

**32nd IAEA Conference – Singapore, 21-26 May 2006
Presentation by Graham Hudson, National Business Development Manager,
Education, DRS Data Services Ltd**

Giving candidates a fairer deal in examinations and tests through electronic marking

Abstract

The presentation will consider how electronic marking brings benefits to candidates taking high-stakes tests and examinations where marking accuracy, reliability and timeliness are paramount. Areas to be covered will include how marking quality is monitored and maintained through multiple marking methods, ensuring anonymity, impartiality and avoidance of bias during marking and how detailed data can provide for improvements in both test design and student learning.

Some discussion of the challenges of deploying an electronic marking system across more than one country will be included, covering data security, resilience and performance matters.

Author

The author of the paper is Graham Hudson. Graham is National Business Development Manager for Education for DRS in the UK. Graham Hudson has over twenty years' experience of implementing and managing large-scale assessments within the UK. His experience covers developing and managing general qualifications since 1983, including the introduction of GCSE examinations in 1988 and the National Curriculum examinations in 1994.

Graham worked at the Qualifications and Curriculum Authority for over three years during which time he ran the external marking and data collection of the Key Stages 2 and 3 tests in England and established a government-funded programme for implementing the use of new technologies in examinations and assessments.

Graham now works for DRS where he has put in place the electronic mark capture and marking of tests for three unitary awarding bodies in the UK capturing 30 million marks in 2005.

Summary

This paper will consider the changes in the e-assessment environment since the IAEA Conference in Abuja. It will update progress on the implementation of electronic marking in the UK and in other countries.

In particular, the paper will examine how electronic marking can bring improved reliability to examination marking – which, it is hoped, will lead to more extensive work investigating between and inter-marker reliability. The volume and detail of the data collected from e-Marker® is extensive and provides a basis for research that has not been available previously.

Discussion of the quality checking methods employed will be supported with some data taken from previous examinations – illustrating how marking agreement can be judged by making use of seeded items.

Interestingly in the UK, the emphasis for change now seems to rest with the awarding bodies themselves, with the Qualifications and Curriculum Authority (QCA) and its subsidiary, the National Assessment Agency (NAA) taking a slightly less robust approach. Examining and testing bodies in Europe and elsewhere in the world see the use of electronic marking as 'innovative' and enabling change both in terms of improved processes and marking quality. The 'modernising' agenda seems to have slightly more constrained boundaries in this context than might have been the case a year ago.

1. Drivers for change

The drivers for change noted at the IAEA Conference in Abuja have continued, supported by those in the industry – both suppliers and users. The emergence of personalised learning and the development in the UK of the Framework for Achievement have begun to break down the existing paradigms relating to teaching and learning and how these are reflected in examinations and assessment.

The prospect of making use of on-line tests on a large scale has begun to raise the profile of graded assessment and testing when ready – not just from the educational perspective but also from the practical management of large-scale testing using PCs in UK general qualifications. In an interesting occasional paper published by *The Exam on Demand Assessment Advisory Group*¹, the prospect of a normal secondary school administering 80 on-line tests per day is raised. Clearly, changes to both the way that learning is organised and the regulatory requirements surrounding examinations will be required!

The development by a number of companies of e-Portfolio systems also raises questions about their nature and scalability especially in the context where learners could maintain portfolios for their lifetimes.

DRS has kept informed about these development whilst moving forward with its present aim of delivering change benefits now to as many students as possible through its e-Marker® applications. Whilst the importance and, it would be true to say now, the inevitability of large-scale on-line testing and portfolio management is not underestimated, the importance of providing now improved management and measurement information together with better and traceable marking reliability should not be overlooked. The current suite of applications is shown in **ANNEX 1**.

2. Developments since Abuja

Since the last paper presented at the IAEA Conference in Abuja, DRS has focused its attention in improving the quality control features of its applications and service. Two areas, in particular, include:

- the management of marking automated items;
- reviewing and controlling the seeding process with markers.

In addition, plans are now in place to process marks for 3.6m candidates in June 2006, with 1.5m being scanned and imaged. Additional functionality for processing examinations conventionally has been added that will reduce much of the administrative work involved in processing audio tapes for modern foreign language speaking components. Further marking of extended answers has taken place, with developments proposed to manage more easily candidates' work spanning many pages.

