

**38<sup>th</sup> ANNUAL CONFERENCE OF THE  
INTERNATIONAL ASSOCIATION FOR EDUCATIONAL ASSESSMENT  
KAZAKHSTAN, 16-21 September 2012**

**Paper Title:** Supporting Quality Control in Examinations through the use of Technology

**Presenter:** Graham Hudson  
Director of e-Assessment, DRS Data Services Limited, Milton Keynes, UK

**Email:** graham.hudson@drs.co.uk

**Abstract**

Awarding bodies globally continually strive to implement methods to improve the quality approaches used in marking tests and examinations. Balancing improvement in quality against the constraints of time and cost can be difficult, but the use of new technologies can be of assistance.

This paper will present different approaches used to monitor and manage the quality of marking outcomes using new technologies. Some potential new approaches will be discussed where feedback to markers is included in the approach.

**Key Words**

Electronic marking; efficiency; marking reliability, high-stakes, quality control, double marking, technology, international

## **Background**

DRS has successfully implemented electronic marking with a number of awarding body clients in the UK, the largest of which is AQA. The general benefits of using electronic marking are becoming more widely recognised both within the UK and internationally.

Key to the approach adopted by DRS and its clients is the focus on improving the quality of marking through the use of technology. Marking judgements made by senior examining personnel, combined with sophisticated algorithms, enable those marking standards to be built into a marking process that continuously checks standards with a regularity that could not be achieved in a paper-based system.

This paper reviews the approaches to quality control that are now possible and discussed their application in a variety of item types.

## **Quality control of marking**

Quality control of marking is of fundamental importance to the reliability of the outcomes of examinations. Conventionally, this is carried out using sampling or double-marking approaches to gather evidence about how well a marker is marking to the standards set by the Principal Examiner for a subject and passed on through team leaders to the markers – the purpose being to reduce to the absolute minimum any variation in marking between the Principal Examiner’s standard and that of each individual marker.

The exact approach will depend upon a number of factors, such as:

- where marking takes place – ie in a marking centre or at the home of each marker
- the question paper layout, type of questions being marked and the associated marking guidance structure
- the size of the examining team
- the view of an awarding body, or national regulator, to the use of statistical approaches to mark adjustment.

Overlaying those factors are the remaining two constraints of the ‘quality triangle’, namely ‘time’ and ‘cost’ and the balance that has to be struck between them – which will be different for each Awarding Body. The procedures used in practice, therefore, represent the outworking of the balance that is acceptable to the whole assessment community, including the regulators and allows the examination results to be delivered when required.

Where examination papers are marked from image, similar techniques are used. However, better outcomes can be achieved within similar ‘time’ and ‘cost’ constraints.

## **Assessing approaches to quality control**

In order to have a process for quality control that produces consistent outcomes, the following key elements need to be included:

- how the ‘true mark’ for each item checked is established
- how the samples of work are selected
- how the judgement about each item is made

- how the judgement about overall marking quality is made, and
- what follow-up actions are taken
- what mechanisms are in place for markers to adapt to feedback?

The true mark is most often established from the decisions made by the Principal Examiner, who is responsible for the preparation of the question paper and marking guidance. It is this standard that is passed down to the marker through the senior examining team and team leaders. Reducing the opportunities for diluting that standard is important as is providing clear guidance on what type of answers can be given credit and to what degree.

Ensuring that markers cannot influence the outcome of the quality control process by being part of the sample selection process also will improve the objectivity of the outcomes. In manual approaches, markers will almost certainly be involved in deciding which examples of marked work are reviewed for quality purposes.

In a manual sampling process the team leader not only has to hold the standard and apply it against the markers in his or her team, but also will be reviewing many different examples of candidates' answers as presented by markers. Reducing the points in the process where such judgements have to be made will improve consistency. Undertaking this manually involves considerable organisational and logistical effort for large marking teams, but can be facilitated readily through the support of technology.

