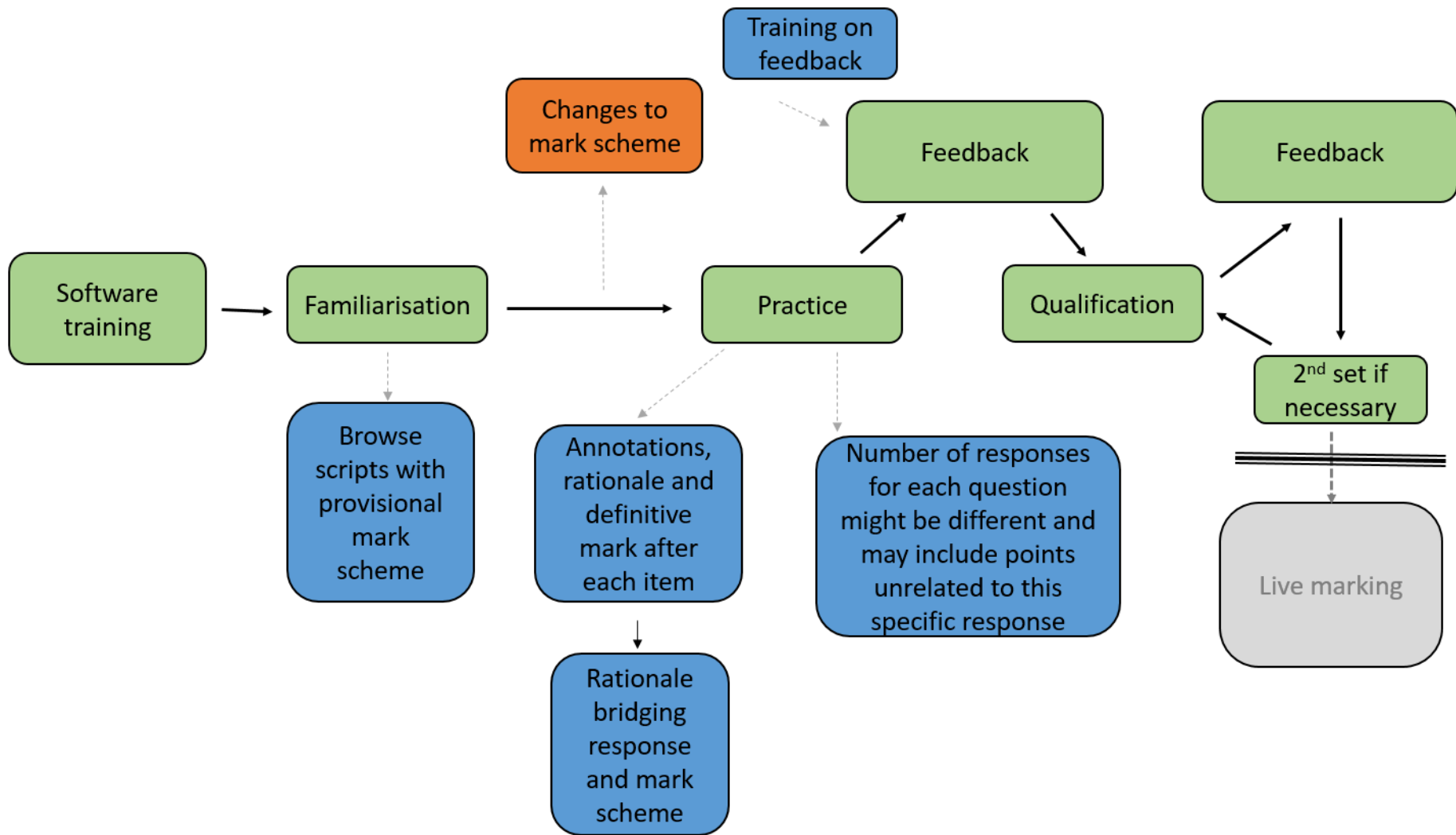


Figure 2. Process map of online standardisation.

## 5.1 Software training

Examiners mark items/scripts on a bespoke marking system. Each system has different functionality and ways of accessing each functionality. Some systems allow the examiners to choose whether they mark a complete script or whether they mark all answers from one question before moving to another question. Examiners also have access to specific annotations (ticks, crosses, etc.) set-up in the system. They can also see the annotations, rationale and definitive mark once they have submitted their practice scripts but each system has a different way of accessing these different features.

Observations of examiners completing the marking of their practice scripts showed that they were not always completely up-to-date with or on top of how to use the system. For example, some could not use the notes function and some were not clear on how to access the annotations and rationales that accompany the practice scripts. Some examiners who completed the survey raised the issue of a lack of training and there was also some evidence of a lack of knowledge about the system during the interviews.

*Examiner 1, male, GCSE, biology*

[interviewer - talking about how the system behaves during the practice phase] What are the red links next to a question? Are they already there on these practice scripts?

Not sure what that means... I thought it was when you put a level of response down.... I'm not sure.

This was also raised in response to the survey:

*Respondent 8868, female, GCSE, chemistry*

This year I asked my team leader where to find the annotations and definitive marks. She said there weren't any. Only when I marked another paper [using the same software] did I realise that I could see them as closed response. This caused massive issues as I didn't learn from the practice scripts as well or as quickly as I could have. Better communication about the practice scripts could have saved me a lot of stress!!

*Respondent 11706, female, A level, history*

No explanation on how to use online system or see qualification items. I lost 2 hours work because I didn't understand the system.

Ideally, it would be necessary to complete the software training before each marking session. Examiners have mentioned the existence of a training package but completion isn't compulsory every year, only when they first start marking.

*Examiner 6, female, GCSE, geography*

[interviewer] Before you start marking, do you have any training on the system?

Yes there is training on the system and I've done it in the past.

[interviewer] So you don't have to redo it every time?

I suppose if I was being totally professional I would probably look through it again.

*Examiner 2, female, GCSE, geography*

I bet the option [to do IT training] is there, I bet if I go into [the system] and have a look. There will be a training option that it would probably let me look at. Again it's not forced on you. And it really should come up on the screen to say you need to have gone through this, please check the box to say you have done so, so the responsibility is on the marker.

From this evidence, it appears that some examiners could benefit from refresher training on how the system works before they start marking each year. A number of examiners mentioned the existence of a training package but there doesn't seem to be a requirement to go through the training package before starting the online standardisation process. Ideally, all examiners would have to complete the training before each marking period so that the system can adequately support their learning of the mark scheme.

## 5.2 Familiarisation phase

Browsing scripts before the practice phase can give examiners an idea of how the questions were tackled by candidates. It also means that examiners can start applying the mark scheme earlier and that they have time to look closely at the mark scheme and potentially ask questions if there is anything they don't understand or they're not sure about. However, this phase is not present for all exam boards.

*Respondent 8409, male, A level, physics*

One advantage of the old examiner's meeting was that examiners had time to see scripts beforehand, and had an opportunity to raise specific problems. Use the dead time between exam & [the standardisation meeting] to supply 20 papers and a draft mark scheme so examiners can try out and see if there are any problems which they can raise [at the standardisation meeting].



*Respondent 8311, male, AS level, statistics*

[Exam board A] do not give assistant examiners sight of the provisional mark scheme until they are ready for us to start on the practice scripts. We are also not able to browse scripts to begin the process of familiarisation with both scripts and mark schemes. So we have to go into the practice scripts cold and the [qualification] scripts only a little bit warmer.

*Respondent 8018, male, A level, physics*

If examiners could view scripts prior to the [standardisation meeting] with access to the provisional mark scheme, they could help to identify areas where adjustments were necessary and identify unexpected responses. Allowing examiners to contribute ideas to their team leaders that might be used at the [standardisation meeting] to produce the final mark scheme would increase their sense of worth and involvement, and mirrors the system that used to exist before on line marking.

One potential risk of the familiarisation phase is that if there are any subsequent changes to the mark scheme after the standardisation meeting, it might be difficult to unlearn the provisional mark scheme and learn the final mark scheme.

*Examiner 7, female, GCSE, mathematics*

Now because I've marked some to one mark scheme, I've now got to unlearn that. I mean that's going to happen, you can't help that, that's just one of those things with this system really.

[later]

[interviewer] So is having two marks schemes confusing?

It sometimes is when one's fairly different than the other one, yeah. It's because you've been practising doing things. You know, you've done at least 10 probably more, so you've got into grips with sort of what you're giving the marks for, and then if it tweaks it slightly then you find you're giving a mark that you weren't giving before in the sense of you're giving the mark for something different than it was before. And I find that difficult really.

[interviewer] Do they tend to add stuff or change what's already there?

Well on this one they've changed some stuff that's there, and they do tend to add. I'm surprised because sometimes you start with a very short mark scheme and it ends up being very [long], pages more.

This suggests that not having access to familiarisation scripts can have a negative impact on examiners' confidence in their marking and their sense of involvement. If possible, some clean scripts and a provisional mark scheme could be made available after the exam is sat so that the examiners can familiarise themselves with the mark scheme and the way candidates have tackled each question. More research would be needed to assess the impact of changes in the final mark scheme on marking accuracy.

### 5.3 Practice phase

From the comments on how to improve the online standardisation process in the survey, a number of examiners suggested that there should be a practice phase after the familiarisation phase. They also mentioned that the definitive mark, annotations and rationale on why the marks were awarded should be available. This practice phase needs the mark scheme to have been finalised.

*Respondent 2930, female, GCSE, physics*

Maybe some practice scripts set out in the same way as online standardisation [qualification] where there are annotations and you can compare your marks with the senior examiner would be useful.

*Respondent 520, female, GCSE, chemistry*

There should be practice scripts AFTER the mark scheme has been finalised and BEFORE the [qualification] so there is more opportunity to discuss issues with a team leader.

The quotes above suggest that the inclusion of a practice phase before the qualification phase, where not already present, offers a major opportunity for learning how to apply the final mark scheme before qualification. The scripts/items in the practice phase should include access to annotations, rationale and the definitive mark once the items have been marked.

#### 5.3.1 Viewing annotations and rationales on answers

From the observations, it appears that some systems are not helping the examiner get the most out of the rationale and annotations. For example, we observed that in one system, the examiner needed to click on a number of tabs in order to see the rationale and, as previously mentioned, not everyone seemed to know how to access them.

We noticed during the observations that some examiners only looked at the rationale when the mark they had awarded was different from the definitive mark. As mentioned in the literature review, Handley, van Outer and Price (2013) suggest that

examiners should also read the rationale and annotations for the questions to which they awarded the same mark as the definitive mark in case their rationale for awarding that mark is incorrect.

However, if the rationale is not available for the items an examiner gets wrong, simply knowing that it was incorrect may not be particularly helpful (Pashler, Cepeda, Wixted & Rohrer, 2005).

*Examiner 2, female, GCSE, geography*

But equally if there's no comments on the ones I've got wrong, that's useless.

In the survey sent to assistant examiners, we asked how frequently the rationale and annotations were presented after marking a response in the practice phase.

Figure 3 shows that 70% of examiners could see a written rationale for all or most items. However, 17% said they could not see any written rationale during the practice phase.

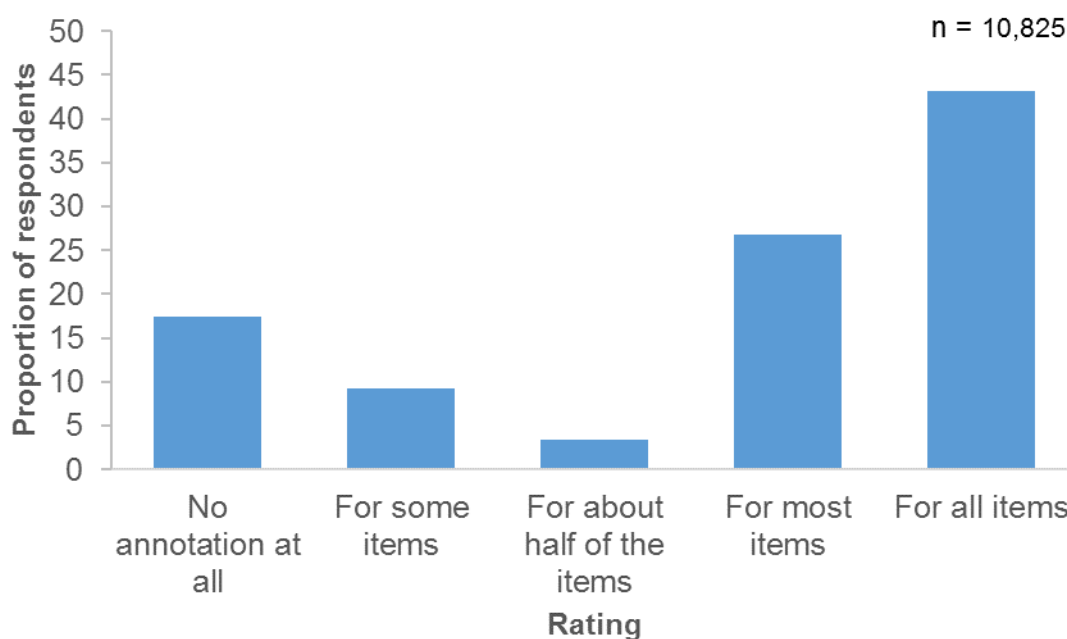


Figure 3. Proportion of examiners who reported that written rationale was present when marking the practice scripts.

A similar question asked how frequently examiners could see the annotation and definitive mark after marking a response in the practice phase. Two thirds of examiners stated that definitive marks and annotations were available for either most items or all items. Twenty-one percent of examiners said that no annotations were available (Figure 4).

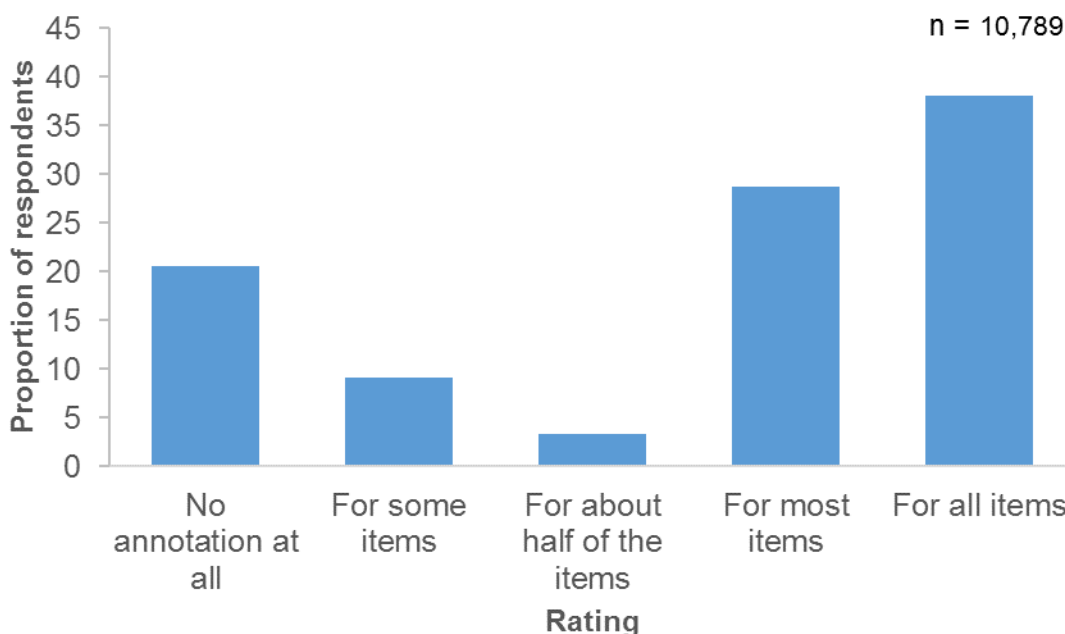


Figure 4. Proportion of examiners who reported that annotations were present when marking the practice scripts.

Finally, the rationale, annotations and definitive mark would ideally be available as soon as possible after the examiner has awarded a mark. We observed that for some exam boards, the examiners can only access the rationale, annotations and definitive mark once they submit a whole script. This means that for examiners marking by item, they will need to complete nearly all of their practice questions before they can have any information on their performance.

*Examiner 2, female, GCSE, geography*

That's the only way to find out the marks that you should have given: the definitive marks. You can only do it on a whole script. Now, to me, that is an absolute weakness of that system.

[later]

But you could only see them after you've finished the entire script, that entire script isn't helpful at all and if they're not there they're not there and there's nothing I can do about it and if I ring up a team leader, if there was no discussion in the room, they may not have any idea why the application happened either.

*Examiner 7, female, GCSE, mathematics [talking about examiners who have access to annotations only after a whole script is completed]*

Yeah, well I wouldn't like to do that ... otherwise you can keep on repeating the same error and by the time you've repeated it often enough you've learned it haven't you?

As mentioned in the literature review, the rationale provided should follow a set format and should bridge how each aspect of the answer is rewardable to the relevant area in the mark scheme (Johnson & Black, 2014). Also, the use of standardised terms in the feedback message have been shown to reduce cognitive load (Coninx, Kreijns & Jochems, 2013).

*Respondent 6186, AS level, accounting*

More quality/detailed comments for marks awarded on practice scripts.  
The reasons for the marks. Accurate placement of annotations eg ticks need to be precise as it is not always clear what point is being rewarded when dealing with a paragraph of words and the tick is somewhere on the right of the page.

Taken together the different strands of evidence suggest that the rationales, annotations and definitive marks should be available for all items, follow a similar structure linking the response to the mark scheme, be easily accessible on the system and as soon as possible after the examiner has marked a response.

### 5.3.2 Number of exemplars and rationales in the practice phase

In the survey, we asked examiners whether the exemplar used in the practice phase had covered a good range of quality. More than 70% answered that they did while 11% said they didn't (Figure 5).

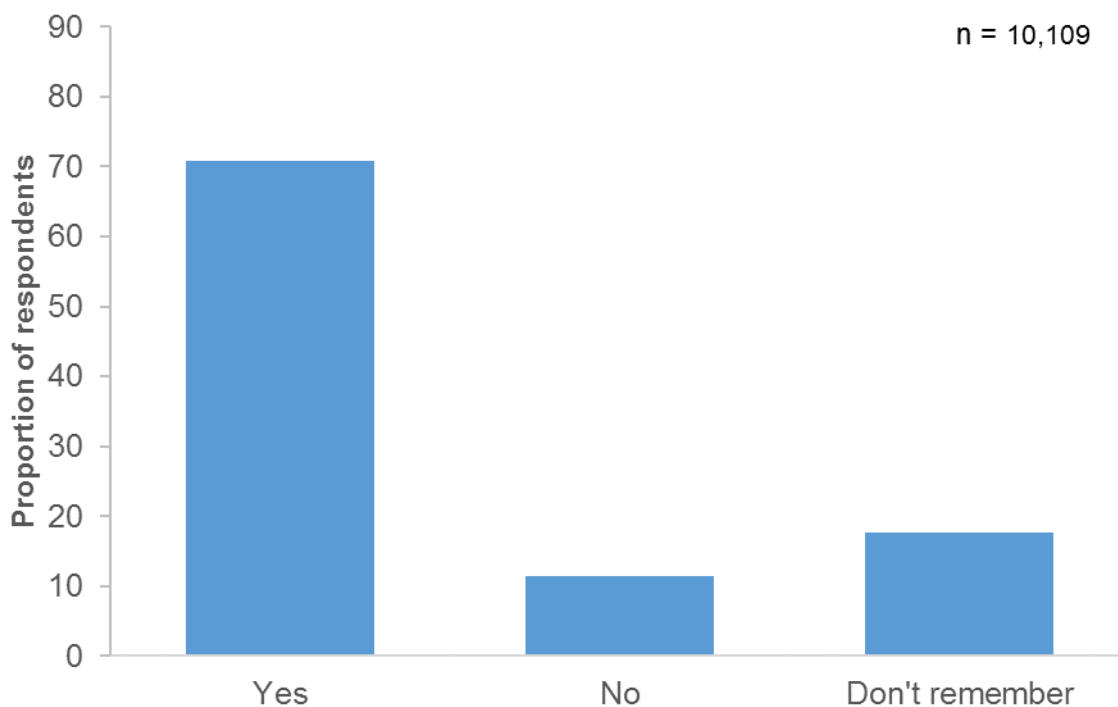


Figure 5. Proportion of examiners who reported whether the range of the quality in practice items/scripts was sufficient.

As previously indicated, it might be that some questions only need a small number of exemplars while others need many more (Suto, Greatorex & Nadas, 2009; Weigle, 1998, 1999). Also, exemplars from across the range of a band should be used and not only exemplars at the top and bottom of the band (Baird, Greatorex & Bell, 2004).

*Respondent 975, male, GCSE, physics*

Some questions could do with more practice questions, others with less. Some are obvious and need 2 examples others could do with 8!!

*Respondent 7958, male, GCSE, chemistry*

More practice scripts. No need to practice where there is only a single acceptable response eg Mg is the only acceptable answer

During the observations, we noted that one exam board included comments in the rationale unrelated to the specific exemplar but that gave additional guidance on how to mark that question. This allowed for more information to be given to the examiners without increasing the number of practice items/scripts. If the number of exemplars is to be kept to a small number, this strategy might be useful.

*Examiner 7, female, GCSE, mathematics*

There's sometimes a comment that might not relate to that particular question, but they've put an additional thing in to remind you, from the marks scheme you have to, you know, if this has happened then, even though it's not in that question. So they're covering more than what's actually that question, yes.