An extensive programme of familiarisation with the software and processes has been put in place by the main awarding body working with DRS which has been facilitated by a separate system being put in place to support training. This system operates independently from the live database and can support a number of management and marking scenarios that can be restored daily enabling over 1,200 examiners to be trained in a 10-day period.

3. Validity, reliability and bias

Every examination and testing organisation wishes to ensure that the assessments that are produced have the highest levels of validity and reliability and that any potential for bias is reduced to the minimum. These factors determine how well an examination is viewed. In this discussion, only between and within marker reliability is considered.

Extensive research has been undertaken in these areas in the context of conventional marking. A review of reliability that remains useful today was undertaken by the University of Nottingham in 1996 for the School Curriculum and Assessment Authority². Techniques for improving reliability are discussed, including (in the context of between and within marker reliability) double or triple-marking and removing as many sources of bias as possible. In the paper script environment this involves 'blind' marking of scripts more than once, random allocation of scripts to markers (to reduce the

effects of a very good or poor centre influencing the marker's judgement) and the removal of names from scripts.

However, there appears to be little research undertaken in the electronic marking environment. The potential for addressing the conventional constraints with electronic marking are considerable with increased flexibility brought about by storing images of candidates' answers which can be accessed simultaneously by more than one marker and routed through different processes.

This paper presents first a view on how electronic marking can improve marking reliability and reduce bias and second some initial findings from item-level data analysis. Proposals for future analytical work are included.

4. Definitions and implications for electronic marking

In the context of this paper, the meanings of the terms are defined as:

Validity	An examination is valid when the results provide information about the knowledge, skills and behaviours that the examination purports to measure.
Reliability	An examination is reliable if results can be replicated on a subsequent occasion. In this context, the interest in reliability relates to an individual marker's marking batch as well as different markers marking the same component.
Bias	An examination outcome is not affected by (a) systematic marking bias generated by one marker marking a batch of scripts and (b) other personal or social factors.

Conventionally, the balance between validity and reliability has been difficult to achieve. On the one hand, the use of multiple choice questions (MCQ) has enabled high reliability to be achieved but on the other hand questions which attempt to increase validity tend to require less structured answers which result in increased variability in marking. Whilst electronic marking does not purport to affect the validity of an item (as it is presented to a candidate in the same form as conventionally) the removal of constraints on how the images of the candidates' answers are processed can affect reliability. So, perhaps, there is a case for having our cake and eating it!

5. Reliability and electronic marking

A comparison of the conventional methods for addressing reliability and those that can be used with electronic marking can be summarised as follows:

Conventional marking	Electronic marking
Use of multiple choice answers removes marker bias and error at the marking stage	Use of 'auto-marking' allows short, single letter or word answers, to be integrated within a standard question paper and still gain the benefits of one marker making all marking judgements
Clerical checking of scripts and mark sheets to correct any errors	Marking of individual items with no clerical work involved. All marks totalled as part of the marking system
Single and double-marking of specific questions by dividing question papers into sections and marking in a marking centre	Segmentation of scripts into individual images for each item enables routing of items to specific markers and simultaneous double marking
Double (or triple) marking of complete scripts by copying candidates' scripts and sending to more than one marker	Segmentation of scripts into individual images for each item enables simultaneous double (or triple) marking (or marking complete scripts in this fashion)
Moving batches of complete scripts between markers and senior markers to check on marking quality periodically	Use of pre-marked 'seeded' items enables regular and real-time checking of marking standards

Clearly, the reduction in paper and paperwork is a welcome relief to many markers and the removal of the clerical tasks ensures that no errors of addition are made. The awarding bodies have also been able to reduce the number of staff carrying out clerical checks as addition no longer has to be scrutinised.

Sending individual items to markers that do not relate to an individual candidate also ensures that marking bias is reduced as no individual candidate's result is determined by the standards applied by

a single marker. In addition, any bias that may be introduced as a result of marking the scripts from a complete centre is removed.

Routing items to specific markers enables marking standards to be focused and tightly managed. With the regular monitoring of marking through the use of 'seeded' items any marker who is straying from the correct standard can be identified quickly and either re-trained or stopped from marking that item. (A fuller explanation of 'seeded' items is provided later in this paper.)

Double (or triple) marking becomes a simple exercise, with appropriate business rules being applied automatically where mark differences between markers exceed any agreed tolerances.