The most difficult aspect of a team leader's role is to make an overall judgement at each sampling point about a marker's marking quality, taking follow-up actions and providing appropriate feedback. This is discussed in more detail in the next section in which the benefits of using technology here will be outlined.

The type of question being marked will, of course, affect the approach to making quality judgements. Setting aside objective test items, which require no such judgements, the scale of complexity increases through short-answer items, based upon empirical content, through longer-answer items to the most challenging, extended and discursive answers which describe or argue a position. For the more complex items, the challenges come not only from the length but also the content, where the variations in the marks awarded by markers from the 'true' mark will tend to be greater.

So, in assessing quality control approaches, what criteria could be used to judge suitability? Based on the key elements above, the following criteria are proposed which could be applied by item type as well as the degree of technological support:

Criteria	Elaboration
Establishing the true mark	What is the basis for establishing the 'true mark'? Is the line of accountability supported by the processes in place?
Sampling approach	What sampling approach is taken? Is the sample chosen without any involvement from the original marker? Can the sampling approach be varied according to question type? Can the sampling approach be adapted to markers who exhibit different competencies in maintaining the marking standard?

Criteria	Elaboration
Keeping to the true mark	What processes are used to ensure that marking can be checked against the true mark? What mechanisms are used to apply consistent tolerances to any marking differences identified?
Providing appropriate feedback	What processes are in place to provide feedback, where appropriate, to markers on marking standards? What processes are in place to ensure that any feedback to markers on marking standards applies to all marking undertaken?
Corrective action	What information is provided to team leaders to help in making decisions about any corrective action required? Are the options for corrective action clearly presented as part of the process?
Post-marking procedures	What post-marking procedures are used to check the marking quality? What adjustments, if any, could be made to marking once completed?
Effort required to manage the quality control	What is the overall level of effort required to manage the quality control approach? What is the level of effort required by team leaders and markers to participate in the quality control approach?

Set out below is an example of how this framework might be used when comparing a manual sampling approach with one form of marking from image. Both processes involve the same standardisation exercise where markers meet and clarify the marking guidance with the Principal Examiner and their team leaders.

The example chosen is based upon the following scenarios, for an examining team of 100 markers, with one Principal Examiner and ten team leaders:

Manual marking of complete scripts for each candidate allocated to a marker	<ul style="list-style-type: none"> <li>• Samples of scripts are sent to a team leader through the post at defined intervals in the marking process.</li> <li>• The first sample includes some common scripts issued to all markers, the remaining scripts are chosen by the marker to defined criteria.</li> <li>• The second sample is chosen by the marker to defined criteria.</li> <li>• Feedback is given after each of these samples and further sampling may be required if the marking is still not deemed to be marking to the true standard.</li> <li>• A final sample is chosen by the marker to defined criteria. No feedback is provided.</li> <li>• A final review of markers' marking takes place to establish if any adjustments are required.</li> </ul>
Marking from image of individual items drawn from the system by the marker none of which are from the same candidate	<ul style="list-style-type: none"> <li>• The Principal Examiner and team leaders meet to choose marked items which will be placed into the quality control queue of the marking process. These items are known as 'seeds' and are presented to markers during the marking process, but are not identified as seeds to the markers.</li> <li>• Markers who exceed the marking tolerance set for the seed are stopped by the marking system.</li> <li>• The team leader speaks to the marker, discusses the seeds and explains the true mark. Markers can be allowed to mark again or, in some cases, will be stopped from marking that item any more. Marks already given can be recycled for marking again.</li> <li>• No final review of marking takes place.</li> </ul>

Example comparison:

Criteria	Manual Marking with Postal Sampling	Marking from Image Using Seeded Items
Establishing the true mark	<p>The true mark is established by the Principal Examiner and transferred to the team leaders.</p> <p>The line of accountability is through the team leaders who apply the standard to the varied samples of work submitted by markers.</p> <p>The process is dependent upon each team leader transferring the standard correctly and consistently in all cases.</p>	<p>The true mark is established by the Principal Examiner and transferred to the team leaders.</p> <p>The line of accountability is through the team leaders who set appropriate seeds into the quality control system and follow up any marking variances. A key element is the fact that the seeds carry the 'true' mark and can be used as a consistent measure, no matter which team leader is reviewing the work of markers.</p>
Sampling approach	<p>Samples of work are taken at defined points in the marking process.</p> <p>Apart from some initial common scripts provided to markers, the remained are all chosen by the markers.</p> <p>The samples will be complete scripts for the candidates and cannot be varied between item types.</p> <p>Additional samples can be requested from markers where further evidence of applying the true marking standards is required.</p>	<p>Seeded items are presented to markers according to a defined algorithm that is part of the marking from image technology.</p> <p>The original marker does not choose the seeded items and does not know which items they are.</p> <p>Different seeded items can be given and mark tolerances set according to item type.</p> <p>The proportion of seeded items given to markers can be varied according to marking competency exhibited.</p>
Keeping to the true mark	<p>Marking can be checked against samples of scripts used during the standardisation of marking phase.</p> <p>Assessing marking variation is undertaken at a total mark level, with tolerances set in terms of the number of mark changes, the different in absolute total mark and the difference in numerical total mark.</p>	<p>Marking can be checked against samples of scripts used during the standardisation of marking phase.</p> <p>Marking variation is assessed regularly and automatically throughout marking against the seeded items at the item level, which include appropriate mark tolerances.</p> <p>Markers' records in terms of the number of seed failures, marking differences and trends can be reported when desired from the marking system.</p>
Providing appropriate feedback	<p>Markers are provided feedback on marking at the end of the first and second sample stages. Issues that arise during marking can be checked with team leaders as a matter of course.</p>	<p>Markers are provided feedback on marking where they fail seeded items. Issues that arise during marking can be checked with team leaders as a matter of course.</p>

Criteria	Manual Marking with Postal Sampling	Marking from Image Using Seeded Items
Corrective action	<p>Where mark changes are required following the first and second samples, markers are instructed to review all previous marking and make changes as necessary. If there is any lingering doubt, a further sample can be required.</p> <p>Establishing whether or not appropriate action has been taken by the marker is difficult to show until further sampling takes place.</p> <p>Markers can be stopped from marking. The scripts that remain to be marked are re-distributed to other markers. A decision is made on whether to adjust or re-mark scripts already marked.</p>	<p>Where there is any doubt about the marking of a batch of items, they can be returned to the marking pool defined by the dates between which the marking took place.</p> <p>Team leaders have the choice only to accept marking or to recommend that the marker is stopped from marking.</p> <p>However, as the markers are marking by item and not the complete script of a candidate, they can continue to mark other items.</p>
Post-marking procedures	<p>Further sampling may be undertaken at the completion of marking as well as taking into account the overall marking performance for each marker when compared with the marking of other markers individually and collectively.</p> <p>This may result in an adjustment of the marks for specific markers or the re-marking of some of the scripts.</p>	<p>No post-marking procedures take place as the marking has taken place at the item level and has been checked regularly throughout the marking allocation.</p>
Effort required to manage the quality control	<p>The overall effort can be judged by establishing how many candidates scripts have been re-marked as part of the quality control process. For a team of ten markers this would equate to approximately <b>320 scripts</b>, assuming that no additional samples were requested.</p> <p>Markers are required to select samples, despatch them, manage the feedback on marking standards and manually adjust scripts where mark changes are required.</p>	<p>The overall effort can be judged by establishing how much effort is required to set up the quality control seeds at the outset. Team leaders set seeds in pairs. Assuming a question paper of 20 items with 60 seeds required for each item, this would equate to approximately <b>120 scripts</b>.</p> <p>Markers are not required to select samples, only to take into account feedback when marking future items.</p>

Whilst time and organisation is required to set seeded items into the system at the start of the marking process, there are specific benefits over and above the manual processes:

- common items, with known marks, are used with all markers;
- the checking process is regular and identifies any unacceptable marking variance during marking;
- feedback on marking standards can be provided, with examples to refer to;
- markers do not know which items are seeded ones and which are not;
- no post-hoc processes are required to confirm the marking standards.