*Respondent 11464, male, GCSE, mathematics from a different exam board*

The practice questions do not always cover the content tested in the qualification questions and therefore do not fully prepare you for qualification or live marking. All areas of marking that will be tested by the qualification questions should have been covered by the practice questions.

In the survey, we asked examiners at what point they felt confident in their marking. Figure 6 shows the percentage of examiners who said they felt confident in their marking at the end of different phases in the process of marking. Thirty-three percent of examiners felt confident in one of the periods up to the end of standardisation, 66% by the beginning of live marking and 96% by the middle of live marking. Finally, 2% of examiners said they never felt confident in their marking. Because there was no option for a period between the beginning and middle of live marking, it is possible that the exact point is closer to the beginning than the middle

of live marking. Either way, it shows that almost a third of examiners don't feel confident in their marking when they start live marking.

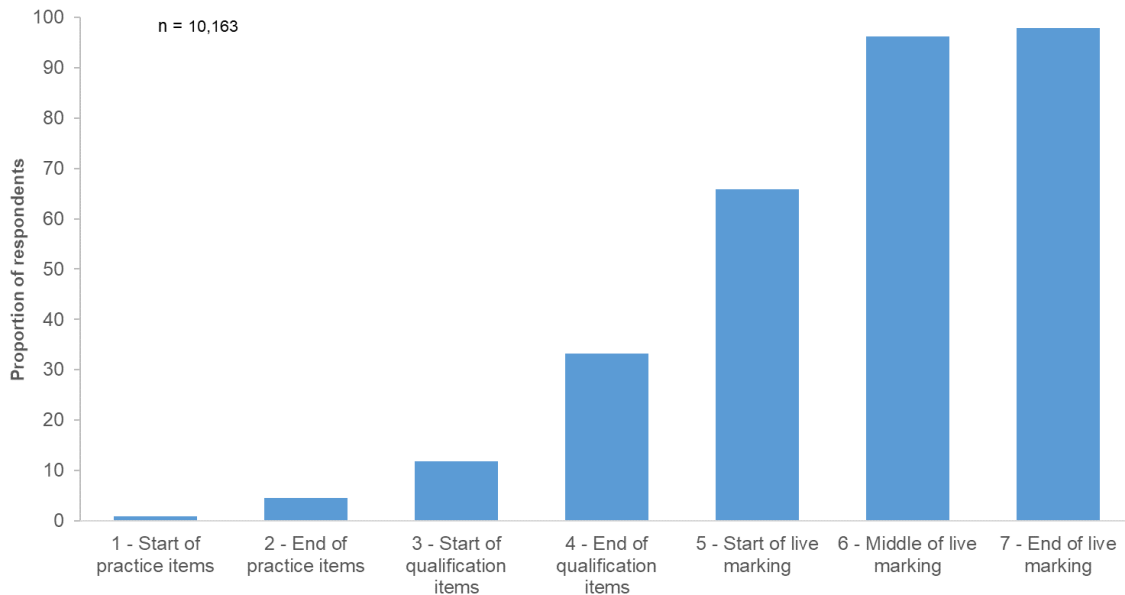


Figure 6. Proportion of examiners stating they felt confident at the end of each phase in the standardisation/marking process.

Given that many examiners only feel confident after the start of live marking, it may be that there should be more items/scripts in the practice phase. However, the fact that examiners do not feel confident in their marking does not necessarily mean that their marking is inaccurate as these two variables can be uncorrelated or negatively correlated (Nadas & Suto, 2008). Moreover, Pinot de Moira, Massey, Baird & Morrissy (2002) and Leckie & Baird (2011) found very little change in the accuracy of marking over the marking period.

In the survey we asked examiners how many practice items/scripts they completed and also how many items/scripts they thought there should be. More than a third of examiners said that there should be more practice items/scripts than there are at the moment.

*Respondent 3736, female, GCSE, biology*

We need many more scripts within on-line standardisation that better represent the range of answers that students give; many, many responses from students are vague and unclear and we need more practice in discerning whether they are mark-worthy or not. Most students don't write in the words of the mark scheme, and similarly, attributing their answer to 'student speak' (ie giving them benefit of the doubt) is not enough either.

*Respondent 3434, female, GCSE, physics*

More scripts to give a wider range of possible problems that you encounter during live marking. Possibly mark 10 scripts then get feedback, discuss problems with team leaders and mark another 10. I'd feel more confident I was marking accurately if I was exposed to more actual scripts before I started marking live.

In addition, where practice materials were based on whole scripts, getting the right coverage of different response types for each item might need more thought, though in available time scales this might sometimes be challenging.

*Respondent 9429, female, GCSE, chemistry*

We could have more practice scripts. The practice 'scripts' could be made from questions that actually come from a variety of scripts, in order to get a much broader range of response for each question.

This suggestion of 'pseudo candidates' is not dissimilar to the use of artificial exemplars which some have indicated might be helpful in speeding up the script selection process (Vijayathanan, Pinot de Moira & Stringer, 2016).

From the evidence presented in this section, this indicates that the number of exemplars for each question could be based on the complexity of the marking strategy, that the exemplars for each question could cover a larger range of quality and that the number of items/scripts included in the practice phase could be increased for some items as this could lead to an increase in examiners' confidence to mark.

## **5.4 Personal feedback**

Personal feedback is the feedback provided by the team leader to the examiner on a one-to-one basis. This feedback should be individualised based on the examiner's performance either in the practice or qualification phase.

### **5.4.1 Personal feedback provided after the practice phase**

In the survey we asked examiners whether they received personal feedback after they had completed their practice scripts. Figure 8 shows that 56% of examiners said that they had received personal feedback from their team leader after the practice phase and 44% of examiners who did not receive any personal feedback.



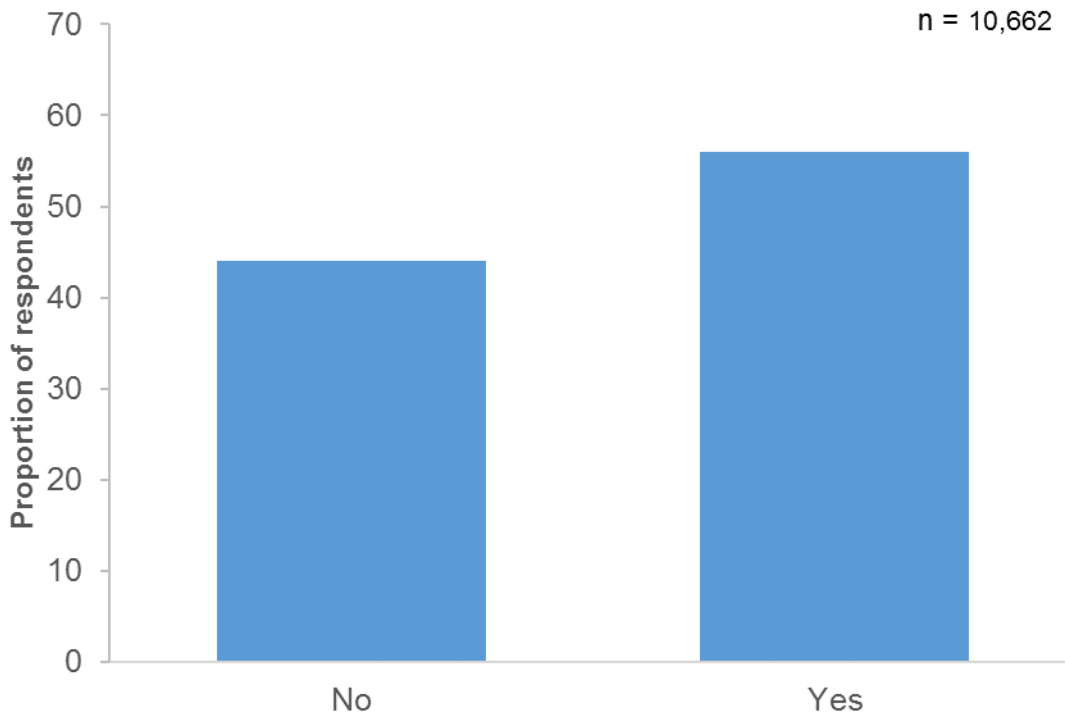


Figure 8. Proportion of examiners who received or did not receive personal feedback after they completed the practice phase.

For at least one exam board, we observed that the examiners could complete the practice items for a specific question and then immediately complete their qualification items without receiving feedback from their team leader, and that some prefer it this way. Below are some comments from the interviews:

*Examiner 2, female, GCSE, geography*

I could have clicked finish on the practice and gone straight into the [qualification] scripts, there's nothing to stop me. Lots of people do that.

*Examiner 7, female, GCSE, mathematics*

They suggest that you might do all the practice first, and then do the qualification afterwards. No!

*[interviewer]* So you do all the practice items for a question and then move to the qualification items for that question?

Well you're not supposed to do that. In theory you're supposed to do all the practice questions first. Then ring your team leader with any queries and then do qualification.

*[interviewer]* You're not convinced?

I don't think many people do it that way. Not when they've done it for a few years because it's clearly advantageous to go on and do qualification

while those practice questions are fresh in your mind, particularly as there's so much pressure on you. At one time you could get two of the 10 qualification ones wrong; now you can only get one wrong [in order to get the go ahead to mark that question].

*Examiner 7, female, GCSE, mathematics*

I mean personally I'm happier doing practice and qualification. If I really can't understand why there's a practice mark then I might not do that question for qualification I might wait and then accumulate all the queries from my questions and then ring the team leader up and then go through those queries and then do the qualification on those questions afterwards as a block.

A number of examiners from across all exam boards also gave their views on the need to receive feedback between practice and qualification phases in the survey.

*Respondent 705, female, GCSE, biology*

Some way of getting queries answered before the [qualification] scripts are attempted would be useful. I find I'm often using the [qualification] script to see if I'm interpreting the mark scheme correctly, when really it should have been made clear before that stage.

*Respondent 9068, female, GCSE, mathematics*

Team leaders should phone markers after doing practice scripts to verbally go through any queries.

*Respondent 11486, female, GCSE, biology*

Possibly to allow for compulsory feedback between practice and qualification

Based on the evidence above, this suggests it would be beneficial to many examiners to receive timely feedback from their team leader before they can access the scripts/items from the qualification phase. It might be that the role of the team leaders has to be reviewed so that they have more time available to deliver feedback to all the assistant examiners they supervise.

#### **5.4.2 Personal feedback provided after the qualification phase**

We have previously covered why personal feedback should be given on the practice phase. It is also important for the team leaders to provide personal feedback on the qualification scripts/items. Given the fact that some boards do not have a practice

phase, it is essential that personal feedback is provided before the start of live marking.

In the survey, we asked examiners whether they had received personal feedback after their qualification scripts. As can be seen in Figure 9, most examiners (90%) received personal feedback from their team leader after the qualification phase and before the start of live marking. However, 10% of examiners did not receive any kind of personal feedback. Also, 8% of examiners stated that they did not receive personal feedback after either of these 2 phases.

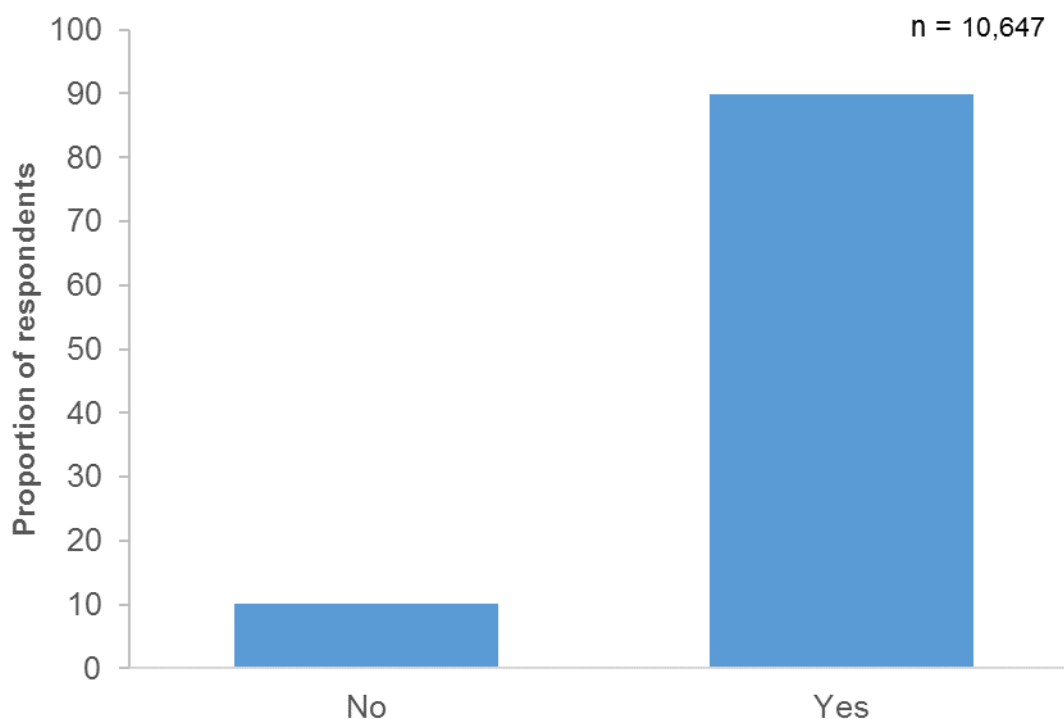


Figure 9. Percentage of examiners who said they had/had not received personal feedback after the qualification phase.

In the survey, some examiners commented on the fact that personal feedback should be provided after qualification.

*Respondent 9147, male, GCSE, geography*

The team leader should have to phone you to discuss the marking of the practice and [qualification] scripts.

*Respondent 11236, male, GCSE, physics*

Once the [qualification] is complete there should be feedback from the team leader.

Based on the evidence available, personal feedback should be available after the qualification phase to ensure that all examiners receive feedback before they start their live marking.

### 5.4.3 Content of personal feedback

Feedback is most effective when it is individual to the recipient (Ahmed et al., 2012). and relates to specific aspects of the examiners performance (Knoch, Read & von Randow, 2007).

Some respondents to the survey made similar observations:

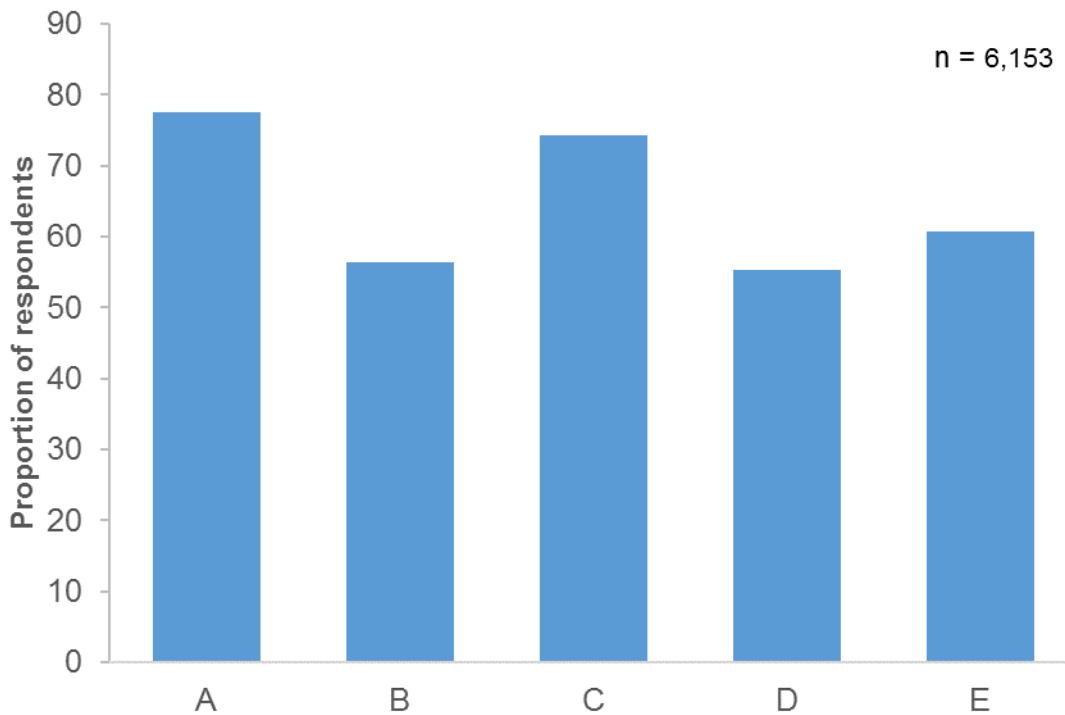
*Respondent 4350, male, A level, geography*

The feedback should be simpler and more personal to the examiner and listen to the concern or thinking of examiner and align it with the principal examiner's thinking.

*Respondent 7379, female, International GCSE, English Language*

There should be more specific feedback from team leaders particularly on reasons for awarding marks for writing. The feedback is always just a verbatim copy from the mark scheme which doesn't always address the issue.

In the survey, we asked examiners whether the personal feedback they received included any of the 5 categories of feedback listed below. As can be seen in Figure 10, the most frequent category was linked to general comments about the quality of their marking (78%) while reference to an individual candidate's response was the least frequent (55%). Sixty-one percent said that the feedback linked the mark scheme and the response.



- A. General comments about the quality of your marking (eg 'your marking is fine')
- B. Trends/patterns in the quality of your marking (eg 'you have a tendency to be a little bit generous in question 4')
- C. Reference to detailed aspects of the mark scheme
- D. Reference to individual candidate's response
- E. Linking between the mark scheme and particular aspects of a candidate's response

Figure 10. Proportion of examiners who said they received each type of feedback during the practice phase.

Based on the evidence presented in this section, there would be scope for an improvement in the personalisation of feedback and the practice of explaining the link between the mark scheme and the candidate's response. This could be tied into the team leaders' training which will be discussed later.

#### 5.4.4 Feedback delivery mode

In the survey, we asked examiners who said they had received personal feedback how they received feedback from their team leader after the practice phase. Overall, examiners have said that the phone was the most frequent way of receiving feedback (62%) followed by email (52%), through the marking system (33%), by text (6%) and finally other methods (2%) (Figure 11).

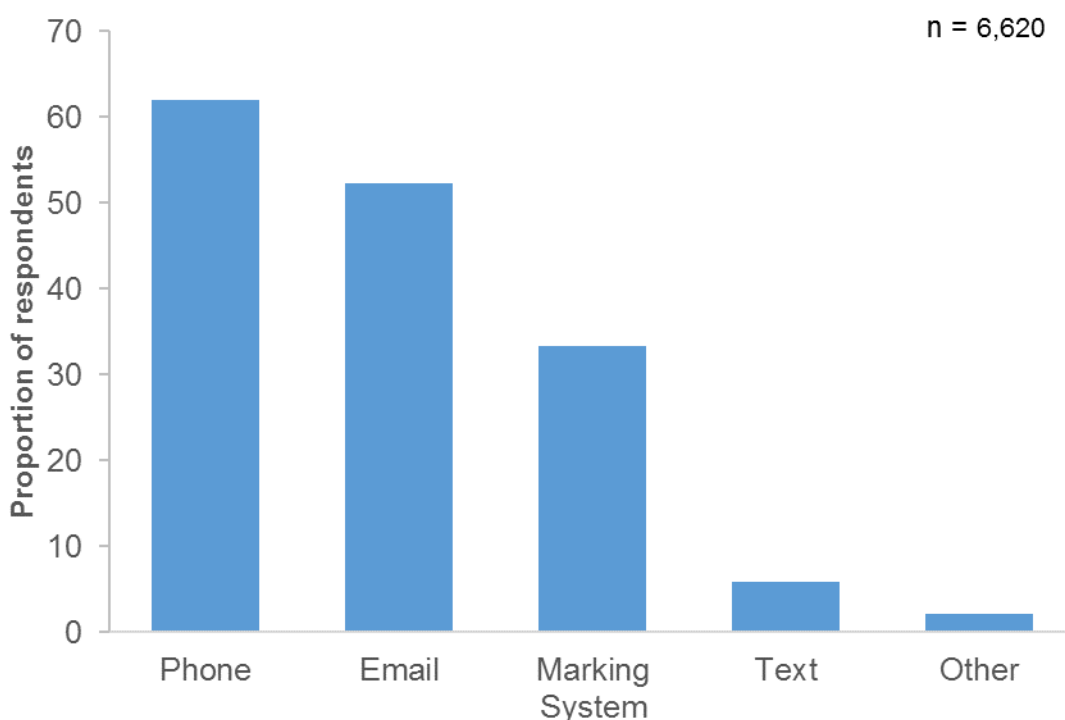


Figure 11. Proportion of examiners who were contacted by their team leader through each of these medium.

The use of video communication to provide personal feedback seems to be extremely rare although based on the literature it could be a very supportive way of providing feedback (Stannard, 2008). However, the fact that phone conversations take place for more than 60% of examiners can be seen as very positive, as phone calls can adequately support discussions around the exemplars.

*Respondent 6464, male, GCSE, ICT*

A meeting via Skype could help the supervisor to explain in more detail why a mark is being given or not being given and also give the opportunity for feedback.

*Respondent 6210, female, International GCSE, physics*

It may be [okay] for very experienced markers to operate by email only, but my experience suggests that phone calls are a much better way of gaining understanding and reaching consensus before live marking begins... calls can be made for no cost at all through WhatsApp, Skype and Viber, and Team Leaders should be encouraged to look at these options.

More than 50% of conversations take place via email, which is more susceptible to misunderstanding.