6. Auto marking

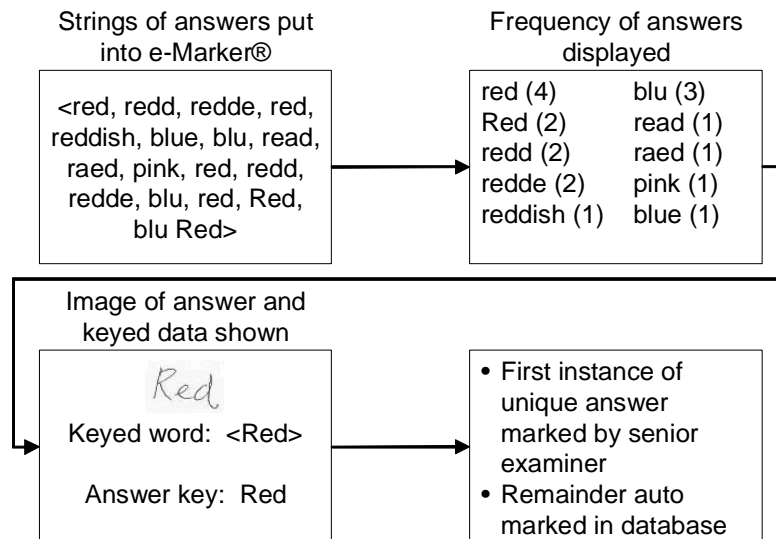
Simple, one word or single letter answers can be marked using imaging and keying techniques. For example, a question such as:

“What colour is used for post boxes in the UK?”

would make a good example of this type of question. The anticipated answer is, clearly, “RED”. However, acceptable answers might include allowable variants (such as “REDDISH”). Such minor variations would be difficult to take into account in a multiple choice item and would be subject to variation if all markers for a component were to mark the answers.

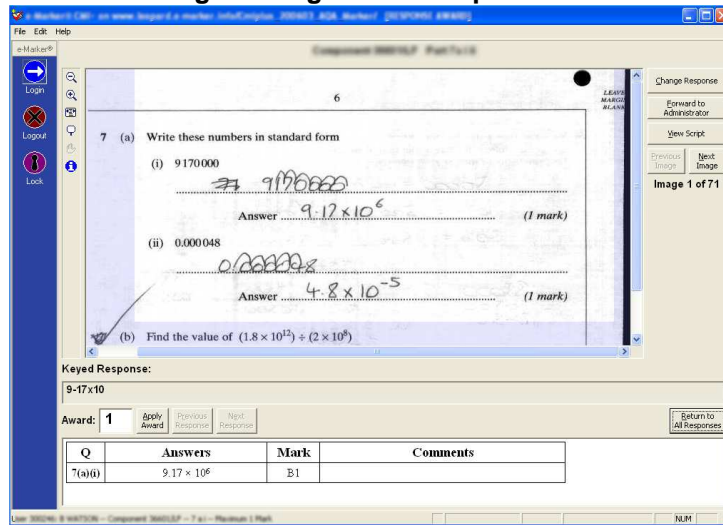
However, if all possible variants for an item were presented to a single, senior marker to make the judgements, acceptable variations in answers would be taken into account with consistent marking carried out by one person.

Auto marking enables this to happen. Images of each answer for all candidates to a question are taken, sent to a keying bureau to key into a data file (with appropriate quality checks) and then imported into the e-Marker® system. A senior marker is provided with a means of viewing the first instance of each unique answer presented and to mark it as correct or not. Once marked by the senior marker, all further instances of the answers are marked automatically by the database without further reference to the senior marker. The following process flow summarises the process.



Simple user interfaces are available, an example of which is shown in **Screen 1** below.

Screen 1
Screen showing marking of first unique auto marked item



As discussed in previous papers, not only is marking reliability enhanced through this mechanism, but information about what answers candidates give to such questions can be collated for both formative feedback to students but also improvements to item design.

7. General and expert marking

Items that cannot be marked using auto marking are routed either to 'general' or 'expert' marking. There is no difference in the functionality of the electronic marking software, but the markers will have different skills, experience and training.

The prime purpose of the division of work in this way is to enable the 'expert' markers to focus on the answers where specialist subject expertise is required. This brings the benefits both of enabling marking to be more consistent and less prone to bias as well as reducing the number of 'expert' markers required. For many awarding bodies this has underpinned their business case for making use of this marking system.