### Quality control processes in marking from image

The above example shows us quality control by using ‘seeded items’. Other methods of quality control can be used and each has its own specific applications and benefits. The following section outline two of these methods are and how they are used.

1. Percentage Double Marking

<b>QC Method</b>	Percentage Double Marking
<b>Marking approach</b>	Marking is undertaken by individual item. Short-answer questions or longer-answer essay questions can be marked using this approach.
<b>QC Principles</b>	The QC principle used in this case is the comparison of one marker’s mark for an item with another marker. The standard is determined by consensus, if the marks of the two markers are within a defined marking tolerance for that item. Rules to enable adjudication of the marking can be put in place where a marking tolerance is exceeded.
<b>QC Description</b>	<p>Based on business rules, which can be varied by the administrator, a marked item from one marker is given to another marker to mark. A comparison of the marks is made by the system and, if a difference is found that is outside an acceptable tolerance, then the marking for each marker is stopped.</p> <p>A senior marker reviews the marking of each marker and makes a judgement about the mark that should be given. Depending on the outcome of the review, the mark of one or other of the markers is confirmed or a new mark given by the senior marker.</p> <p>Where a marker exceeds a tolerance for an item repeatedly, the marker is stopped and discussion with the senior marker takes place.</p> <p>Variants on this approach can include the comparison of marking by three markers before being sent to a senior marker. In addition, automatic adjudication can take place by choosing the highest mark or lowest mark or by the system computing an average.</p>
<b>Benefits</b>	<p>This approach enables:</p> <ul style="list-style-type: none"> <li>• longer-answer questions to be marked, which would not be suitable for creating seeds because of their length;</li> <li>• marking standards to be monitored regularly, as with seeding;</li> <li>• the degree of double-marking can be varied;</li> <li>• feedback on marking standards can be given;</li> <li>• items checked without the marker knowing which ones are involved;</li> <li>• no post-hoc processes are required to confirm the marking standards.</li> </ul>

2. Batch Sampling

<b>QC Method</b>	Batch Sampling
<b>Marking approach</b>	A complete script for a candidate is presented to a marker. All items for that script are marked by one marker, with the marks for each item being captured individually. The system totals the marks once the script is submitted by the marker.
<b>QC Principles</b>	Common, pre-marked scripts are issued to markers at known points during the marking process. These scripts are issued in defined marking batches, so that marking quality can be checked periodically. The same scripts are given to each marker and represent the true mark as defined by the Principal Marker for the subject.
<b>QC Description</b>	<p>At the start of marking, common scripts are given to each marker to mark. Depending on the outcome of the marking, markers are allowed to move onto live marking of candidates’ scripts.</p> <p>The marking is divided up into batches and common, pre-marked scripts are issued to the markers in each batch. The marking quality for markers is judged against the pre-marked scripts. Depending upon the outcomes, different interventions can be made by the senior marker that could result in either the marker continuing or being stopped from marking.</p>

<b>Benefits</b>	<p>This approach enables:</p> <ul style="list-style-type: none"> <li>• the work of a complete candidate to be assessed by one marker;</li> <li>• marking standards to be monitored regularly, against a common marking standard;</li> <li>• the number of pre-marked scripts issued in a batch can be varied;</li> <li>• feedback on marking standards to be given;</li> <li>• access to more comprehensive statistical analyses because of the use of common per-marked scripts;</li> <li>• complete scripts are checked without the marker knowing which ones are involved;</li> <li>• no post-hoc processes are required to confirm the marking standards.</li> </ul>
-----------------	--

Both approaches have common benefits, but with different marking approaches. The degree of monitoring provided would not be possible without working from the images of scripts. However, an important aspect of this approach is providing regular, helpful information to senior markers on the marking quality so that good decisions about individual marker’s marking can be made. This is discussed further in the next section.

**Quality of Marking Decisions**

An important factor in ensuring consistent control of the quality of marking is the information that can be made available to the decision-makers.