*Respondent 4093, female, GCSE, history*

There should be conversations by phone not e-mail as e-mail is not dialogue

*Examiner 8, female, GCSE, mathematics [talking about feedback by email as a team leader]*

That's a lot of emails going backwards and forwards and a lot of misunderstandings.

And a lot of wrong things and people don't like to get things wrong. So you explain something to them by email and they misinterpret it. And then they get it wrong and then you've got to tell them that they're wrong again.

Given the evidence reviewed here, the use of a mode of communication allowing for instant interaction (phone or video) should be favoured as they allow for discussions around the exemplars and can help the examiners' learning of the application of the mark scheme.

#### **5.4.5 Timing of feedback**

As outlined in the literature review, there seems to be some advantages to receiving feedback quickly after submission of the practice or qualification scripts (Johnson & Black, 2012; Rovai, 2000). This suggests that the quicker the feedback is delivered, the more powerful the feedback will be and the more positive examiners will feel. Making sure that all examiners have access to feedback within 24 hours of submitting their scripts/items would be helpful as it may be difficult for examiners to remember how they have applied the mark scheme to specific responses when feedback is delayed.

In the survey, we asked examiners who had received personal feedback how long after submitting their scripts/items did they receive feedback. Most examiners (59%) receive feedback within 6 hours of submitting their practice scripts. A further 26% of examiners receive feedback between 6 and 24 hours after submitting their practice scripts. Finally, around 15% of examiners have to wait more than 24 hours before receiving feedback (Figure 12).

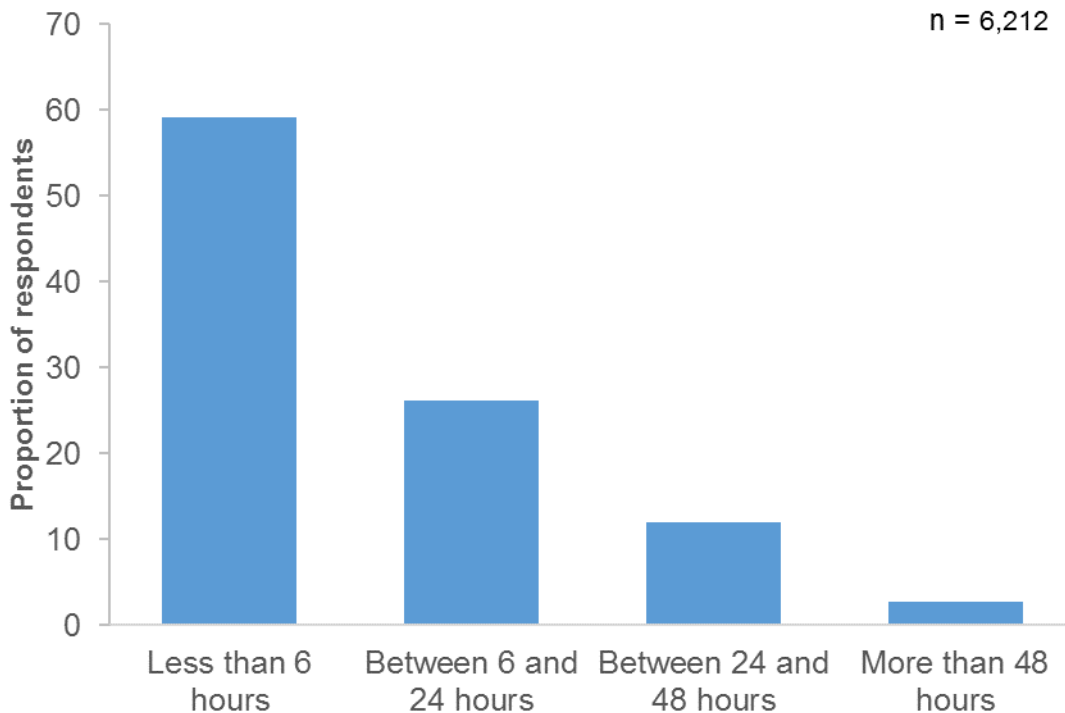


Figure 12. Time period for receiving feedback, for those examiners who received personal feedback.

Many respondents to the survey made the observation that prompt feedback would be helpful.

*Respondent 10269, female, A level, mathematics*

The team leader needs to promptly come back with answers to questions about items from the practice or qualification sets that the examiner considers as incorrectly marked on the system.

*Respondent 4226, female, A level, government and politics*

Make team leaders/supervisors available throughout the whole time of standardisation so questions and feedback can be dealt with - I had to wait 2 days before I received feedback and could be cleared, because my Team Leader was away.

Given the previous research, it might be important to aim for the feedback to be given within 24 hours of submission, so that examiners can clearly remember what their rationale was at the time they marked the item. However, in practice, when team leaders have large teams, it might be difficult for timely feedback to be provided for all team members.



### 5.4.6 Training on providing/receiving feedback

#### a) Providing feedback

As mentioned earlier, Johnson and Black (2012) identified 2 types of team leaders: evaluators and comparators. The distinction between these 2 types of team leaders could help define the training required by team leaders in order to provide the most effective feedback. Johnson (2016) suggested that team leaders could benefit from training on how to deliver feedback to the examiners they supervise. In a different context, Grant (2007) showed that a 2-day coaching skills training programme could improve coaching skills and hence feedback delivery. To our knowledge, team leaders do not receive specific training on how to deliver feedback.

*Examiner 6, female, GCSE, geography*

[interviewer] Any suggestions on how to improve the process?

Maybe teaching the effective feedback?

[interviewer] Do you mean to team leaders?

Yes team leaders, principal examiners. I'm sure they've probably already had it, I don't know. How to give and maybe how to receive feedback, not just the team leaders, but examiners, so they do a course on how to give and take feedback.

Survey respondents made many observations around the feedback received and the interactions they had with their team leaders.

Some were very positive:

*Respondent 3345, female, A level, geography*

My team leader this year was amazing and we had a really detailed discussion of the mark scheme but I know other team leaders that are less "user friendly".

*Respondent 7008, female, International A level, psychology*

Making sure all team leaders have a consistent approach to standardisation/supervision. As I said my team leader was excellent but on other units they can take days to get back with queries or provide minimal feedback.

Some mentioned the lack of interaction with their team leader when asked to briefly describe their suggested changes to the system:

*Respondent 10092, female, GCSE, mathematics*

Team leaders who care more. They should constantly get in touch with their team, as this is why they get paid more. They should be available more often and help and advise with practice and qualification in order to apply the mark scheme correctly. I felt very alone this summer. I complained but no one got back in touch.

Some commented on the fact that there is variation in the level of support provided by different team leaders:

*Respondent 126, female, GCSE, English*

There needs to be some review of the team leader role. Over the years I have worked for many team leaders. Some have been excellent. However, others can take ages to get back to me when there is an issue and don't really appear to take their duties seriously enough. Sometimes, I have had to wait days for the team leader to get back to me and that puts pressure on the marking deadlines. There is no mechanism for an assistant marker to review the support they have received on a year by year basis.

*Examiner 4, female, GCSE, English language [talking about team leaders giving feedback]*

It depends on how good your team leader is. Some team leaders are very, very good at it; some team leaders are less good at it.

But I would say the level of contact varies. A colleague who's also a marker has had one email off her team leader. [Exam board b] never. I never speak to a team leader. I get emails, and it makes it, it actually makes the stress worse and makes you feel more isolated.

Given the evidence above, in order to increase examiners' confidence and marking ability, it might be useful for team leaders to receive training and/or information on how to give personalised feedback and to give positive feedback not only when things go wrong (Ahmed et al, 2012).

b) Receiving feedback

For assistant examiners, there may be scope to provide them with information about why feedback is important and why it is given. A number of examiners have mentioned feeling that their team leader was there to outline their errors instead of being there to help them align their views with the principal examiner's views. As above, this could be partially addressed by team leaders providing feedback on good marking, but it would also be helpful for assistant examiners to be given clearer information on what to expect from their team leader.

*Examiner 6, female, GCSE, geography*

I used to find it terribly demoralising. I think most teachers do, don't they? They don't like to be told they're doing things wrong. When I first started I thought I was no good at all. The first lot of feedback I got I felt like, well I thought there's no way I can do this, I've done so badly and I'm so awful. And then I understood that what they were trying to do was standardise us, not persecute me. I think I used to take it very personally. And this is something that all new examiners need to be given some training on, I think, that the PE or the team leader isn't out to get them as it were; that the aim of it is to standardise.

*Examiner 4, female, GCSE, English language [talking about team leaders]*

I don't see them as the enemy, I see them as supportive. But that's taken me a lot of years to get there.

This suggests it could be helpful if examiners are given more information about the aim of feedback in order to facilitate discussions with their team leader.

## 5.5 Qualification phase

### 5.5.1 Second set of qualification scripts

Once the examiners have marked their qualification scripts/items, their team leader will look at their performance and decide whether their performance is close enough to the definitive marks for them to start live marking. The tolerance and tolerance rules (eg how many scripts/items out of tolerance) are different for different subjects and different exam boards. When examiners are deemed to be “out of tolerance” on too many items/scripts they may be asked to complete a second set of qualification items/scripts. There are some differences in if or under what conditions a second set of qualification scripts might be required.

Below are some comments from the interviews:

*Examiner 6, female, GCSE, geography*

...on this system there is a second [qualification] set that you can mark, it depends on how out you are and they can give you the same ones back again and say well correct them in line with my feedback, and as long as you correct them in line with their feedback and you're good enough, you're let through.

There are more [qualification] scripts available, but I've never been asked to do additional [qualification] scripts for anything.

*Examiner 5, male, GCSE, English language*

*[speaking as a team leader]*

Well you would take a second lot of five if you felt there were real problems. But only if you felt there were real problems after you'd had a chat with them.

The second quote indicates that in some instances qualification can take place on a relatively small number of items (eg 5), and that even if performance is not good on these, examiners may be cleared to mark if the examiner is satisfied after a discussion.

It seems likely that having scope for examiners to undertake a second set of qualification materials, and clear guidance around this, could be useful in ensuring that those cleared to mark have met the required standard and have had greater scope for more feedback.

## 5.6 Other relevant findings

Some additional questions of the survey are relevant to understanding the effectiveness of online standardisation in promoting the required learning, marking consistency and giving examiners sufficient motivation and satisfaction to continue in the role.

In this section, we will briefly outline these other relevant findings.

### 5.6.1 Importance of practice and qualification phase

In the survey, we asked examiners how important they thought the practice and qualification scripts/items were for giving them sufficient understanding of how to apply the mark scheme accurately.

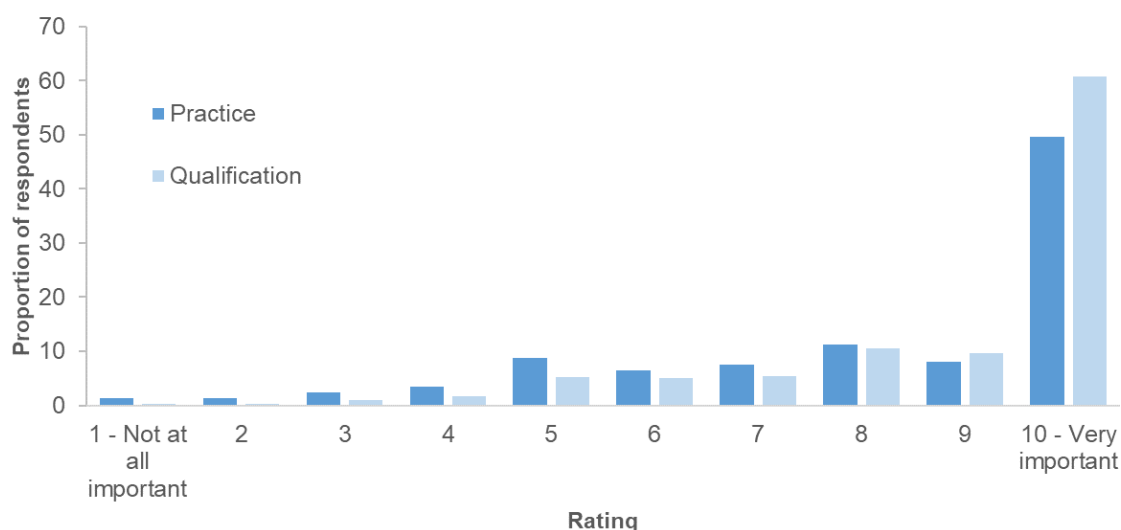


Figure 13. Examiners' perception of the importance of practice and qualification items/script in understanding how to apply the mark scheme.

As seen in Figure 13, half of the examiners stated that the practice scripts/items were very important for giving them sufficient understanding to apply the mark scheme accurately (providing a rating of 10, where one is not at all important and 10 is very important). Twenty-four percent of examiners gave a rating of one to 6 while 27% gave it a rating of between 7 and 9.

For the qualification phase, more than 60% of examiners said that the qualification scripts/items were very important (again, a rating of 10) in giving them sufficient understanding to apply the marks scheme accurately whilst only 14% gave a rating of between one and 6.

The comparison between these 2 figures is interesting in itself as it shows that examiners think that the qualification phase is more important than the practice phase for learning to apply the mark scheme. When looking at individual examiners

we find that 57% of them think that both phases are as important (same rating to both questions), 31% think that the qualification phase is more important while only 13% think that the practice phase is more important. This is at odds with the view that the practice phase is the main learning phase in the online standardisation process. This could be due to the fact that not all exam boards include rationale and annotations in the practice phase or provide personal feedback for example in the form of a telephone call from the team leader.

### 5.6.2 Examiners' satisfaction

Given what was said previously about the link between examiners' satisfaction and their willingness to continue examining, we also asked the examiners about their satisfaction of the online standardisation process.

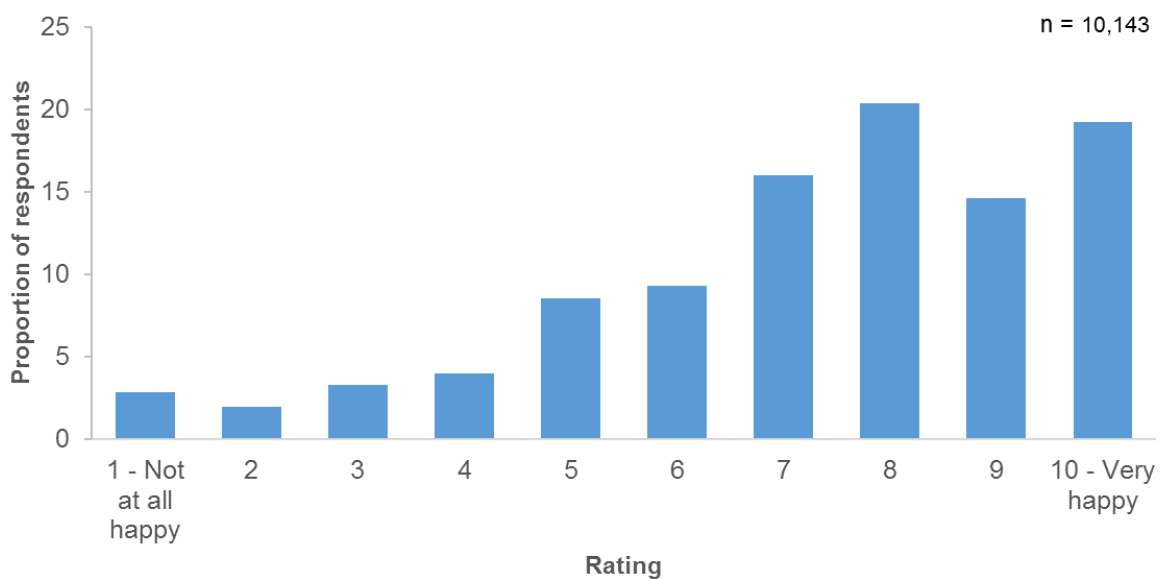


Figure 14. Examiners' level of satisfaction with the online standardisation process.

Thirty percent of examiners gave a rating between one and 6 where one is not happy at all and 10 is very happy with the current online standardisation system. Fifty-one percent gave a rating between 7 and 9 while 19% said they were very happy with the current online standardisation system (Figure 14).

We also asked examiners whether they had taken part in a face-to-face standardisation meeting in the past. The interaction of these 2 questions is shown in Figure 15. Examiners who had previously attended a face-to-face standardisation meeting showed lower levels of satisfaction than the examiners who had never attended such a meeting. For those who had previously attended a meeting, 62% reported a level of satisfaction above 6 compared to 78% of examiners who had not attended a meeting.

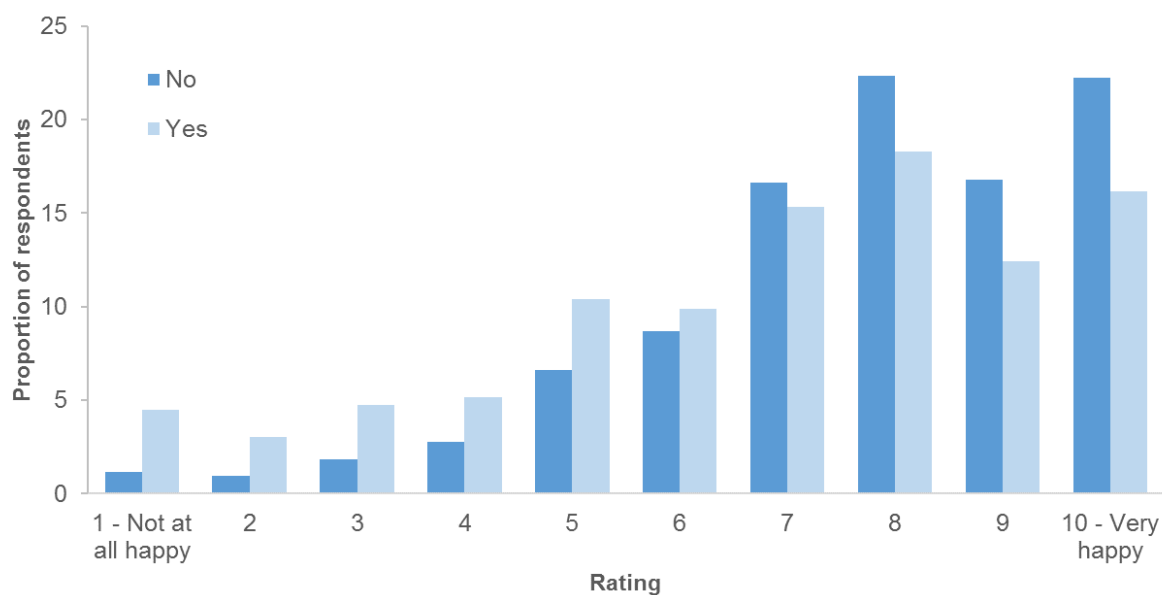


Figure 15. Examiners' level of satisfaction with the online standardisation process as a function of whether or not they have attended a face-to-face standardisation meeting in the past.

In the context of examiners' retention, it might be important to identify whether some of the factors discussed in this report could improve the satisfaction of examiners who have previously taken part in face-to-face standardisation meetings.

### 5.6.3 Perceived advantages and disadvantages of online standardisation

In the survey, the examiners were asked to select all the advantages and disadvantages presented in a list that they thought applied (for example, they could select more than one). The 2 principal advantages were related to the fact that they could choose when to do the practice scripts/items (70%) and that there is no travel required (63%).

On the other hand, examiners thought that the main 2 disadvantages were the fact that they did not have a chance to meet with their team leader (52%) and that there was no in-depth discussion about the mark scheme (48%). The latter point could possibly be improved by an increase in the frequency and quality of the feedback provided although more research would be required in order to monitor whether these variables have an impact on the examiners satisfaction.

## 6 Discussion

The main aim of this project was to investigate how online standardisation is implemented across the exam boards, how it is experienced and perceived by examiners and where there may be opportunities to increase its effectiveness. The proposed process map (page 15) pulls together a number of "good practices"

observed across the exam boards as well as some rooted in the feedback and learning literature. We observed that these “good practices” were sometimes implemented inconsistently within and between boards. Part of this inconsistency is because setting up online standardisation materials for any particular exam (of which there are around 18,000 each year) is necessarily completed in a very tight timeframe of typically one or two days. During this period, senior examiners have to select suitable standardisation scripts/items and arrive at definitive marks. They may also need to amend the mark scheme and provide ‘definitive’ annotation and/or a rationale for the marks awarded.

Contrary to the survey carried out in 2014 looking at the quality of marking (Ofqual, 2014), examiners in the current survey were fairly positive about the online standardisation process. However, almost half of them also mentioned that not having a chance to meet their team leader was a disadvantage and that the process could lead to less in-depth discussions around the mark scheme. On the other hand, most examiners also said that being able to carry out the practice scripts/items when they wanted and that not having to travel to standardisation meetings were 2 important advantages of the process.

One of the important findings was that it is important for examiners to know not only if their marks are ‘correct’ (in that they awarded the same mark as the definitive mark) but also if they are awarded on the same basis as was intended. Annotation, written rationales of definitive marks and personalised feedback can be helpful. However, less than half of examiners said they could view clear rationales for how marks were awarded. Moreover, the annotation/rationales were not always viewed: examiners were likely only to view them if their marks disagreed with the definitive mark, despite the possibility that their marks had been awarded on the wrong basis.

Personalised feedback after the practice phase occurred around half of the time and it was more likely to occur after the qualification phase. Feedback was given mostly by telephone or electronically but very few team leaders seem to use video to deliver feedback which could potentially be a useful mode of communication to use in order to support discussion. Most feedback was received within 6 hours of completing the practice items/scripts which is positive and can contribute to reducing the isolation examiners sometimes mentioned.