Reliability is improved by:

- maintaining consistent marking judgements over a period of time for each item, and
- not allowing the marking to be affected by answers to other questions in the question paper.

Some examples are given from the AQA *ICT A Full Course 3521/F* examination taken in June 2005 and marked using CMplus. Question 12d was marked by 'general' markers and question 16a was marked by 'expert' markers.

Question 12d – General Marker

(d) Give **two** differences between a LAN and a WAN.

Difference 1

.....

Difference 2

.....

(2 marks)

Question 16a – Expert Marker

16 A new Animal Care Centre is going to collect and store data on its clients. They have decided to store the data in three linked tables. The three tables they will use are:

- Owner Details
- Animal Details
- Billing Details.

(a) Explain why these tables should be linked together.

.....

.....

(2 marks)

In addition, ‘general’ and ‘expert’ marking both address the issue of bias mentioned earlier in more than one respect. First, systematic leniency or severity in marking will not affect individual candidates or groups of candidates in a batch of marking in the same way – as each candidate is marked by many markers. Second, marking judgements will not be influenced by name, location or social factors (as the items are anonymous). Current papers methods for hiding candidate details are cumbersome and wasteful of printing space. Electronic methods only present the candidates’ answers without any personal information associated with it.

8. Quality control with CMIplus

CMIplus makes use of seeded items to monitor marking quality. A seed is a live item that is given a mark by a senior marker at the start of the marking process and which is then used to gauge markers’ marking accuracy by presenting it to be marked without identifying its source. The application then makes a comparison between the seeded mark and that given by the marker.

The CMI+ application uses seeded items as follows:

- at the start of each marking day for each item;
- during the marking process for each item.

At the start of a marking day, a marker or general marker must qualify to mark on each item for that day. A number of seeded items are presented to the marker and a set number must be marked correctly to allow access to the queue of live items to mark. If the qualification criteria are not met, the marker is returned to the item selection menu where another item can be chosen – and qualification started. Once qualified, a marker can exit an item and return to it later in the same day without having to qualify again.

Once qualified, the marker starts marking live items, and pairs of seeded items are presented to a marker at intervals during the marking of the item queue. The intervals are set by the awarding body. The business rules governing criteria for seed failure are also set. In addition, the tolerance allowed between the mark given by the marker and the seeded item mark value can be varied depending upon the maximum mark available for the item and the nature of the subject being marked. An average mark difference can be derived together with an absolute mark difference as another measure of marking accuracy.

The application also allows for single and double-marking to take place (up to three times) together with adjudication on items where disagreement remains (a fourth view). Qualification and seeding can be added to this, if desired. A table of seeding combinations and marking strategies is given in **ANNEX 2**.

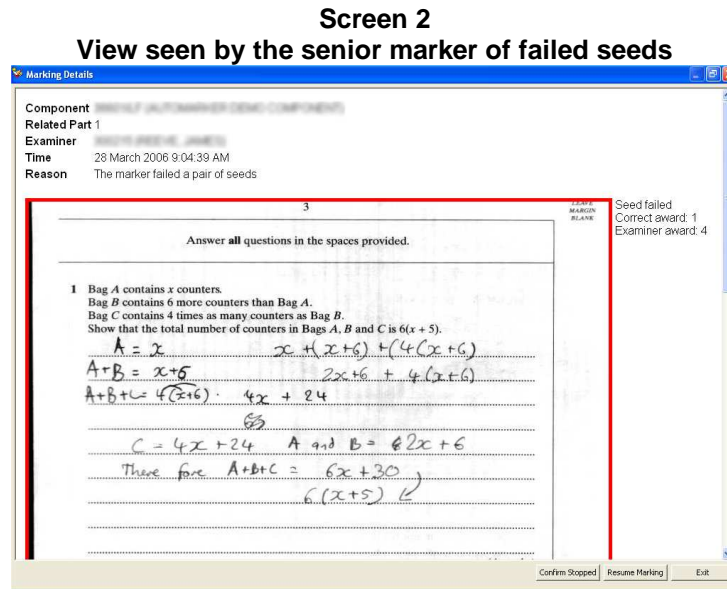
Where a marker fails the criteria for qualification or seeding during marking, the administrator receives notification. The details of the seeded item are provided together with the mark given. The seeded item can be viewed simultaneously by both the marker and the senior marker and contact made between the two to discuss the marking differences. If the senior marker is satisfied that the marker

understands why the marking difference occurred, he or she can reset the marker to be able to mark that item again.