With the manual process, the information that can be used during marking is limited and generally comprises:

- a review of the marks, by item, of a sample of scripts;
- identifying mark differences by item and by complete paper;
- identifying absolute mark difference by complete paper;
- a review of trends, general severity or leniency.

The information is often presented in written form, requiring the senior marker to complete his or her marks on a mark sheet and calculate differences and absolute differences. Judgements are then made based on this information, but normally without any analyses of performance across the team or the whole marker population.

However, with marking from image, the fact that marks have been captured electronically and data analyses performed, provides a much more powerful basis for making decisions. How this applies to the difference quality control approaches is summarised below:

<b>QC Method</b>	<b>Information Available for Decision-Making</b>
Seeding	<ul style="list-style-type: none"> <li>• The ‘true mark’ for the seed, given by the senior marking team.</li> <li>• Details of any marking annotation given by the marker.</li> <li>• The mark given by the marker.</li> <li>• Access to ‘Quality of Marking’ reports that enable a review of marking undertaken by a marker over time.</li> </ul>



Percentage double marking	<ul style="list-style-type: none"> <li>• The marks given by each marker.</li> <li>• Details of any marking annotation given by each marker.</li> <li>• Access to ‘Quality of Marking’ reports that enable a review of marking undertaken by a marker over time.</li> </ul>
Batch sampling	<ul style="list-style-type: none"> <li>• The marks given by the marker.</li> <li>• Details of the marking annotations given by the marker.</li> <li>• Access to mark difference and absolute mark difference reports for the pre-marked scripts in each batch.</li> <li>• Reports based upon comparisons of the marking of the pre-marked scripts across different sub-groups of the marking population or the whole marking population.</li> <li>• Trend analyses of the marking.</li> <li>• Access to ‘Quality of Marking’ reports that enable a review of marking undertaken by a marker over time.</li> </ul>

As information is gathered about marking quality over time for an individual marker and for groups of markers (depending upon the quality control approach taken), more comprehensive and useful information can be provided to a senior marker to support marking decisions. Decisions about marking quality can be taken at a time when interventions can be made to correct any divergence in standards. Information over different marking series can also be captured at a detailed level, providing helpful information to administrators for establishing who should be appointed as a marker in the future.

In the manual marking approach, this kind of information would only be available once the majority of marking had taken place. At this stage, the opportunity to intervene and correct marking standards has passed.

### **Summary**

The use of a framework, as proposed, could be extended and would enable comparisons of the methodologies to be compared both within a particular marking system and across difference marking systems. The features of the different methodologies could be identified as well as the management effort involved in putting each approach into practice. It could be applicable across a range of assessment types set in the international context.

More comprehensive information of this kind can also help the regulators of qualifications to engage with how awarding bodies are managing the reliability of marking outcomes and provide better information for assessing risk of failure in this respect.

Without the use of technology, combined with the use of real-time reporting approaches, the use of such quality control approaches would not be possible.

### **Contact**

If any aspect of this paper is of particular interest, please contact:

Graham Hudson, Director of Electronic Assessment, DRS Data Services Limited  
 Email: [graham.hudson@drs.co.uk](mailto:graham.hudson@drs.co.uk); Mobile: +44 7766 831578

## **ANNEX 1 – Assumptions made in effort to manage quality**

### **Question paper**

A paper comprising 20 items

### **Marker Team**

Team of 100 markers

Ten Team Leaders

One Principal Examiner

### **Manual marking**

Four samples taken:

Sample 1: 10 scripts, 3 common scripts, 7 chosen by the markers

Sample 2: 10 scripts, all chosen by them marker

Sample 3: 10 scripts, all chosen by the marker

Review: 5 scripts, chosen by the marker team

This equates to a total of **320 scripts** re-marked by each team leader.

### **Marking from image**

Seeding meeting with Principal Examiner and 10 Team Leaders

Seeds set for each of 20 items

60 items for each seed created for seed bank

Team Leaders working in pairs to select seeds

This equates to a total of **120 scripts** re-marked by each team leader.