Examiners tended to believe that the qualification phase was more important than the practice phase in helping them learn to apply the mark scheme correctly. This could be due to board and subject differences in implementing the different phases of online standardisation, such that the practice phase items sometimes do not have annotations, rationale and definitive marks. It is easy to see how, in those circumstances, examiners would think that the qualification phase was the important phase.

While confidence is not always associated with competence, a third of examiners reported not feeling confident in their marking at the point at which they started live marking.

It is not possible from this study to know whether this is higher or lower than levels of confidence where there has been face-to-face or another type of standardisation. Nevertheless, it may indicate scope for more learning and/or confidence building prior to live marking.



Finally, it would be important to gather data in order to verify whether the “good practices” presented here lead to an increase in accuracy or satisfaction.

## **7 Conclusion**

On balance, examiners are positive about the online standardisation processes. There are differences between boards and subjects about the exact phases and the exact form online standardisation might take. However, good practices regarding personal feedback, annotation, use of mark rationales were observed and reported.

## References

- Ahmed, M., Sevdalis, N., Paige, J., Paragi-Gururaja, R., Nestel, D. & Arora, S. (2012). Identifying best practice guidelines for debriefing in surgery: a tri-continental study. *The American Journal of Surgery*, 203, 523–529.  
<http://doi.org/10.1016/j.amjsurg.2011.09.024>
- Baird, J.-A., Grooten, J. & Bell, J.F. (2004). What makes marking reliable? Experiments with UK examinations. *Assessment in Education*, 11, 333-347.  
<http://doi.org/10.1080/0969594042000304627>
- Baird, J.-A., Meadows, M., Leckie, G. & Caro, D. (2015) Rater accuracy and training group effects in Expert- and Supervisor-based monitoring systems, *Assessment in Education: Principles, Policy & Practice*, 24:1, 44-59.  
<http://doi.org/10.1080/0969594X.2015.1108283>
- Billington, L. & Davenport, C. (2008). *Online standardisation trial, Winter 2008: Evaluation of examiner performance and examiner satisfaction*. Manchester: AQA Centre for Education Research and Policy. Retrieved from  
<https://cerp.aqa.org.uk/research-library/online-standardisation-trial-winter-2008-evaluation-examiner-performance-and-examiner-satisfaction>
- Bramley, T. (2008). *Mark scheme features associated with different levels of marker agreement*. Paper presented at the annual conference of the British Educational Research Association, September 3-6, Edinburgh, UK. Retrieved from  
<http://www.cambridgeassessment.org.uk/Images/109770-mark-scheme-features-associated-with-different-levels-of-marker-agreement.pdf>
- Bull, P. (2001). State of the art non-verbal communication. *The Psychologist*, 14, 644-647. Retrieved from <https://thepsychologist.bps.org.uk/volume-14/edition-12/state-art-nonverbal-communication>
- Chamberlain, S. and Taylor, R. (2011). Online or face-to-face? An experimental study of examiner training. *British Journal of Educational Technology*, 42, 665–675.  
<http://doi.org/10.1111/j.1467-8535.2010.01062.x>
- Coninx, N.S., Kreijns, C.J. & Jochems, W.M.G. (2013). The use of keywords for delivering immediate performance feedback on teacher competence development. *European Journal of Teacher Education*, 36, 164-182.  
<http://doi.org/10.1080/02619768.2012.717613>
- Daft, R.L. & Lengel, R.H. (1986). Organizational information requirements, media richness and structural design. *Management Science* 32, 554-571.  
<https://doi.org/10.1287/mnsc.32.5.554>
- Elder, C., Knoch, U., Barkhuizen, G., & von Randow, J. (2005). Individual feedback to enhance rater training: Does it work? *Language Assessment Quarterly*, 2, 175-196.  
[http://doi.org/10.1207/s15434311laq0203\\_1](http://doi.org/10.1207/s15434311laq0203_1)
- Fearnley, A. (2005). *An Investigation of Targeted Double Marking for GCSE and GCE*. London: QCA. Retrieved from <http://dera.ioe.ac.uk/id/eprint/9450>

- Grant, A.M. (2007). Enhancing coaching skills and emotional intelligence through training. *Industrial and Commercial Training*, 39, 257-266. <http://doi.org/10.1108/00197850710761945>
- Handley, K., den Outer, B., & Price, M. (2013). Learning to mark: Exemplars, dialogue and participation in assessment communities. *Higher Education Research and Development*, 32, 888-900. <http://doi.org/10.1080/07294360.2013.806438>
- Johnson, M. (2015). Articulation work: Insights into examiners' expertise from their remote feedback interactions. *Communication & Language at Work*, 4, 28-52. <http://doi.org/10.7146/claw.v1i4.20771>
- Johnson, M. (2016). Feedback effectiveness in professional learning contexts. *Review of Education*, 4, 195–229. <http://doi.org/10.1002/rev3.3061>
- Johnson, M. & Black, B. (2012). Feedback as scaffolding: senior examiner monitoring processes and their effects on examiner marking. *Research in Post-Compulsory Education*, 17, 391-407. <http://doi.org/10.1080/13596748.2012.738965>
- Johnson, M. & Black, B. (2014). *How do examiners align their marking judgements remotely? Insights into examiners' feedback interactions for remote standardisation*. Paper presented at annual conference of AEA Europe, 6-8 November, in Tallinn, Estonia.
- Knoch, U., Read, J. & von Randow, J. (2007). Re-training writing raters online: how does it compare with face-to-face training? *Assessing Writing*, 12, 26-43. <http://doi.org/10.1016/j.asw.2007.04.001>
- Kruger, J., Epley, N., Parker, J., & Ng, Z. (2005). Egocentrism over e-mail: Can we communicate as well as we think? *Journal of Personality and Social Psychology*, 89, 925-936. <https://doi.org/10.1037/0022-3514.89.6.925>
- Leckie, G., & Baird, J.-A. (2011). Rater effects on essay scoring: A multilevel analysis of severity drift, central tendency, and rater experience. *Journal of Educational Measurement*, 48, 399–418. <https://doi.org/10.1111/j.1745-3984.2011.00152.x>
- Mathieson, K. (2012). Exploring student perceptions of audiovisual feedback via screencasting in online courses. *American Journal of Distance Education*, 26, 143-156. <http://doi.org/10.1080/08923647.2012.689166>
- Meadows, M. (2004). *Measures of examiner job satisfaction as predictors of intention to continue examining*. Manchester: AQA Centre for Education Research and Policy. Retrieved from <https://cerp.aqa.org.uk/research-library/measures-examiner-job-satisfaction-predictors-intention-continue-examining>
- Nádas, R. & Suto, W. M. I. (2008). An exploration of self-confidence and insight into marking accuracy among GCSE maths and physics markers. *Research Matters: A Cambridge Assessment Publication*, 5, 9-15.
- Ofqual. (2014). *Review of quality of marking in exams in A Levels, GCSEs and other academic qualifications; Findings from survey of examiners, May 2013*. (Ofqual Report 14/5382). Coventry: Ofqual. Retrieved from <https://www.gov.uk/government/publications/survey-of-examiners-on-quality-of-marking-in-gcses-and-a-levels>
- Pinot de Moira, A. (2005). *Do examiner characteristics affect marking reliability?*

RPA\_05\_APM\_RP\_04. Manchester, UK: AQA Centre for Education Research and Practice.Internal.

Pinot de Moira, A., Massey, C., Baird, J.-A, & Morrissy, M. (2002). Marking Consistency over Time. *Research in Education*, 67, 79-87.  
<https://doi.org/10.7227/RIE.67.8>

Price, M. (2005). Assessment standards: The role of communities of practice and the scholarship of assessment. *Assessment & Evaluation in Higher Education*, 30, 215-230. <http://doi.org/10.1080/02602930500063793>

Rovai, A. P. (2000). Building and sustaining community in asynchronous learning networks. *Internet and Higher Education*, 3, 285–297. [http://doi.org/10.1016/S1096-7516\(01\)00037-9](http://doi.org/10.1016/S1096-7516(01)00037-9)

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13, 191–209. <http://dx.doi.org/10.1080/0305498870130207>

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30, 175-194.  
<http://dx.doi.org/10.1080/0260293042000264262>

Stannard, R. (2008). A New Direction in Feedback. *Humanising Language Teaching*, 10. Retrieved from <http://www.hlomag.co.uk/dec08/mart04.htm>

Stephenson & Cadwallader (2014). Trialling a system of enhanced team leader feedback for examiners. Manchester: AQA Centre for Education Research and Policy.

Suto W. M. I., Greatorex J. (2008). A quantitative analysis of cognitive strategy usage in the marking of two GCSE examinations. *Assessment in Education: Principles, Policy and Practice*, 15, 73–90. <http://doi.org/10.1080/09695940701876177>

Suto, W. I., Greatorex, J. & Nádas, R. (2009). Thinking about making the right mark: using cognitive strategy research to explore examiner training. *Research Matters: A Cambridge Assessment Publication*, 8, 23-32.

Vidal Rodeiro, C. (2007). Agreement between outcomes from different double marking models. *Research Matters: A Cambridge Assessment Publication*, 4, 28-34.

Vijayathanan, M., Pinot de Moira, A., & Stringer, N. (2016). *Judging the efficacy of artificial exemplar material*. Manchester: AQA Centre for Education Research and Practice. Retrieved from <https://cerp.aqa.org.uk/research-library/judging-efficacy-of-artificial-exemplar-material>

Weigle, S.C. (1998). Using FACETS to Model Rater Training Effects. *Language Testing*, 15, 263–287. <http://doi.org/10.1177/026553229801500205>

Weigle, S.C. (1999). Investigating Rater/Prompt Interactions in Writing Assessment: Quantitative and Qualitative Approaches. *Assessing Writing*, 6, 145–178.  
[https://doi.org/10.1016/S1075-2935\(00\)00010-6](https://doi.org/10.1016/S1075-2935(00)00010-6)

# Appendix A – Survey

## Survey methodology

The survey was sent out electronically between July and September 2016. We asked awarding organisations to email a hyperlink to all of their examiners who standardised online. The survey was open to all examiners who had carried out online standardisation in summer 2016, but only those who had marked GCSEs, AS or A level or their recognised equivalent were used in the analysis. This is because the observations and interviews were carried out with examiners who had marked paper included in those qualifications. Because some examiners marked more than one unit, we asked them to select one to answer the survey and list the level and subject of the unit.

The sample of respondents who completed the survey was fairly large compared to the population but because it involved a self-selection bias (examiners decided whether or not to respond) we cannot ascertain whether it is representative or not of the population of examiners. This report includes a summary of the quantitative responses alone. The survey did include some open response questions but these qualitative responses have been incorporated into the companion report on interviews and observations of examiners.

## Survey results

### Respondents' characteristics

#### **Respondents**

A total of 12,681 examiners completed the survey.

#### **Gender of respondents**

Of the 12,681 examiners who completed the survey, 64% were women and 36% were men and just under 1% preferred not to disclose their gender.

#### **Age of respondents**

Respondents were equally spread between 4 of the 5 age categories (between 20% and 24% of candidates in the last 4 categories while the 21 to 30 category had around 9%). This is due to the very small number of respondents in the 21 to 25 years (1.2%). This might be explained by the fact that teachers have to complete a one-year PGCE or on the job training after their degree as well as a one-year induction period when they start teaching. This means that by the time teachers are in a position to apply to be an assistant examiners, they are most probably in the 26- to 30-year-old category. However, given the sampling technique used here (self-selecting), we would not expect the sample to necessarily be representative of the population of examiners.

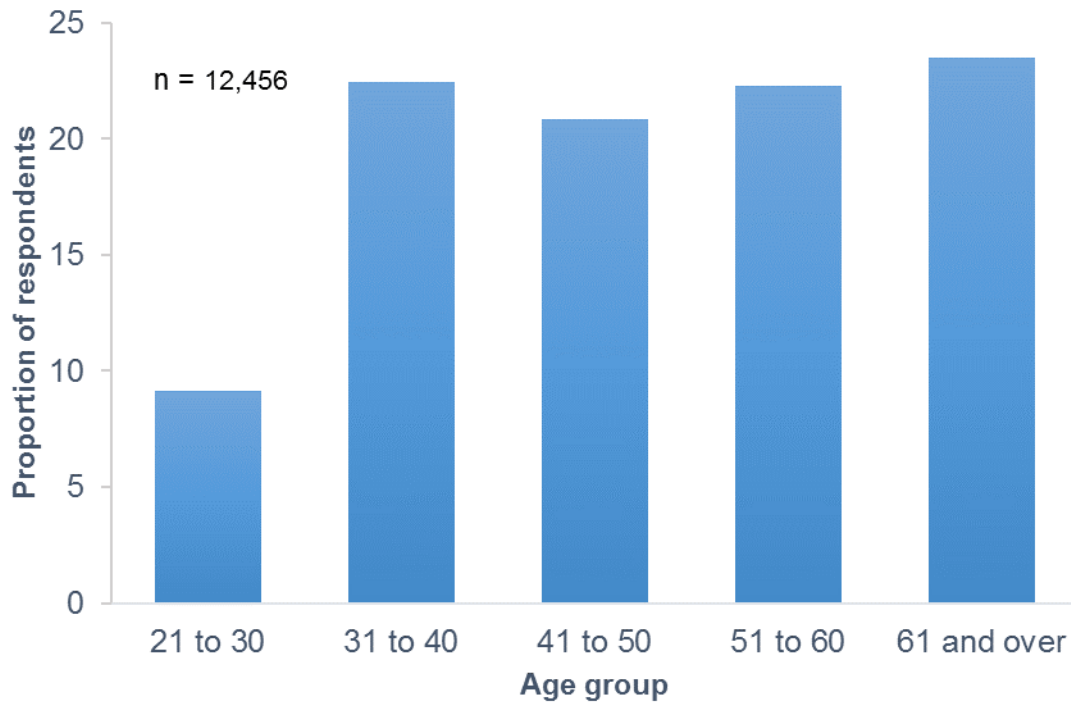


Figure 16. Proportion of respondents by age group

**Age and gender**

The proportion of male respondent gets larger in each age category whereas the proportion of female decreases as the age category increases except for a lower proportion of female respondents in the 21- to 30-year-old category.

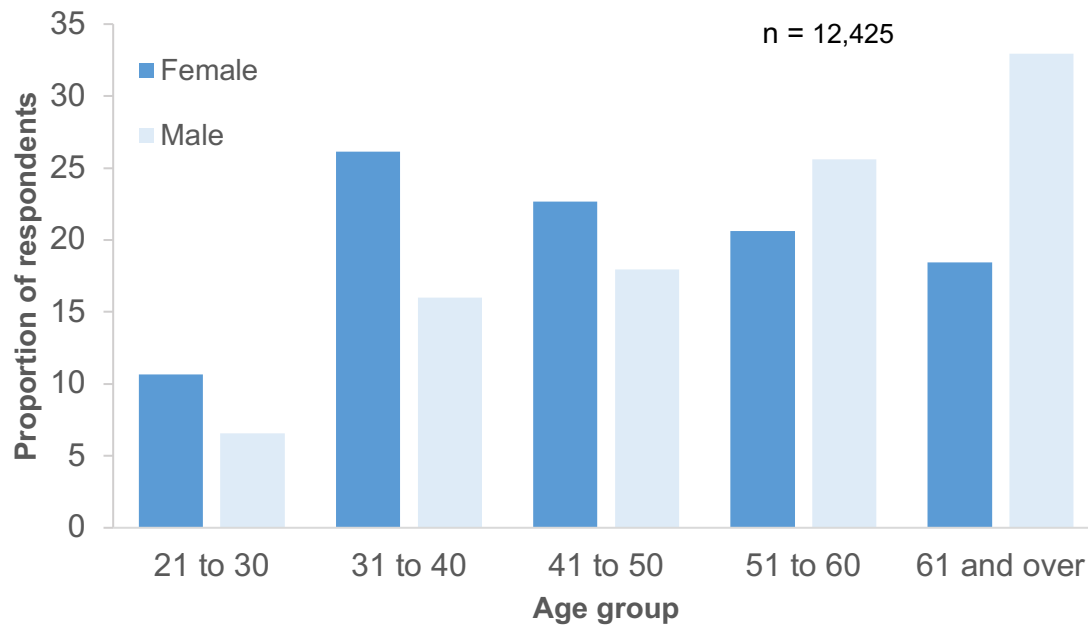


Figure 17. Proportion of respondents by age group according to gender.

**Qualification level**

Two-third of respondents marked GCSE qualifications or their equivalent while 6% marked AS level or equivalent and 27% marked A level or equivalent.

n = 12,681

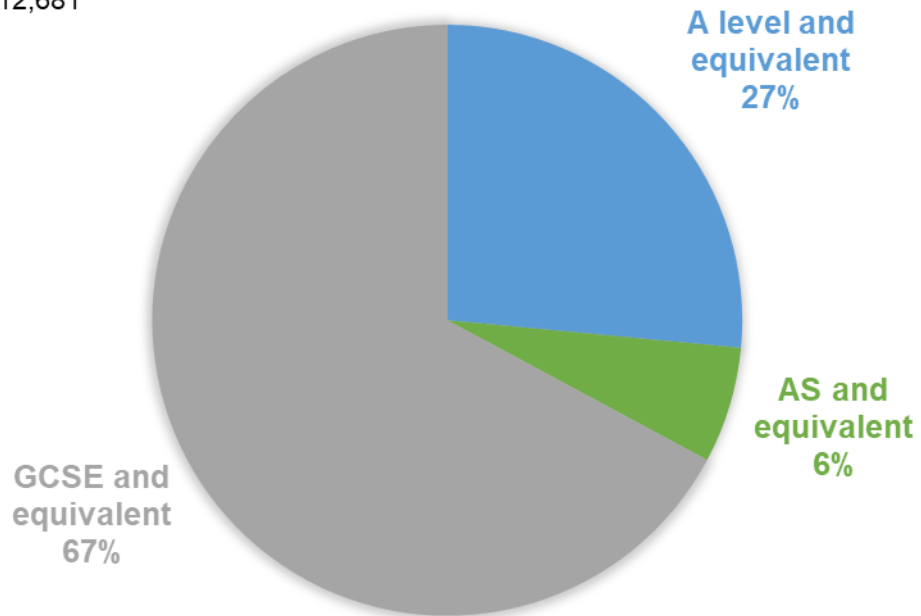


Figure 18. Proportion of respondents marking each qualification level.



**When was the first time you marked for an exam board?**

Around 38% of respondents have been examiners since before 2005 whereas around 26% have been examiners from 2015 or 2016. The sample of respondents was therefore broadly divided between those who have been examining for an extended period, and those who are relatively new to marking, with only a small number of examiners with an intermediate amount of experience

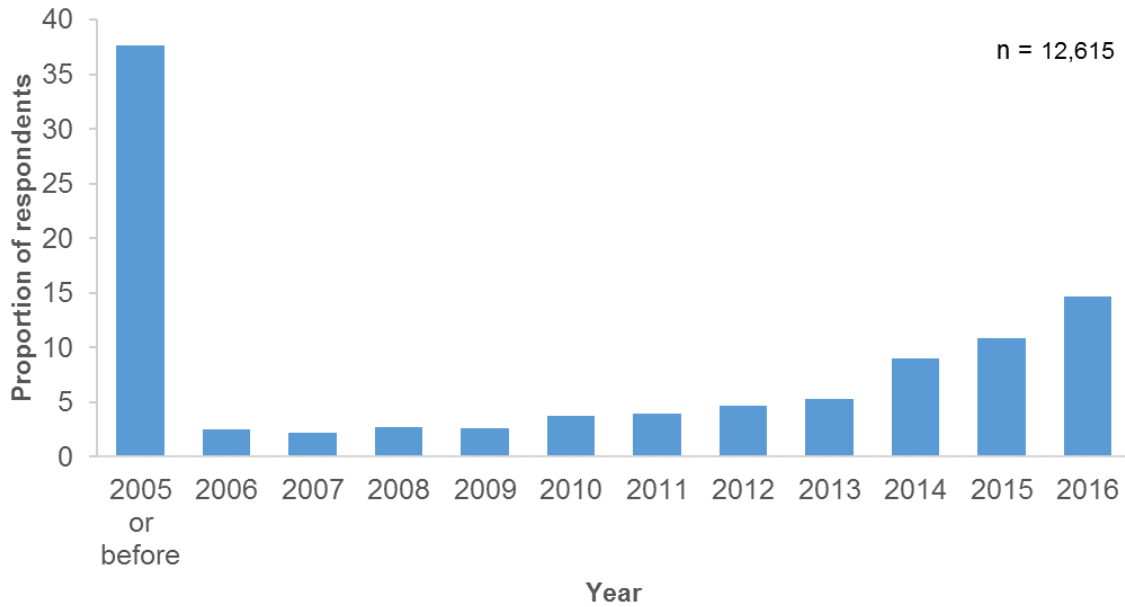


Figure 19. Proportion of respondents who started marking in each year.

In our sample men have marked for longer in general than women. More specifically, a larger proportion of men than women have started marking in 2005 or before (46% vs 33%) whereas larger proportions of women than men have started marking between 2015 and 2016.