This method of quality control allows:

- regular monitoring of marking standards;
- immediate feedback when a seeded item is failed;
- marking to continue with items where no quality problems exist with only specific problem item types being removed from markers;
- more comprehensive management data to be compiled to gauge overall marking quality – for quality control and for feedback to the markers.

Screen 2 shows one of the screens used by the senior marker to review the seeds failed by the marker.



9. Use of seeding data as a measure of marking accuracy

The capture of markers' marks for each seed presented provides a significant set of data with which to gauge marking accuracy. DRS has just started some analysis of data from the June 2005 examinations, some of which is presented below.

Each instance of a seed being presented to a marker has been recorded and the marking differences between the marker and seed mark noted. Any marking tolerance allowed has been identified. Initially, a simple analysis comparing the number and percentage of agreement has been carried out. Further, more extensive, analysis is planned. **Tables 1 and 2** below refer to the items given earlier in the paper.

Table 1
Marker marks compared with seed marks for Question 12d

Question 12d											
Tolerance = 0											
Variation from seed value	Examiner 1	Examiner 2	Examiner 3	Examiner 4	Examiner 5	Examiner 7	Examiner 8	Total	Percentage	Totals	
-3	0	0	0	0	0	0	0	0	0.00%	2.14%	
-2	0	0	1	0	1	0	1	3	0.19%		
-1	3	0	9	2	3	10	3	30	1.95%		
0	286	137	249	143	155	334	157	1461	94.87%	94.87%	
1	9	5	4	7	3	12	5	45	2.92%		
2	0	0	0	0	0	1	0	1	0.06%		
3	0	0	0	0	0	0	0	0	0.00%	2.99%	
Total	298	142	263	152	162	357	166	1540	100.00%		

Table 2
Marker marks compared with seed marks for Question 16a

Question 16a												
Tolerance = 0												
Variation from seed value	Examiner 1	Examiner 2	Examiner 3	Examiner 4	Examiner 5	Examiner 6	Examiner 7	Examiner 8	Examiner 9	Total	Percentage	Totals
-3	0	0	0	0	0	0	0	0	0	0	0.00%	1.50%
-2	0	0	0	0	0	0	0	0	0	0	0.00%	
-1	1	2	1	1	0	1	4	0	2	12	1.50%	
0	62	71	116	71	78	73	100	89	124	784	98.12%	98.12%
1	0	1	1	1	0	0	0	0	0	3	0.38%	
2	0	0	0	0	0	0	0	0	0	0	0.00%	
3	0	0	0	0	0	0	0	0	0	0	0.00%	0.38%
Total	63	74	118	73	78	74	104	89	126	799	100.00%	

Both questions have a tolerance of zero marks – in other words, any difference in mark between the seed value and the marker’s value would trigger the seeding rules to be applied.

The interesting points to note are:

- all instances of mark checking have been noted;
- the mark checking represents the total process undertaken at key points during marking and any seed differences would have been identified at the time that they occurred;
- the measure of marking agreement is very high.

10. Management information from CMiplus

General information that can be derived from CMiplus includes the:

- marking progress as marks are submitted online;
- number of each item marked by marker;
- number of seeded items marked by each marker by item;
- number and proportion of seeded items failed by each marker by item;
- number and proportion of items adjudicated (where double-marking is used);
- number and proportion of problems escalated by a marker to an adjudicator;
- progress of the completion of each item as marking progresses;
- various other management and exception reports to keep track of marking progress of all item types.

Some example of current management reports are provided in **ANNEX 3**.

Item level information not only can be used to judge marker performance and provide feedback at the end of the marking process, but it can also provide more information about how items perform in terms of maximum marks achieved and give the mean and standard deviation by item as well as the discrimination.

11. Further work

DRS hopes to engage in further work in this area of examining, using item response theory, and examine the reliability of dichotomous and non-dichotomous items. Given the nature of the mark capture, the use of very large data sets is possible which should provide useful and convincing evidence of the benefits of marking examinations in this way.

12. Summary

The progress since September 2005 has been considerable not only in the development of the marking system but in the recognition of the value of the information being derived. There is continued and growing interest in the approach in Europe and the rest of the world which is underpinned with a successful large-scale implementation in 2005.