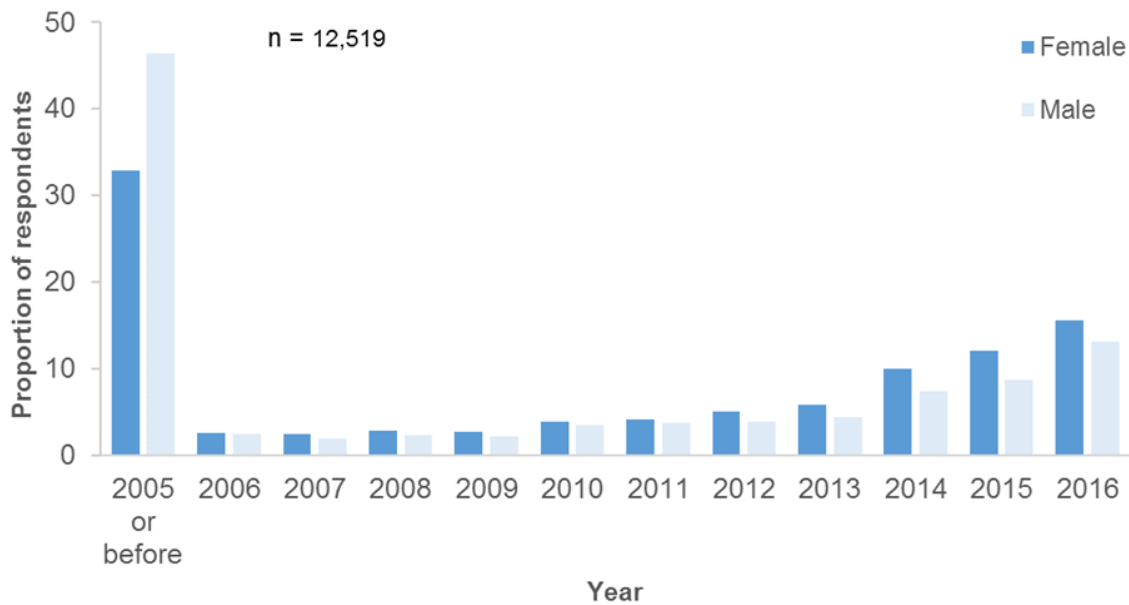


Figure 20. Proportion of respondents who started marking in each year according to gender.

**Did you attend a face-to-face standardisation meeting in the past?**

Overall, half of the respondents who completed the survey (50%) attended a face-to-face standardisation meeting in the past. (n = 12,624)

The proportion of men who attended a face-to-face standardisation meeting in the past is slightly higher than for women (54% vs 48%). This is probably due to the fact that male examiners have generally been marking for longer, before online standardisation was introduced.

### How many practice scripts/items did you have to complete?

The two most frequent responses were 10 practice scripts/items (30%) and 5 practice scripts (25%). The third most frequent answer was more than 10 (24%).

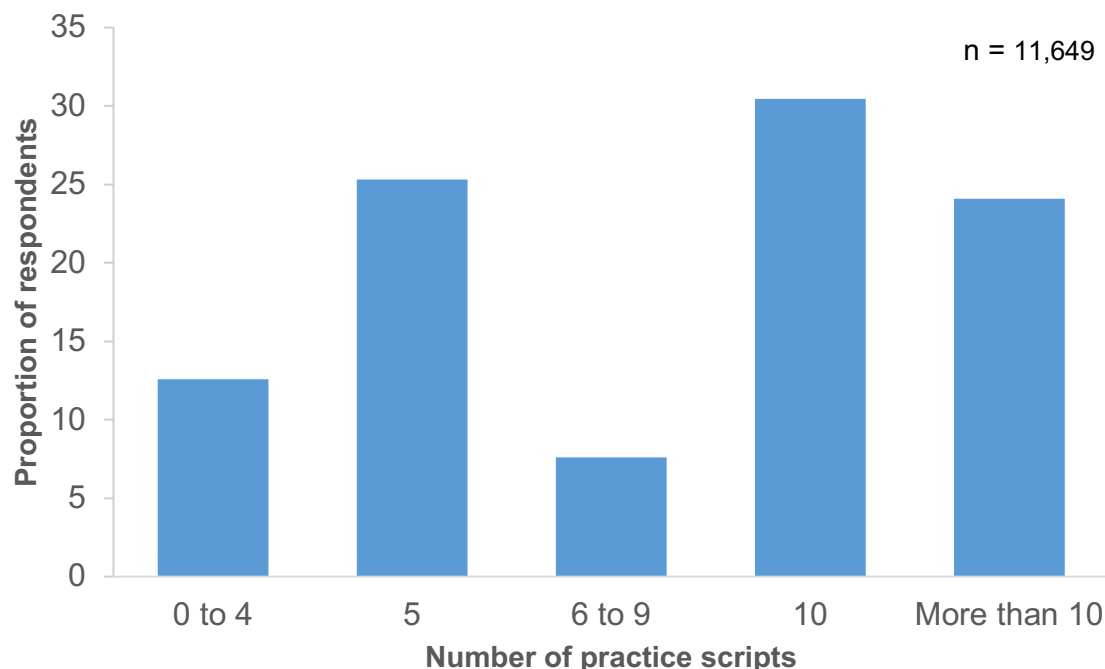


Figure 21. Proportion of respondents who completed different number of practice scripts/items.

Because the different awarding organisations have different methods of presenting material to examiners, it is possible that some respondents may have interpreted this question differently. For respondents who marked by scripts, the number they gave is the number of scripts or answer for each question. For examiners who mark by clips, the number they gave could have been the total number of clips and not the number of clips per question. To support this, the most frequent response from respondents marking for awarding organisations using clips was more than 10. The most frequent response from respondents marking by scripts was 10 for one awarding organisation and 5 for the other which is in line with the suggestion that the question may have been misunderstood.

### Practice phase

#### How important do you think the practice scripts are for giving you sufficient understanding to apply the mark scheme accurately?

Half of the examiners stated that the practice scripts/items are very important for giving them sufficient understanding to apply the mark scheme accurately. Twenty-four percent of examiners gave a rating of one to 6 where one is not at all important and 10 is very important. Finally, 27% gave it a rating of between 7 and 9.

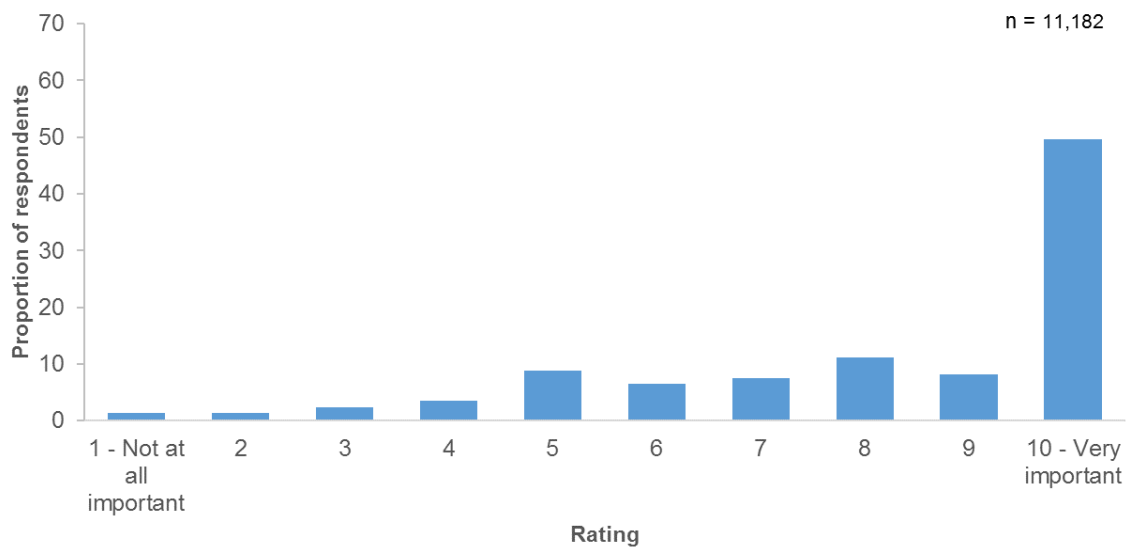


Figure 22. Proportion of respondents who gave different ratings of how important the practice scripts are at giving sufficient understanding to apply the mark scheme accurately.

**After you had marked the practice scripts, could you see written rationale (explanation on each question usually underneath the response) given for the definitive mark?**

Seventy percent of respondents said that they could see written rationale for all or most items. However, 17% of examiners said they could not see any written rationale.

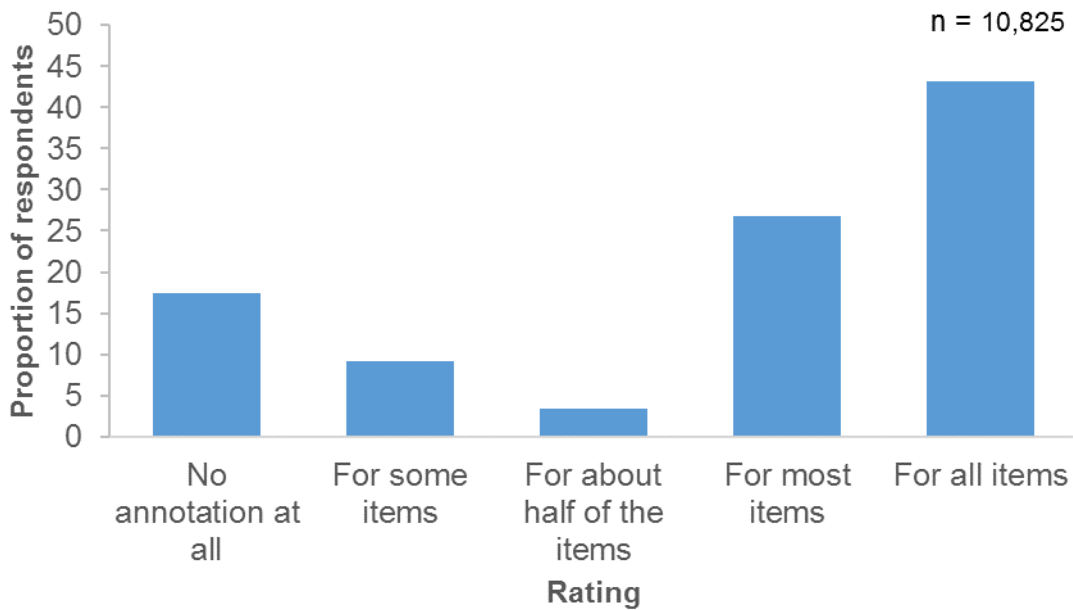


Figure 23. Proportion of respondents who could see different amount of written rationale during the practice phase.

**If written rationales were presented, to what extent were they clear and helpful in providing you with an understanding of where and how marks were awarded for individual responses?**

A quarter of respondents gave a rating of 10 (very clear and helpful), half gave a rating between 7 and 9 whilst a quarter gave a rating from one to 6 (where one is not at all clear and helpful).

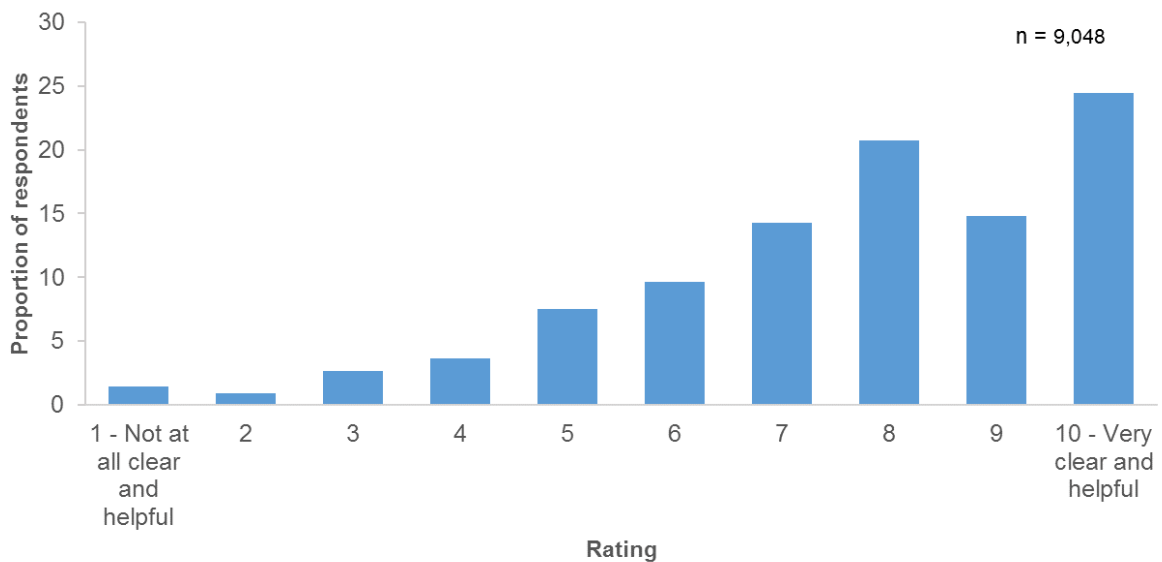


Figure 24. Proportion of respondents rating the extent to which rationale was clear and helpful in providing an understanding of where and how marks were awarded for individual responses.

**After you had marked the practice scripts, could you see the 'definitive' annotations, for example, indicating where credit had been given?**

Two-third of respondents stated that definitive annotations were available for either most items or all items. Twenty-one percent of respondents said that no annotations were available. Annotations appear to be included a little more frequently than written rationales.

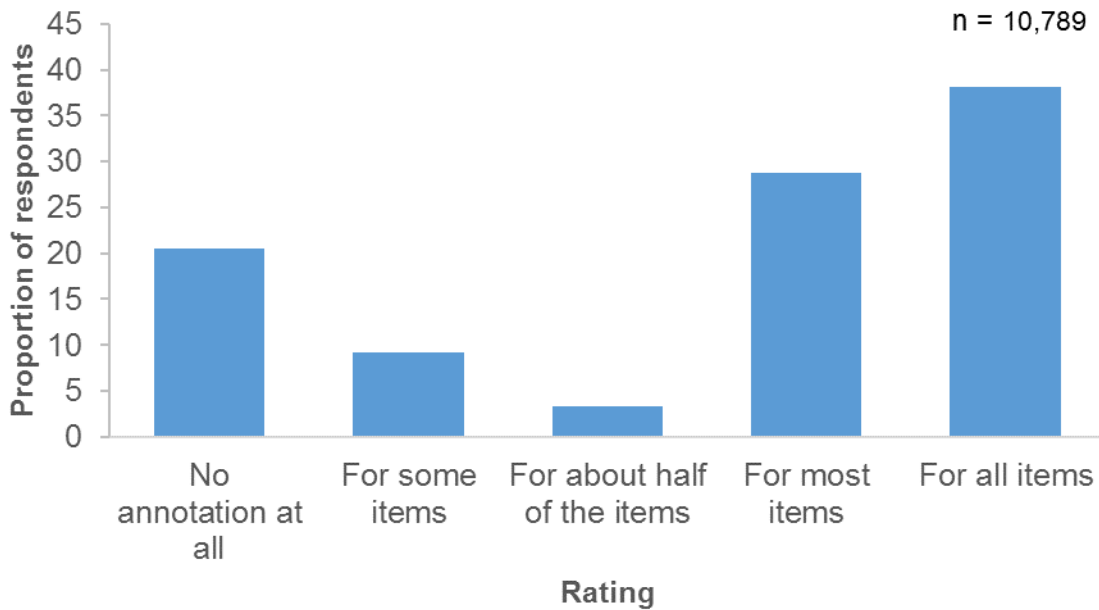


Figure 25. Proportion of respondents who could see different amount of annotations during the practice phase.

**To what extent were the annotations clear and helpful in providing you with an understanding of where and how marks were awarded for individual responses?**

For those examiners who said that there was some form of annotation available for the practice scripts, 24% thought that the annotations were very clear and helpful. A further 51% rated the quality of annotations between 7 and 9 while 25% rated the quality between one (not at all clear and helpful) and 6. These were very similar to the rated usefulness of written rationales.

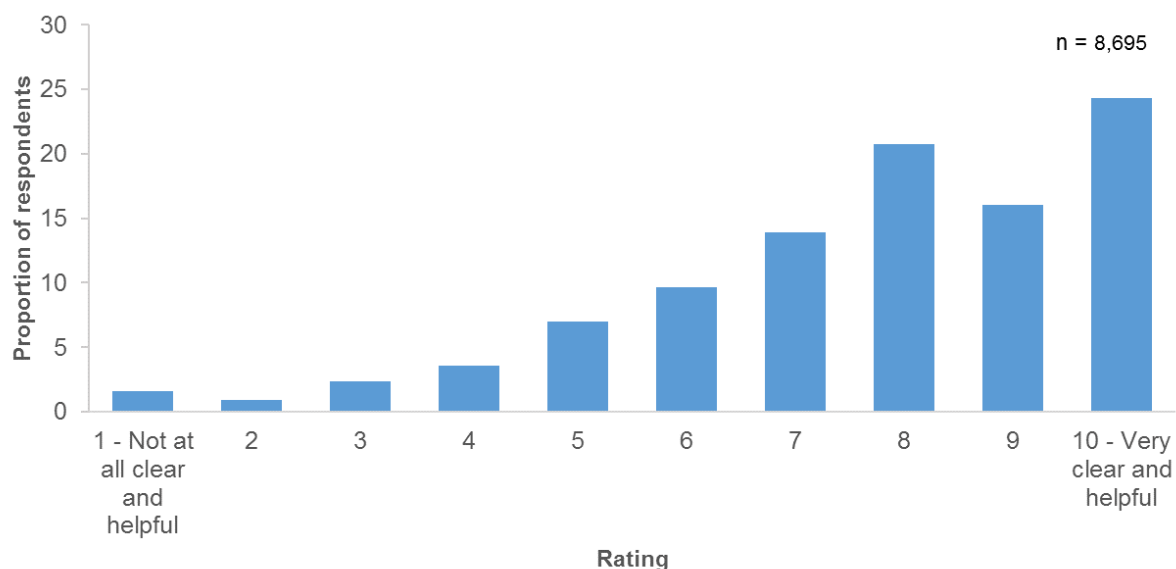


Figure 26. Proportion of respondents rating the extent to which annotation was clear and helpful in providing an understanding of where and how marks were awarded for individual responses.

**Besides any standard information on the scripts/responses on the system, did you receive any personal feedback from your team leader/supervisor after you submitted your practice scripts and before you started your qualification scripts?**

Fifty-six percent of examiners said that they received personal feedback from their team leader after the practice scripts/items. (n = 10,662)

A quick win would be to make sure as many examiners as possible get personal feedback after their practice scripts.

**Did your team leader/supervisor contact you or did you initiate the contact?**

The vast majority of respondents who received feedback were contacted by their team leaders (84%) while 16% said that they contacted their team leader. (n = 6,294)

**How did you get feedback after submitting your practice scripts (tick all that apply)?**



Overall, respondents said that the phone was the most frequent way of receiving feedback (62%) followed by through email (52%), through the marking system (33%), by text (6%) and finally other methods (2%).

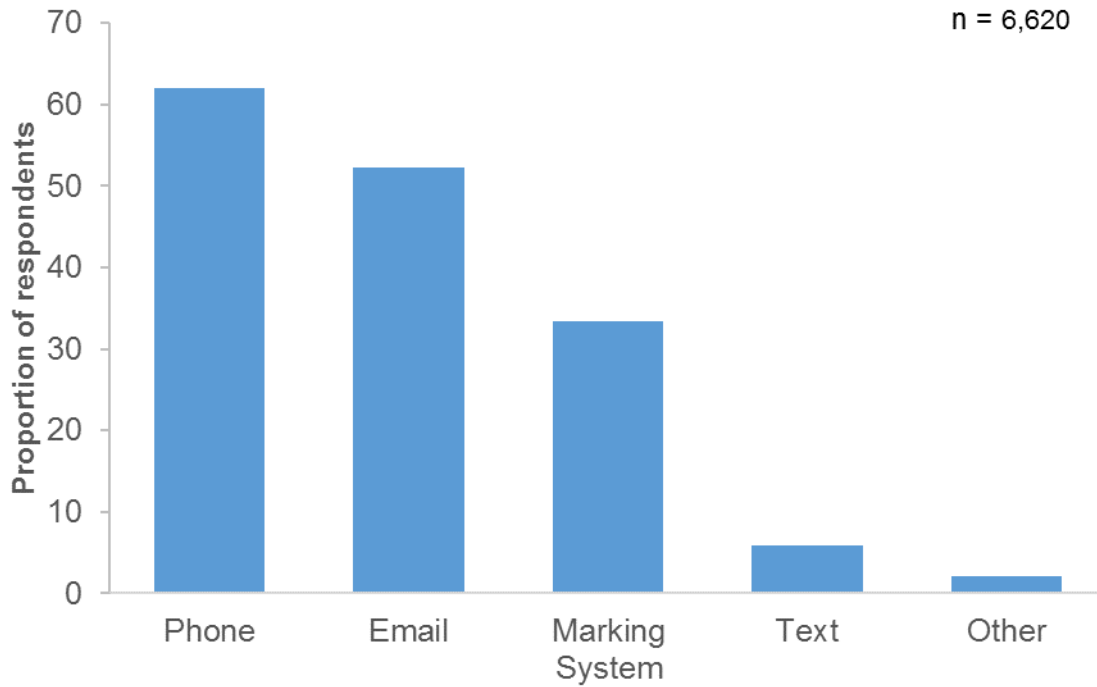


Figure 27. Proportion of respondents who said they received feedback through different media.

**How quickly did you receive feedback once you submitted your practice scripts?**

Most respondents (59%) received feedback within 6 hours of submitting their practice scripts. A further 26% of respondents received feedback between 6 and 24 hours after submitting their practice scripts. Finally, around 15% of respondents had to wait more than 24 hours before receiving feedback.

Given what we know from the literature, it might be important to aim for feedback to be given within 24 hours of submission so that examiners can clearly remember their rationale for awarding marks to an item (Johnson & Black, 2012).

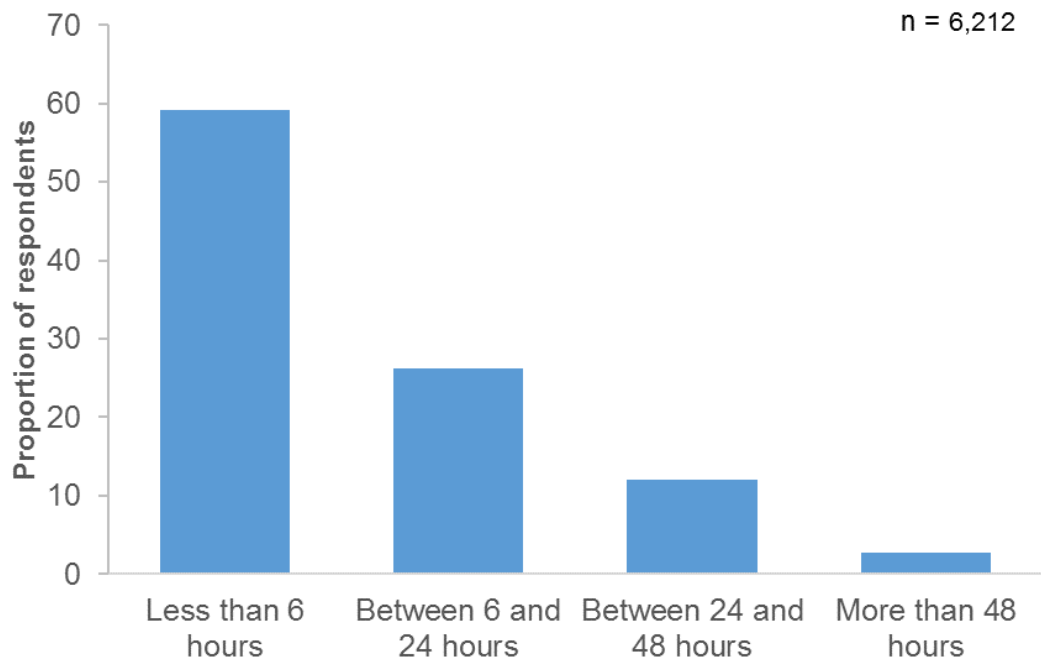


Figure 28. Proportion of respondents who received feedback within each different timeframes.

**To what extent was this personal feedback clear and helpful in providing you with an understanding of where and how marks were awarded for individual responses?**

Forty percent of examiners rated the personal feedback they received as very clear and helpful (rating of 10). Forty-three percent of examiners gave a rating of between 7 and 9 while 16% rated the feedback they received between one (not at all clear and helpful) and 6. Overall the quality of feedback, when given, was rated quite positively.

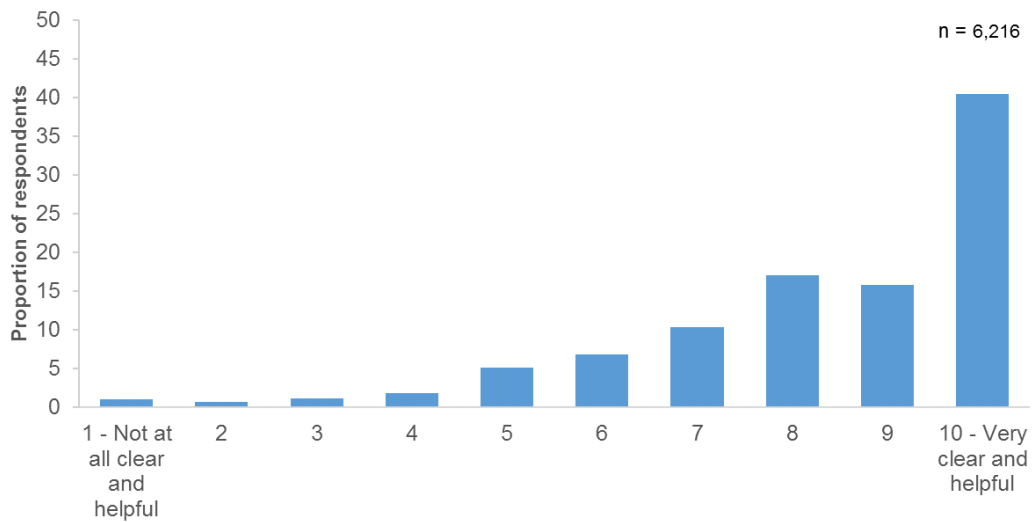
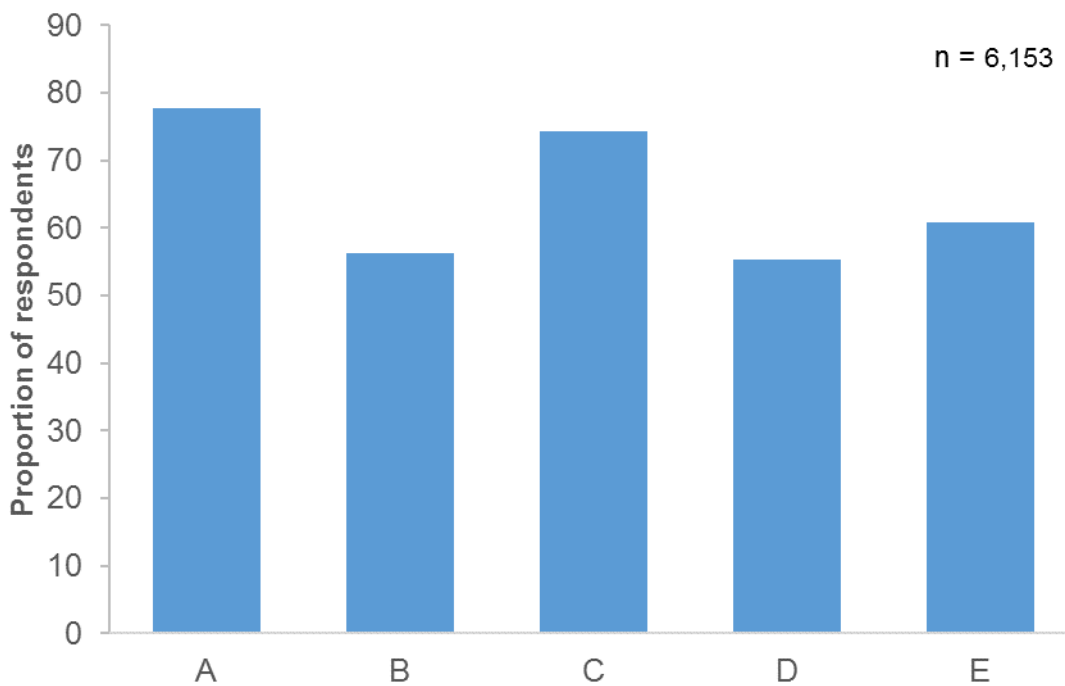


Figure 29. Proportion of respondents who rated how clear and helpful the individual feedback was after submitting the practice scripts.

**Did the feedback include the following (tick all that apply)?**

Seventy-eight percent of respondents said that the feedback they received after completing the practice scripts/items included general comments about the quality of their marking. Seventy four percent said that it included references to detailed aspects of the mark scheme whilst 61% said that the feedback linked the mark scheme to particular aspects of a candidate's response.



- A. General comments about the quality of your marking (for example 'your marking is fine')
- B. Trends/patterns in the quality of your marking (for example 'you have a tendency to be a little bit generous in question 4')
- C. Reference to detailed aspects of the mark scheme
- D. Reference to individual candidate's response
- E. Linking between the mark scheme and particular aspects of a candidate's response

Figure 30. Proportion of respondents who said the feedback they received after completing their practice scripts/items included the different elements listed above.

**Was the feedback tailored to the issues you encountered whilst marking your practice scripts?**

The large majority of respondents (93%) said that the feedback they received after completing their practice scripts/items was tailored to the issues they encountered while marking the practice scripts. (n = 6,200)

**To what extent did you discuss and question the issues raised in the feedback with your team leader/supervisor?**

Twenty percent of examiners who received feedback after submitting their practice scripts said they had in depth discussions with their Team Leader. Forty-four percent rated the extent of the discussion between 7 and 9 while 36% rated it between one (not at all) to 6. This may suggest that the feedback from team leaders, although generally useful, was not sufficiently in depth.

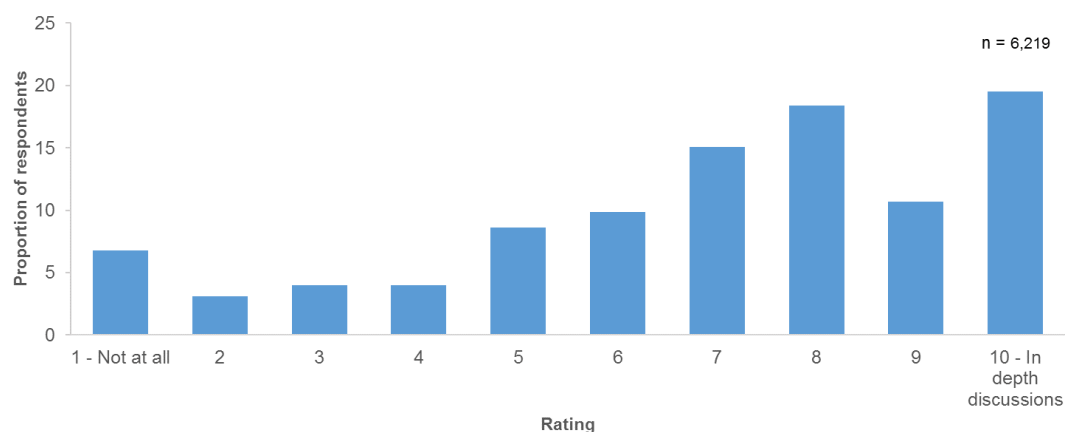


Figure 31. Proportion of respondents rating the depth of discussions with their Team Leader after submitting their practice scripts.

## Qualification phase

### How many qualification scripts did you have to work through?

Most examiners said that they completed 10 qualification scripts/items (50%) while 22% said they completed 5 scripts/items. The final 28% was distributed somewhat evenly between the other categories.

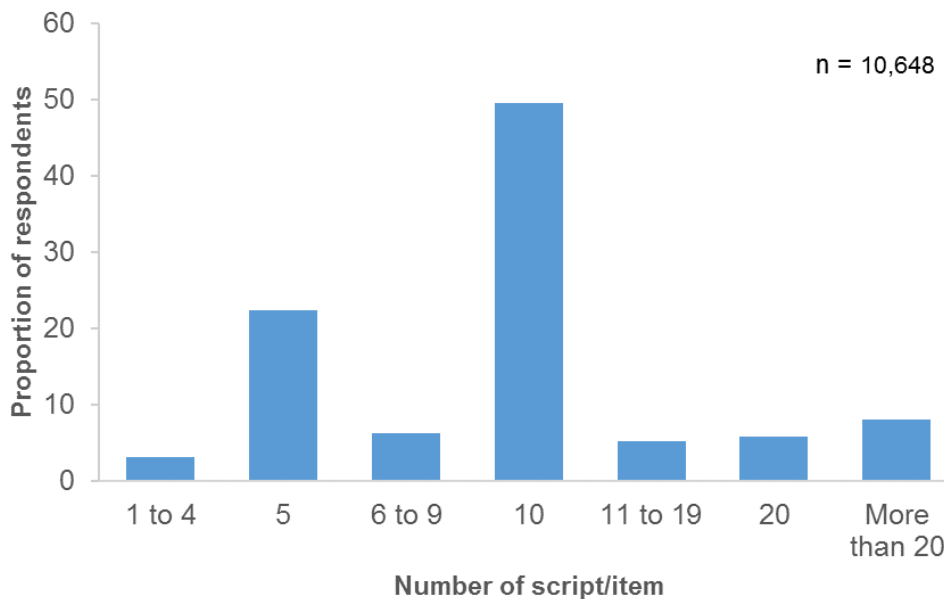


Figure 32. Proportion of respondents completing different numbers of qualification scripts.

### How important do you think the qualification scripts are for giving you sufficient understanding to apply the mark scheme accurately?

More than 60% of respondents said that the qualification scripts/items were very important in giving them sufficient understanding to apply the marks scheme accurately (rating of 10).

The comparison with the ratings for the practice scripts is interesting as it shows that examiners think that the qualification phase is more important than the practice phase for learning to apply the mark scheme. When looking at individual examiners we find that 57% of them think that both phases are as important (same rating to both questions), 31% think that the qualification phase is more important while only 13% think that the practice phase is more important. This is at odds with the view that the practice phase is the main learning phase in the online standardisation process. However, this could be due to the fact that not all exam boards include rationale and annotations in the practice phase. It may not be optimal to combine qualification to mark with the primary learning phase.

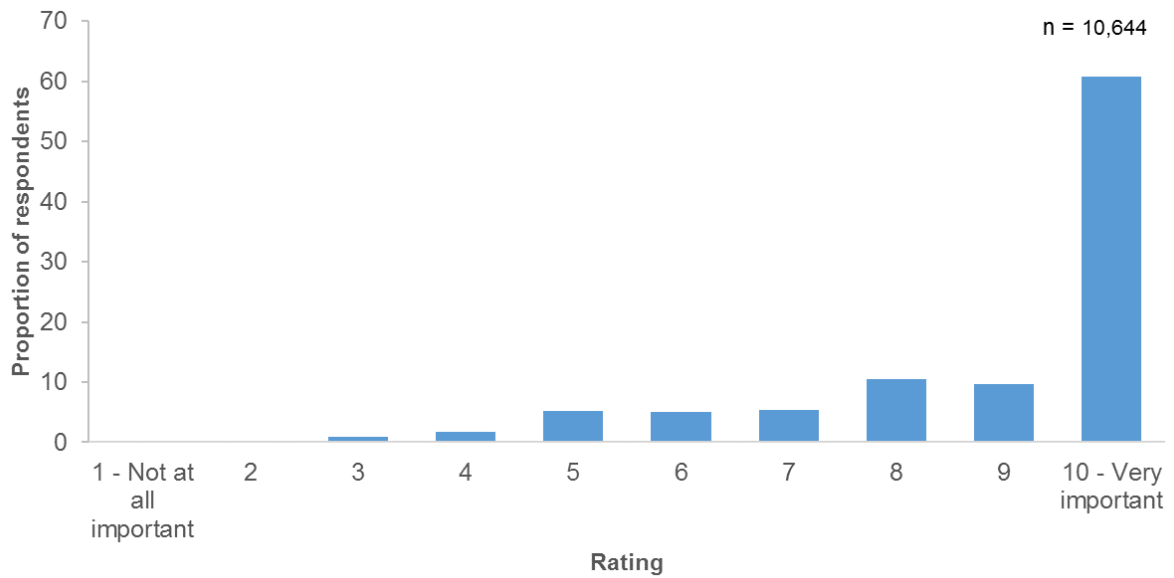


Figure 33. Proportion of respondents who rated the importance of the qualification scripts in giving them sufficient understanding to apply the mark scheme accurately.

**Did you get feedback on your qualification scripts?**

Ninety percent of respondents said they received feedback after submitting their qualification scripts (n = 10,647). Given that this is the final step in being qualified to mark, we might expect a somewhat higher percentage, even for those who easily pass qualification to mark.

**How did you get feedback after submitting your qualification scripts?**

Fifty-seven percent of respondents said they received feedback by phone after submitting their qualification scripts. Forty-four percent of respondents received feedback through the marking system while 40% of respondents said that email was used to provide feedback. Finally, 5% of respondents said they were contacted through other means (text and other).

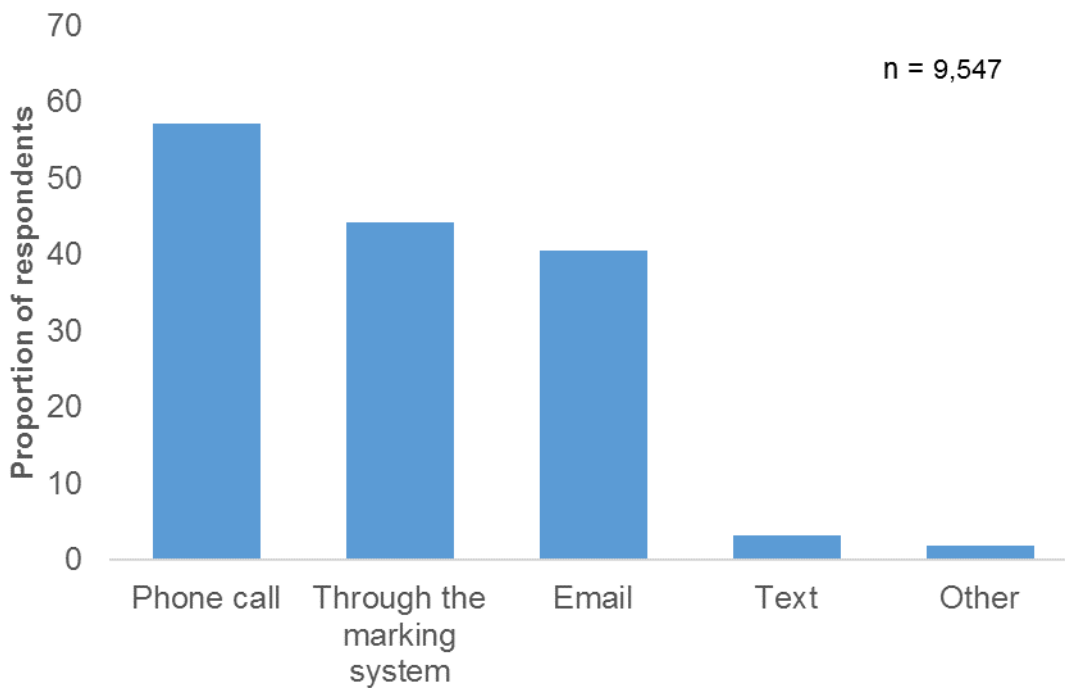


Figure 34. Proportion of respondents who were given feedback through the different media.



**How quickly did you receive feedback once you submitted your qualification scripts?**

Eighty-five percent of respondents said they received feedback within 24 hours of submitting their qualification scripts with 57% receiving it in less than 6 hours. This is very similar to the feedback speed after submission of the practice scripts.

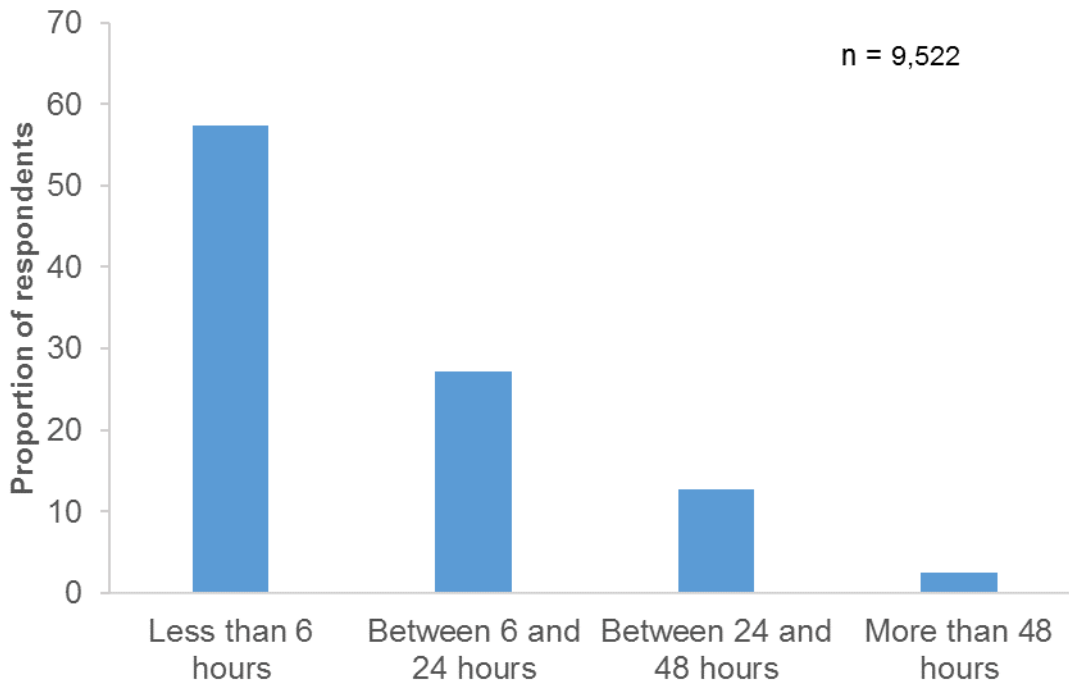


Figure 35. Proportion of respondents who received feedback after submitting their qualification scripts within the different timeframes.

**To what extent was this personal feedback clear and helpful in providing you with an understanding of where and how marks were awarded for individual responses?**

Forty-three percent of examiners said that the feedback was very clear and helpful (a rating of 10) while 42% rated the feedback between 7 and 9 (where one is not at all clear and helpful). Finally, 14% of examiners rated the feedback between one and 6. This is very similar to the ratings given for the feedback in the practice phase.