DRS looks forward to helping deliver benefits to more organisations in the future, with targets in the UK alone amounting to over 4 m candidates marks to be captured in 2006 – of which over 1.5m will be using imaging.

Whilst e-Assessment and e-Portfolios are being scoped and developed for future use, benefits from electronic mark capture continue to be delivered now.

ANNEX 1

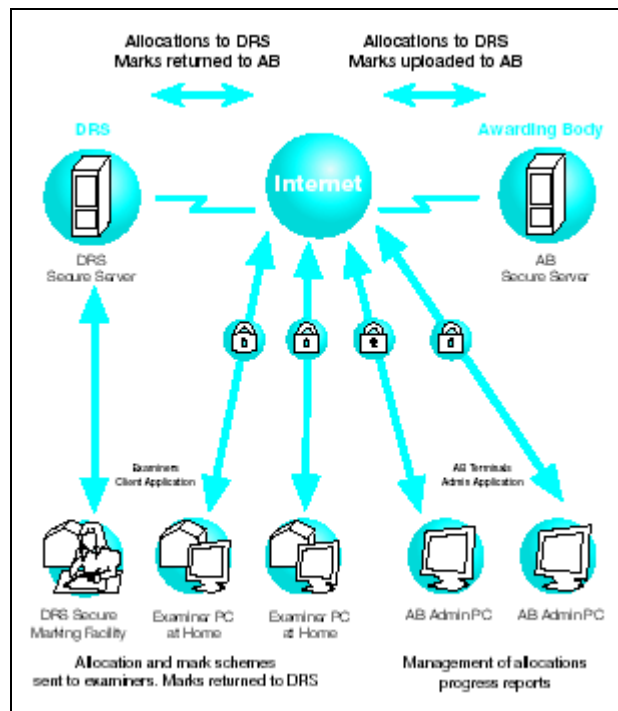
Description of e-Marker® applications

Capabilities

The applications have been designed to fit with awarding bodies needs – whatever the number of examinations or candidates are being marked. The internet suite of applications has been extended for 2006 and can be summarised as below:

On-Screen Marksheets (OMS)	<i>Allows the input of total component marks direct onto screen, replacing paper-based mark capture forms</i>
Question Marking from Script (QMS)	<i>Allows the input of item marks direct onto screen, once scripts have been marked</i>
Computer Marking from Multimedia (CMM)	<i>Similar to QMS, but allows the input of marks from audio tapes for speaking components</i>
Computer Marking from Script (CMS)	<i>Allows the direct marking of scripts onto screen, capturing item marks directly</i>
Computer Marking from Image (CMI)	<i>Allows the direct marking of images of complete scripts onto screen, capturing item marks directly.</i>
Computer Marking from Image ⁺ (CMI ⁺)	<i>Allows the direct marking of individual items directed to specific markers determined by marking capability and item type.</i>

An overview of the current system is shown in the following diagram:



Benefits for markers and awarding bodies

A summary of benefits of all applications mentioned is shown in the table below. The major benefits realised in 2005 relate to the detailed management information that can be derived from the CMI⁺ application. The item level data provides information for awarding bodies that was available

previously. A change to the way that the quality of marking is judged has also provided much closer control over marking standards in real time, as well as providing a more detailed analysis of marking quality.

Benefits	OMS	QMS	CMS	CMI	CMI ⁺
Real-time marking management	■	■	■	■	■
Identify anomalies and missing scripts earlier	■	■	■	■	■
Regular performance monitoring		■	■	■	■
No postage delays returning scripts to the awarding body				■	■
Faster transfer of marks	■	■	■	■	■
Auto totalling of marks		■	■	■	■
No answers can be missed		■	■	■	■
Mark parameters handled		■	■	■	■
Centralised mark schemes			■	■	■
Full image of script available				■	■
e-Sampling and seeding capabilities				■	■
No paper script sent to markers				■	■
Electronic re-allocation of scripts and items				■	■
Improved support for grade awarding				■	■
Item specialisation					■
Less call on expert marking					■
Automatic marking					■
Increased general marking					■
Escalation of marking problems to an adjudicator					■

A key benefit that underpins the business case for electronic marking is the ability to differentiate item marking by type and marking approach. This allows for the differentiation of the cost of marking as well as providing more information on the marking process.