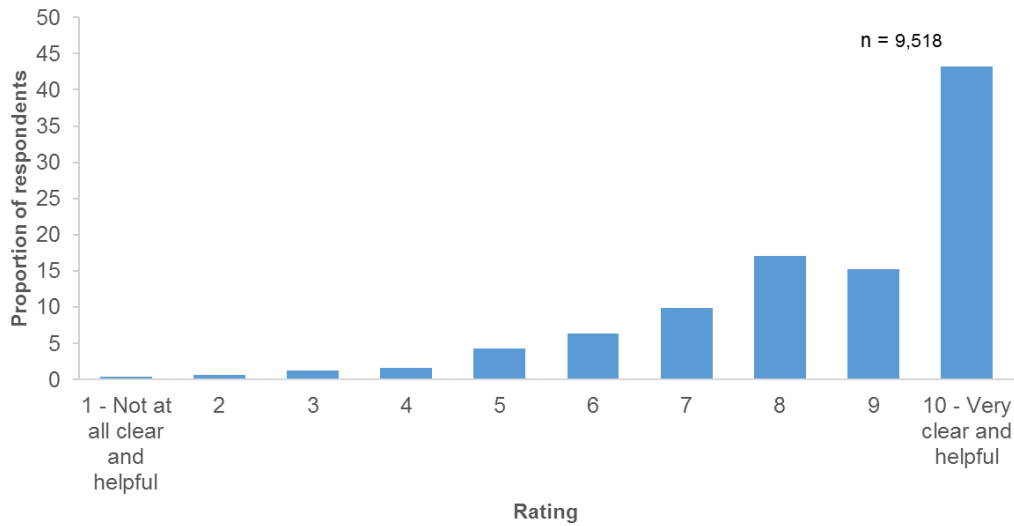
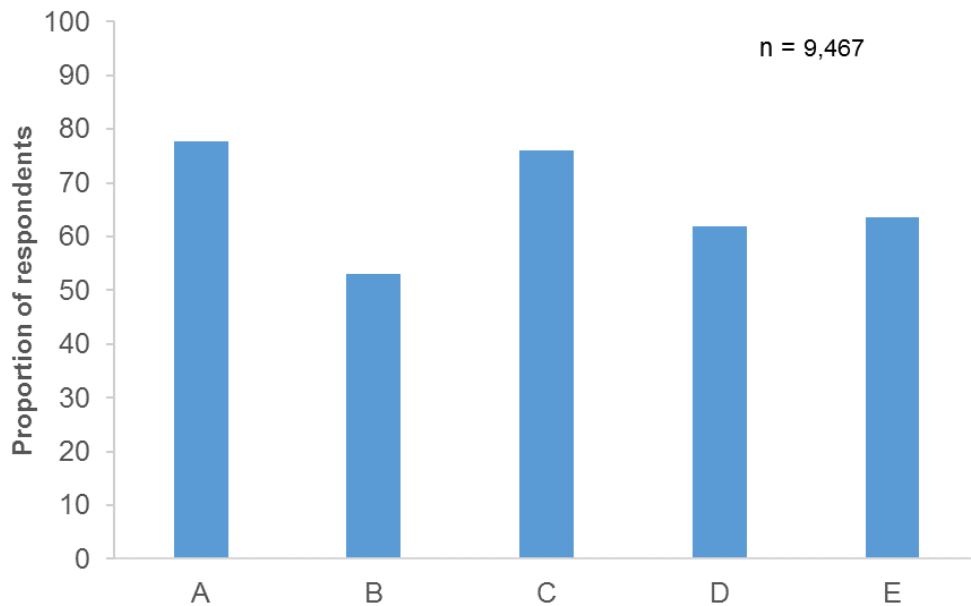


Figure 36. Proportion of respondents who rated how clear and helpful the individual feedback was after submitting the qualification scripts.

**Did the feedback include the following (tick as many as apply)?**

Seventy-eight percent of respondents said that the feedback they received after completing the qualification scripts/items included general comments about the quality of their marking. Seventy-six percent said that it included references to detailed aspects of the mark scheme while 64% said that the feedback linked the mark scheme to particular aspects of a candidate's response. These percentages are very similar to those from the practice phase.



- A) General comments about the quality of your marking (eg 'your marking is fine')
- B) Trends/patterns in the quality of your marking (eg 'you have a tendency to be a little bit generous in question 4')
- C) Reference to detailed aspects of the mark scheme
- D) Reference to individual candidate's response
- E) Linking between the mark scheme and particular aspects of a candidate's response

Figure 37. Proportion of respondents who said the feedback they received after completing their qualification scripts/items included the different elements listed above.

**Was the feedback tailored to the issues you encountered whilst marking your qualification scripts?**

Ninety-four percent of examiners said that the feedback was tailored to the issues they encountered whilst marking the qualification scripts. (n = 9,496)

**To what extent did you discuss and question the issues raised in the feedback with your team leader/supervisor?**

Respondents were asked to rate the extent to which they discussed and questioned the issues raised in the feedback they received after submitting their qualification scripts on a scale from one (not at all) to 10 (in depth discussions). Twenty-one percent of examiners who received feedback said they had in depth discussions with their team leader. Forty-three percent rated the extent of the discussion between 7 and 9 while 36% rated it between one (not at all) to 6. These percentages are very similar to those from the practice phase.

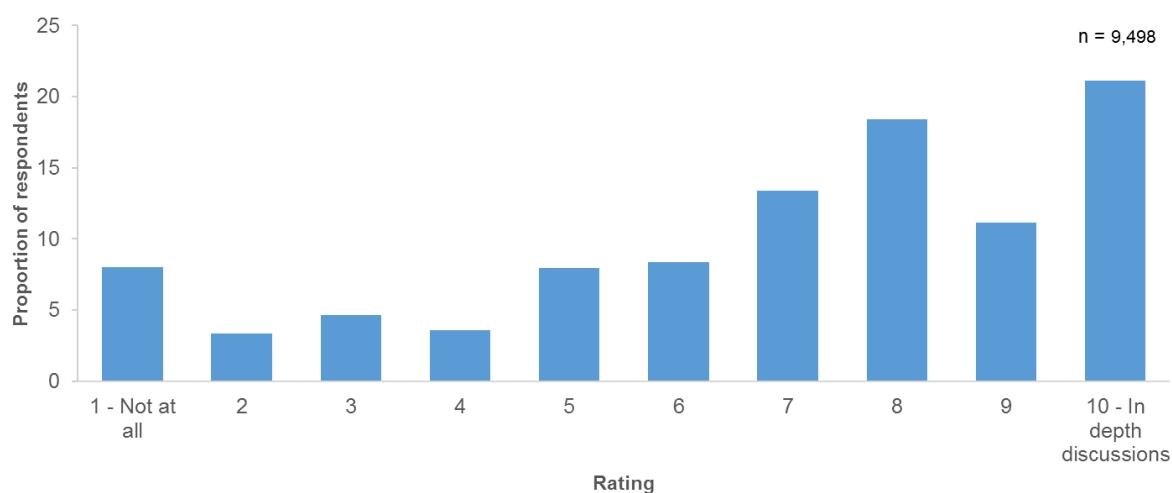
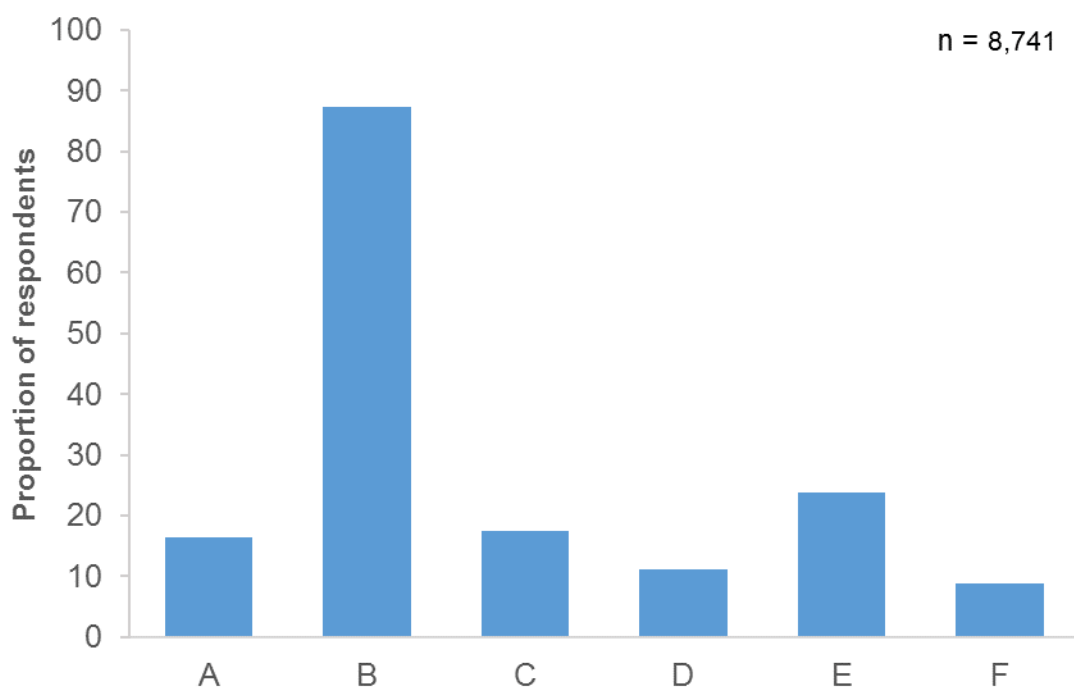


Figure 38. Proportion of respondents rating the depth of discussions with their team leader after submitting their qualification scripts.

## Overall experience

**When you receive feedback on the practice scripts, how do you make sure you integrate/remember it? (tick all that apply)**

Most respondents<sup>7</sup> (87%) said that they make notes of the feedback they receive on their mark scheme in order to remember it.



- A) I mainly just remember the feedback  
 B) I make notes on my copy of the mark scheme  
 C) I print the feedback  
 D) I print the candidate's answer  
 E) I make a note of the feedback on a piece of paper  
 F) Other

Figure 39. Proportion of respondents who use each of the above strategies to remember feedback given after the practice scripts.

**Do you feel that the practice scripts/responses had a good spread of quality (ie marks) for each question?**

Most respondents (71%) said that they felt that the practice scripts/responses had a good spread of quality for each question.

<sup>7</sup> Note that this item was in the survey for 3 of the 4 awarding organisations which explains the smaller number of respondents.

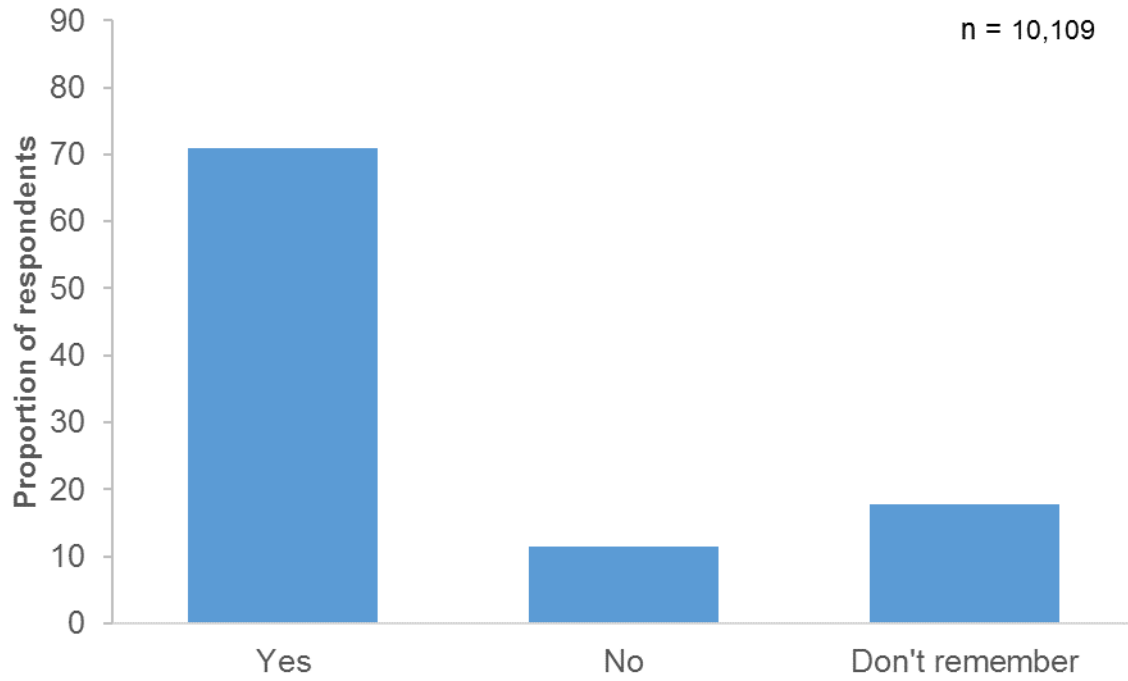


Figure 40. Proportion of examiners who felt that the practice scripts had a good spread of quality.

**Do you feel that the qualification scripts/responses had a good spread of quality (ie marks) for each question?**

Eighty-two percent of examiners said that the qualification scripts had a good spread of quality for each question. Given the survey was completed a number of weeks after the end of marking, a number of examiners could not remember whether there was a good spread of quality in both the practice scripts (18%) and the qualification scripts (8%). It is worth noting that the percentage who thought the scripts were not representative was low and did not differ much from the qualification phase.

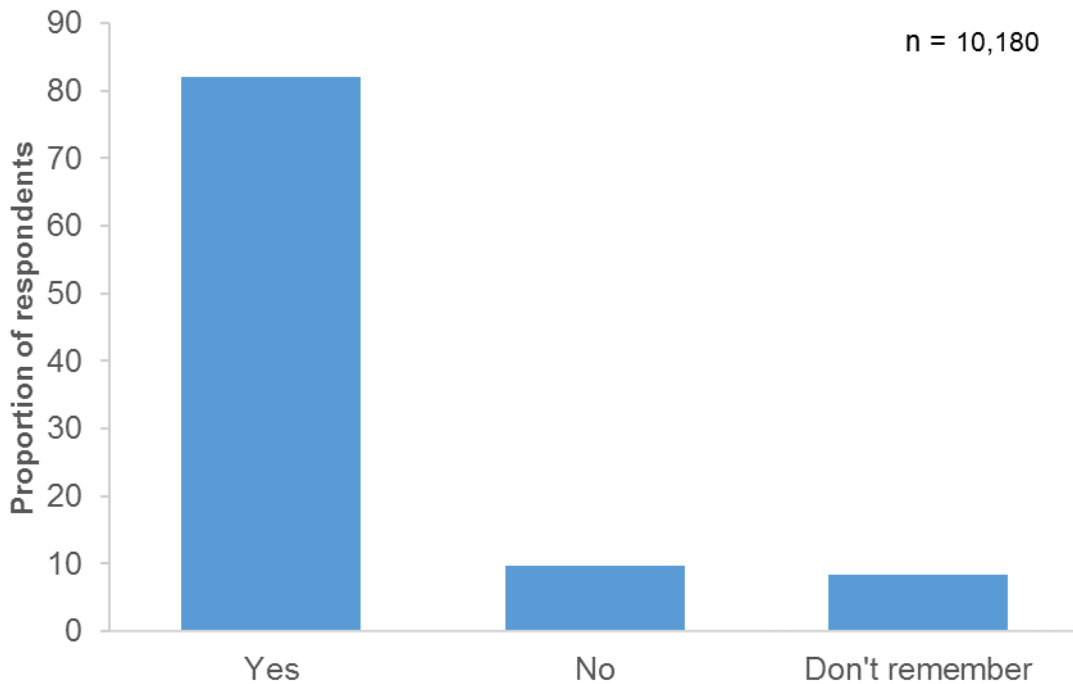


Figure 41. Proportion of examiners who felt that the qualification scripts had a good spread of quality.

**Do you feel that the practice scripts/responses were representative of your own marking allocation?**

Seventy-three percent of examiners felt that the practice scripts/responses were representative of their marking allocation while 13% felt they were not representative. The remaining 13% did not remember whether the responses were representative of their marking allocation. As discussed previously, the survey was filled a number of weeks after the end of marking and as the practice scripts were the first to be completed, a number of examiners could not remember them very well (18%).

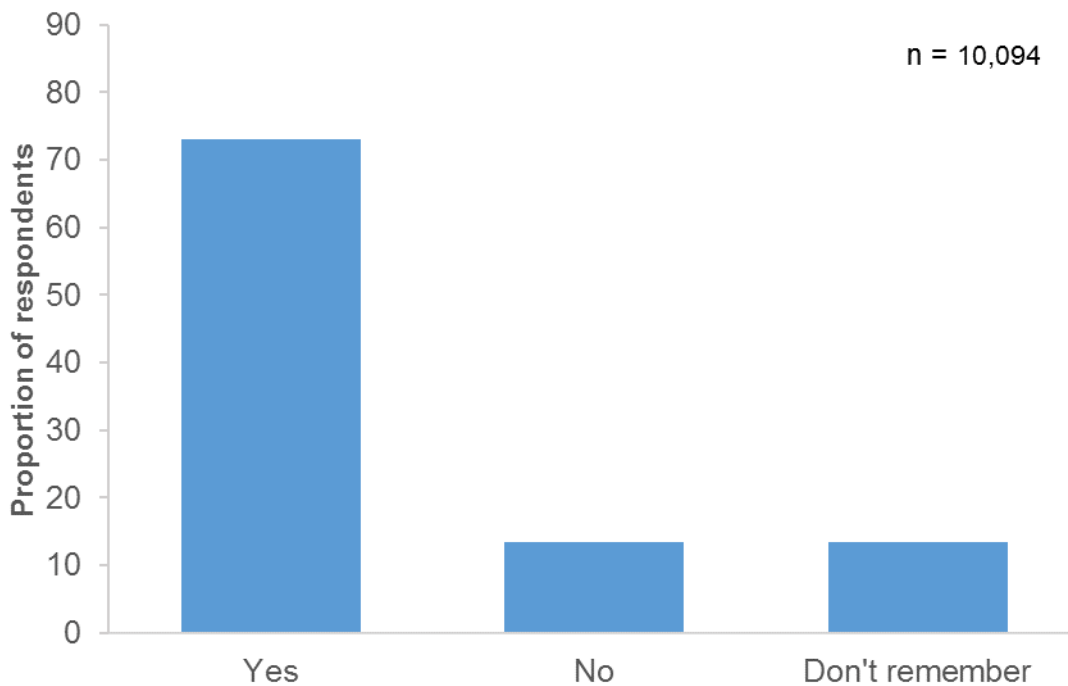


Figure 42. Proportion of respondents who felt that the practice scripts were representative of their marking allocation.



**Do you feel that the qualification scripts/responses were representative of your own marking allocation?**

For the qualification scripts/responses, 83% of examiners felt that they were representative of their marking allocation while 12% thought they were not. The remaining 5% of examiners could not remember whether the qualification scripts were or not representative of their marking allocation.

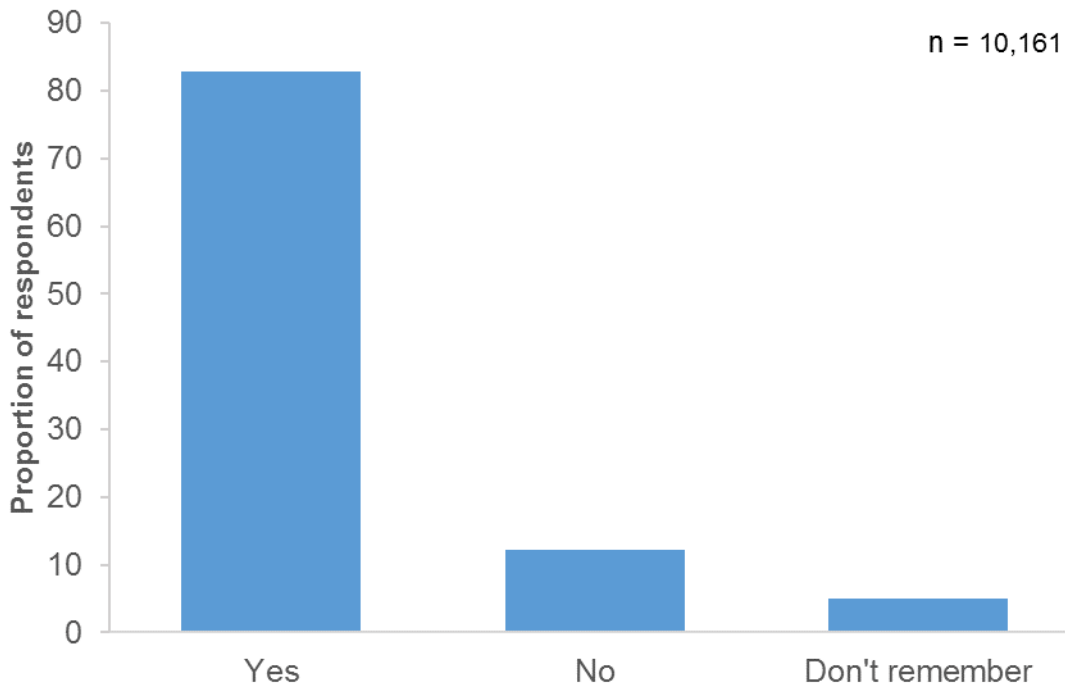


Figure 43. Proportion of respondents who felt that the qualification scripts were representative of their marking allocation.