The use of the administration application provided to awarding bodies provides access to detailed operating and quality information that leads to other benefits, as follows:

Script-based and CMI⁺ components

- set up of component parameters, marker types and rank and administrators;
- tracking of marking by total marks;
- tracking of sampling;
- matching of unexpected candidates with entry details.
- exporting of completed marks.

CMI⁺ components only

- tracking of marking by item;
- direct management of marking quality through seeding;
- image viewing for awarding and other purposes.

ANNEX 2

Description of seeding and marking combinations

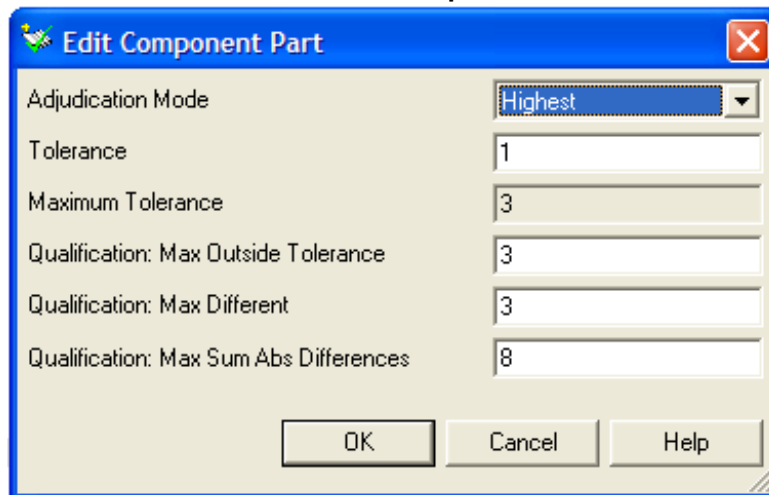
Seeded items are used in two ways:

- as part of the qualification process which a marker has to pass on each item on each new day's marking, and
- as part of the marking process where seeded items are introduced to the marker's queue of items to be marked (once qualification has been completed).

Table 1 summarise the types of uses for seeds.

Type	Detail of usage
Qualification	<p>A set number of seeded items are presented to a marker. Business rules are agreed with the awarding body on the number and criteria for success. For example, out of ten items presented, 7 out of 10 must be marked correctly to enable the marker to qualify.</p> <p>Other values relating to the number of qualification seeded items that can be marked differently from the seed value in a session and the maximum sum of the absolute differences between marks and seed values in a qualification session can also be set.</p> <p>Screen Example 1 shows how this is set up within the CMI+ administration application.</p>
Marking	<p>Pairs of seeded items are presented to the marker during the marking session. The 'gap' between the presentation of the seeded items can be set within the administration function. Two different business rules can be applied:</p> <ul style="list-style-type: none"> • rule 1 – where both seeded items have to be marked correctly to continue. If one of the pair is failed, then the marker is stopped; • rule 2 – where a set number of seeds has to be marked correctly from a group of pairs marked. For example, out of the last 10 seeded items marked, 7 must be marked correctly. <p>The parameters for setting the seed window values are expressed as a percentage, for example:</p> <ul style="list-style-type: none"> • 50% gives 2 items to mark then 2 seeded items; • 20% gives 8 items to mark then 2 seeded items; • 5% gives 38 items to mark then 2 seeded items.

Screen Example 1



A number of different marking strategies can be supported. In summary, these are:

- single marking – where only one marker's mark is captured. This can make use of qualification and seeding during marking as the quality control mechanism;
- double-marking – where more than one marker's mark is captured and compared with another marker – or more than one marker. This strategy also allows for adjudication where the values of the marks differ between markers and are outside tolerance.

Adjudication can be set to:

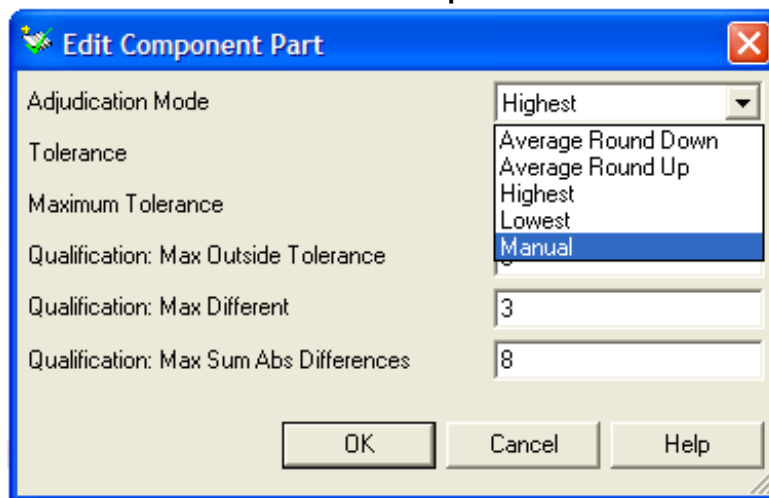
- manual – that is a senior marker views the item and previous markers' marks and makes a final decision;
- highest or lowest of the values given by the markers, when compared;
- the average of the two values given by the markers either rounded up or rounded down.