**How many practice scripts do you think there should ideally be?**

Forty-four percent of examiners think that there should be 10 practice scripts/responses and this was the highest percentage for all the boards. Twenty-two percent thought there should be 5 practice scripts/responses.

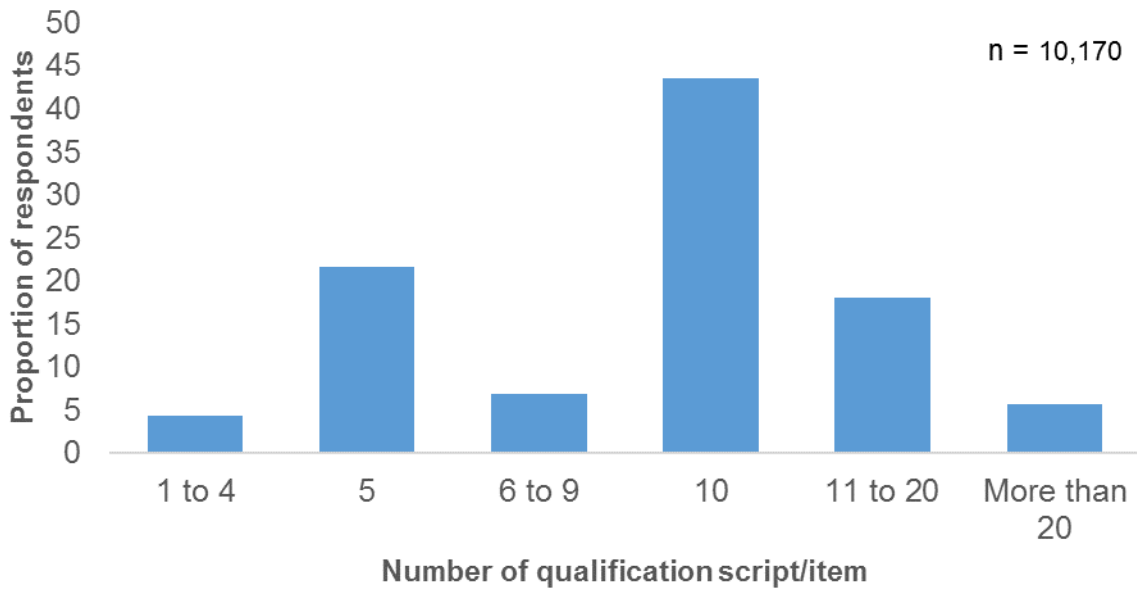


Figure 44. Proportion of respondents who think there should be these different numbers of qualification scripts.

**Marking well involves learning how to interpret candidate responses and applying the mark scheme consistently. If 100% is the total amount of learning you have done throughout the marking window, in what phases and in what amounts did that occur?**

On average, respondents said that 41% of their learning occurred during the qualification phase, 26% during the practice phase and around 33% during live marking.

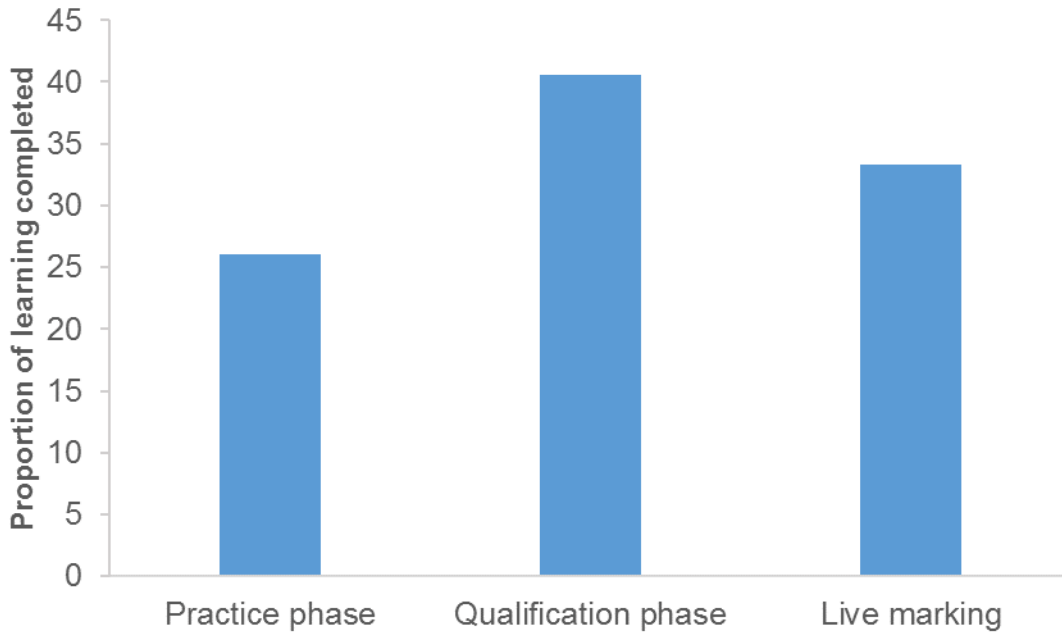


Figure 45. Average proportion of learning completed during each phase.

**At what point did you feel confident that you were marking accurately?**

Sixty-six percent of respondents felt confident in the accuracy of their marking by the start of live marking while 30% said that they felt confident in the middle of live marking. Although this sounds like a lot of markers made it half way through their marking before they felt confident, in fact if respondents did not feel confident at the very start of live marking, the next available response was the middle of live marking. We do not know how many became confident soon after starting their live marking. A further 2% said that they felt confident at the end of the qualification phase whilst 2% said they never felt confident. As noted earlier, these questions are about confidence in their own marking, and not necessarily about competence.

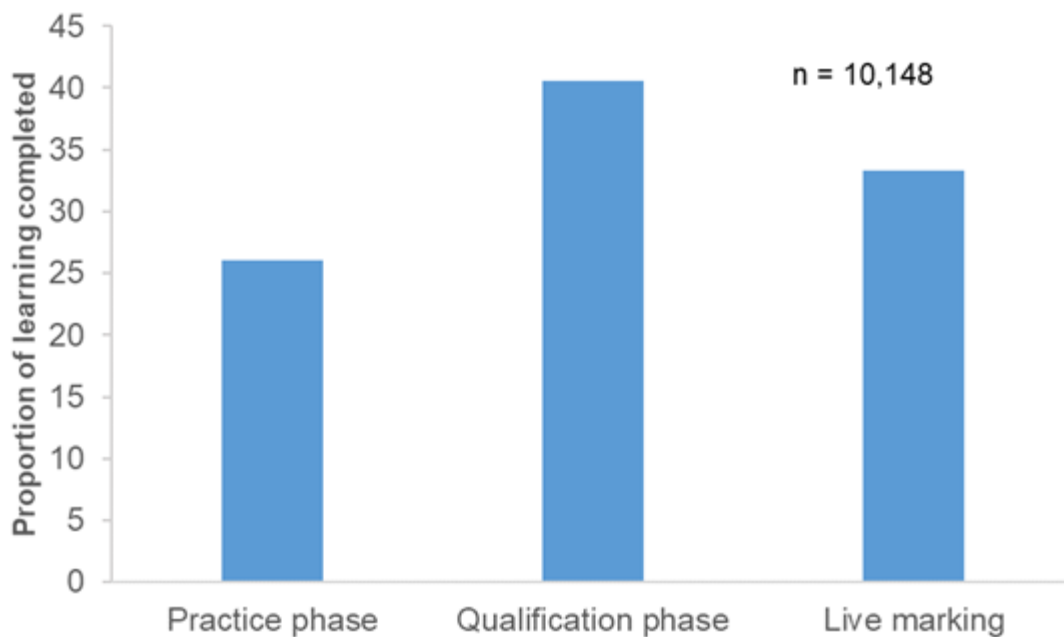


Figure 46. Proportion of respondents who felt confident in their marking by the end of each phase of the standardisation/marking process.

**One view of good marking is that it requires all the examiners to mark like the principal examiner. To what extent do you feel you were marking like the principal examiner during live marking?**

Although it is hard to know how people use the rating scale, and there is a large element of individual personality in determining confidence, markers felt they were generally marking similarly to the principal examiner. Eighteen percent of examiners rated their marking between a 1 (did not mark at all like the Principal Examiner) and 6 where a rating of 10 is marking the same as the Principal Examiner. Seventy-two percent of examiners rated their marking between 7 and 9 while 10% rated it as 10 – the same as the principal examiner.

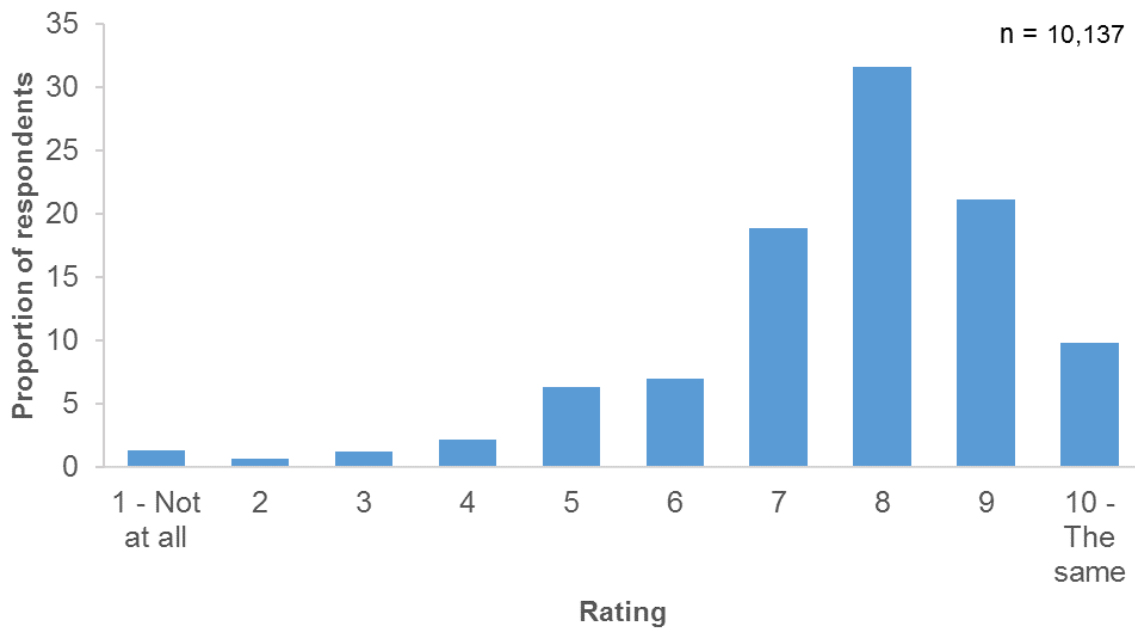


Figure 47. Proportion of respondents for each rating of how they thought they marked like the principal examiner.

**Did the practice scripts remain available on the system throughout the marking session?**

Sixty-one percent of examiners said that the practice scripts remained on the system throughout the marking session. Thirteen percent said they were not available on the system while 26% could not remember whether they were or not available during the marking session.

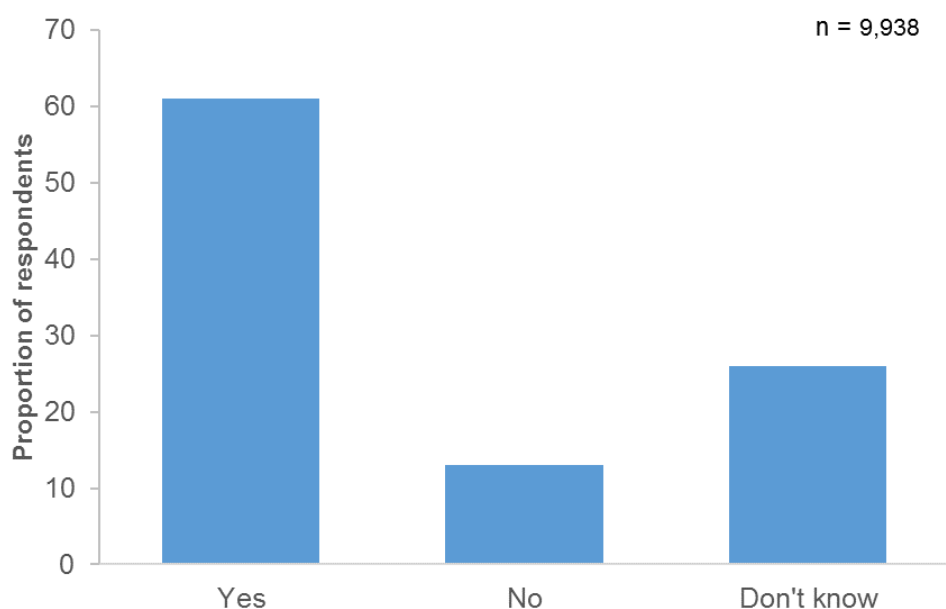


Figure 48. Proportion of respondents who said the practice scripts were/were not available throughout the marking session.

**Did you go back to the practice or exemplar scripts during live marking?**

Of the examiners who could access the practice scripts during the marking session (6,041), 72% say they went back to them while 28% did not.

**Did the qualification scripts remain available on the system throughout the marking session?**

A slightly larger percentage of examiners (72%) said that the qualification scripts remained on the system throughout the marking session compared to the practice scripts. Eight percent said they did not remain on the system while 20% did not know.

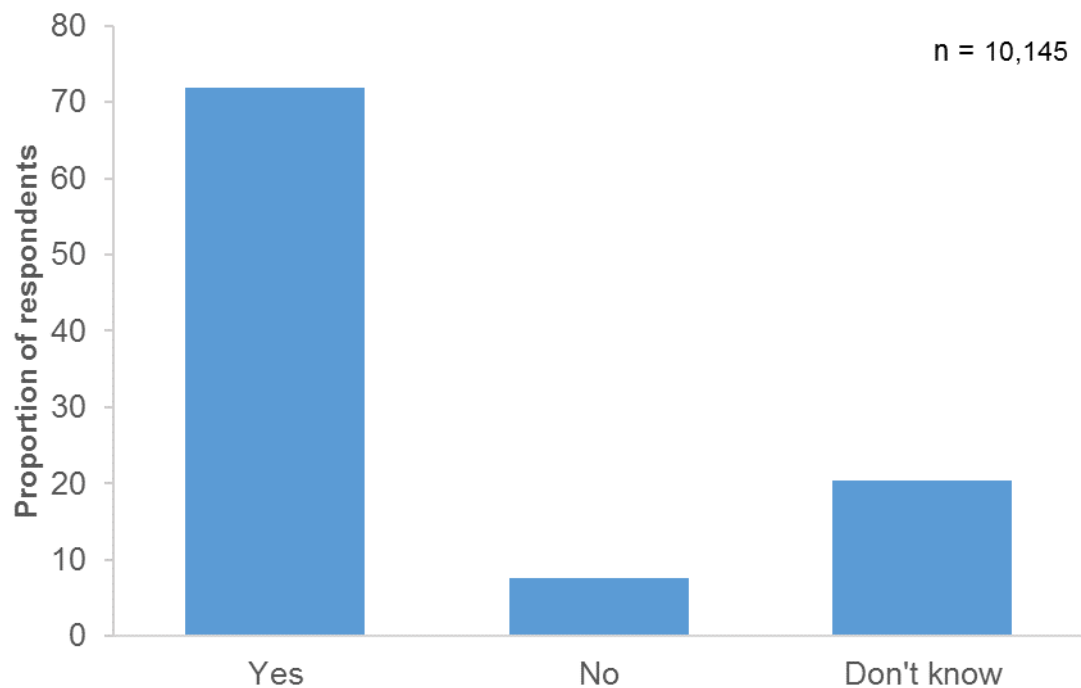


Figure 49. Proportion of respondents who said the qualification scripts were/were not available throughout the marking session.

**Did you go back to the qualification scripts during live marking?**

Of those who said that the qualification scripts were available throughout the marking session (n = 7,268), 83% said they went back to them while 17% said they did not.

**How happy are you with the current online standardisation system?**

Although most examiners were at least quite happy with online standardisation (a rating of 7 or above), there were a significant number who were not, to some extent reflecting the earlier results in the Quality of Marking report. Thirty percent of examiners gave a rating between one and 6 where one is not happy at all and 10 is very happy with the current online standardisation system. Fifty-one percent gave a rating between 7 and 9 while 10% said they were very happy with the current online standardisation system.

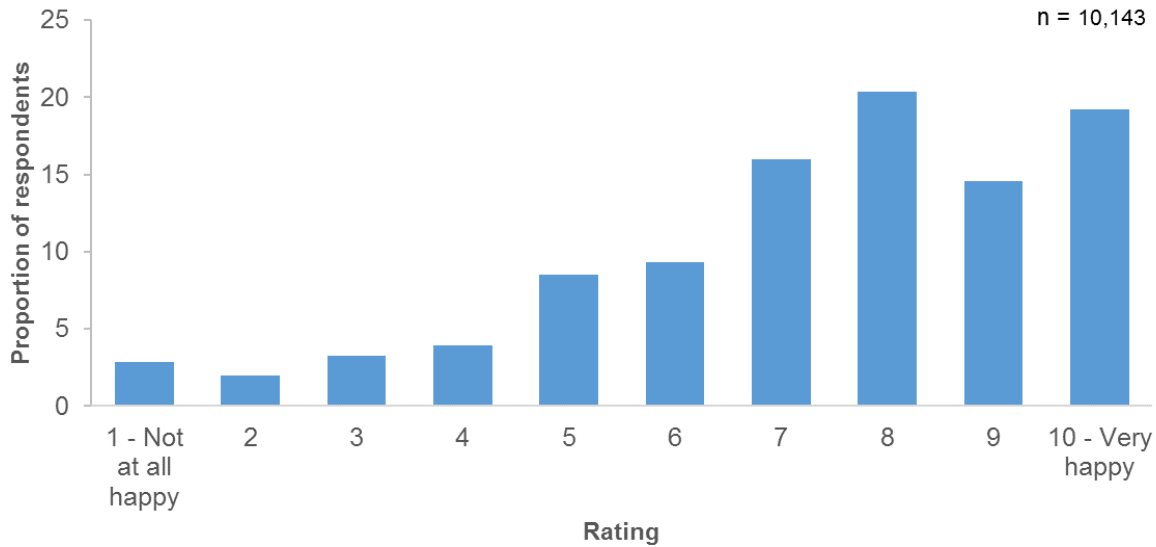


Figure 50. Proportion of respondents for each rating of how happy they are with the current online standardisation system



## Advantages and disadvantages of online standardisation

**In your experience, what are the advantages of being standardised (ie doing the practice scripts/items) online (tick all that apply)?**

This question was included in the survey of three of the four awarding organisations (n = 8,707). Respondents were asked to choose as many statements as applied. There were a wide range of factors which examiners felt positive about, but the most frequently cited were those centred on convenience.

The percentages of respondents who chose each statement are:

- 70% I can choose when I do the practice scripts
- 63% I don't have to travel to a training meeting
- 60% I get familiar with the marking system whilst doing the practice scripts
- 59% I can easily refer back to the practice scripts
- 54% I can choose where I do the practice scripts
- 52% I have enough time to reflect
- 42% I like to be able to do the practice scripts question by question
- 36% It is an effective way for me to learn
- 35% I get feedback promptly on the practice scripts
- 35% I don't get distracted by other examiners
- 32% I get feedback that is personalised on the practice scripts
- 32% I get detailed feedback on the practice scripts
- 31% It is quick
- 30% I know I get the same annotations/rationale on the system as all the other examiners
- 18% I can ask a lot of questions
- 5% No advantages
- 4% Other advantage

<p><b>In your experience, what are the disadvantages of being standardised (ie doing the practice scripts/items) online (tick all that apply)?</b></p>
--

This question was included in the survey of 3 of the 4 awarding organisations (n = 8,670). Respondents were asked to choose as many statements as applied. The majority of comments revolved around not meeting their fellow examiners and the lack of in-depth discussion and understanding of the Principal Examiner's standard.

The percentages of respondents who chose each statement are:

- 52% I don't have a chance to meet my team leader/supervisor
- 48% There are no in-depth discussions of the mark scheme around the practice scripts
- 47% I don't have enough opportunity to discuss in a group
- 46% I don't have a chance to meet the other examiners
- 37% I don't get a sense of common purpose with other examiners
- 31% I feel remote from the principal examiner and their way of thinking
- 31% It can feel impersonal
- 21% I have experienced IT problems (eg software or internet problems) in accessing the practice scripts
- 20% I sometimes have to wait for feedback on the practice scripts
- 18% After the practice scripts, I am still not confident
- 18% I don't feel I have enough influence on the mark scheme
- 15% The annotations/rationales on the practice scripts are not enough for me to learn from
- 12% No disadvantages
- 11% There are not enough practice or exemplar scripts
- 11% It is time consuming
- 10% Other disadvantage
- 8% It is not an effective way for me to learn
- 8% The feedback I receive from my team leader/supervisor on the practice scripts is not detailed or clear enough
- 3% I find the system difficult to use



© Crown Copyright 2018

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated.

To view this license, visit

[www.nationalarchives.gov.uk/doc/open-government-licence/](http://www.nationalarchives.gov.uk/doc/open-government-licence/)

or write to

Information Policy Team, The National Archives, Kew, London TW9 4DU

Published by:

**ofqual**

Earlsdon Park  
53-55 Butts Road  
Coventry  
CV1 3BH

0300 303 3344  
[public.enquiries@ofqual.gov.uk](mailto:public.enquiries@ofqual.gov.uk)  
[www.gov.uk/ofqual](http://www.gov.uk/ofqual)