The different marking strategies are summarised in **Table 2** below:

Marking Strategy	Seeding Type	Adjudication Mode	Detail of usage
Single Marking	With or without qualification and marking seeding	Not applicable	Only one marker marks the item. Quality control is maintained by the qualification and seeding if it is used.
Double marking	With or without qualification and marking seeding	Mode set as required	The marks of two markers are compared and adjudication is either automatic (according to the mode set) or manual – where an adjudicator views any mark differences outside tolerance
Peer double-marking	With or without qualification and marking seeding	Mode set as required	The marks of two markers are compared. If they differ, the item is sent to another marker of the same rank. If two markers from the three agree, that mark is captured. If all three differ, the item is referred to adjudication (either automatic or manual).

Screen Example 2 shows how the adjudication mode is set.

Screen Example 2



ANNEX 3

Simplified examples of management reports

Component part completion report

Question	Marker Type	Number to be marked	Number marked	Number remaining	Average marking time (s)
A 1 a	General	6863	6852	11	1,636
A 1 b	Examiner	6863	6863	0	4,319
A 2 a	General	6863	6792	71	2,182
A 2 b i	General	6863	6791	72	2,214
A 2 b ii	General	6863	6861	2	1,263
A 3 a	General	6863	6863	0	1,923
A 3 b	General	6863	6795	68	879
A 4 a	Examiner	6863	6863	0	3,692
A 4 b	Examiner	6863	6863	0	2,599
A 5	Examiner	6863	5933	930	4,251
A 6	Examiner	6863	6863	0	2,843

Tracks the overall completion of marking by item and shows the average marking time for that question.

Quality of marking by marker

Marker Number	Question	Quota	Number marked	Number of seeds marked	Number of seeds failed	% of seeds failed	Number of times stopped for a question
000001	A1b	750	701	48	3	6.25	0
000001	A4a	750	693	50	0	0	0
000001	A4b	750	701	48	1	2.08	0
000001	A5	750	700	49	2	4.08	0
000001	A6	750	702	48	5	10.42	0
000001	A7b	750	702	48	1	2.08	0
000001	A7c	750	701	49	4	8.16	0
000001	A8a	750	688	62	4	6.45	1
000001	A8b	750	698	50	0	0	0
000001	A9	750	701	48	0	0	0
000001	A10 a b	750	702	48	2	4.17	0

Tracks the quality of marking as judged by the number and proportion of seeds failed by the marker, by each marker.

Quality of marking by related part

Question	Marker Type	Marker Number	Number of questions marked	Number of seeds attempted	Number of seeds failed	% of seeds failed	Number of times stopped for a question
A1a	General	000001	4790	265	0	0	0
A1a	General	000002	1856	105	0	0	0
A1a	General	000004	24	14	0	0	0
A1a	General	000006	217	33	0	0	0
A1a	General	000010	63	1	0	0	0
A1a	General	000011	0	0	0	0	0
A1b	Examiner	000012	0	0	0	0	0
A1b	Examiner	000013	701	48	3	6.25	0
A1b	Examiner	000015	690	60	6	10	1
A1b	Examiner	000016	681	67	2	2.99	0
A1b	Examiner	000017	0	0	0	0	0
A1b	Examiner	000018	686	50	2	4	0
A1b	Examiner	000019	688	62	3	4.84	0
A1b	Examiner	000020	0	0	0	0	0

Tracks the quality of marking as judged by the number and proportion of seeds failed by each question.

¹ *The development of e-assessment 2004 to 2014* by The Exam on Demand Assessment Advisory Group, September 2005.

² *A Review of Research into the Reliability of Examinations* by John Wilmut, Robert Wood and Roger Murphy, May 1996, University of Nottingham.