34th International Association for Educational Assessment (IAEA) Annual Conference – 7 to 12 September 2008

Supporting valid assessments in large-scale examination processing

Graham Hudson, DRS Data Services Limited, UK

1. Abstract

This paper will discuss how test and examination items can be managed in an electronic marking environment that enables more efficient large-scale marking processes to be operated without materially affecting validity.

This will be reviewed at the level of individual item types, including changes that have been made to recent examination papers. In addition, we will explore samples of examination types from a number of European countries to demonstrate the potential benefits to marking quality that this approach can bring when applied more widely.

The presentation will have a practical flavour, borne of experience of processing nearly 8,000,000 examination papers using this approach.

The paper will be given by Graham Hudson, National Business Development Manager for Education in DRS. Graham has twenty-five years experience of implementing and managing large-scale assessments within the UK, including the national curriculum tests and establishing a government-funded programme for implementing the use of new technologies in marking.

2. Background

2.1 Electronic marking provides a means of presenting candidates answers to examinations and assessment on-screen to enable a marker working in a marking centre or at home to award marks.

2.2 Candidates answers can be imported into the marking system from a variety of electronic sources. The most common are images generated from scanning paper examination scripts.

2.3 Electronic marking enables enhanced quality control to be applied as the restrictions of paper-based marking are removed. The use of images enables individual questions to be identified and electronically split (segmented) and marked separately, bringing the following benefits:

- **improved marking quality** – markers are able to concentrate on one question at a time;
- **marking bias is reduced and reliability increased** – more than one marker marks an individual candidates' script;
- **anonymity is maintained** – as markers do not know whose answer they are marking;
- **markers who mark outside set tolerances are stopped from marking** individual questions until they can be retrained;
- **improved feedback** to students, examining centres and awarding bodies.

2.4 More awarding bodies are turning to the use of electronic marking as the impact of these benefits is felt. Better management and transparency of marking is achieved as well as improved overall throughput. Evidence will be put forward in this paper for an improvement in marking quality as well as a consequential improvement in the setup and layout of examination and assessment materials.

2.5 **Research evidence** – research conducted by DRS and the NFER[1] on five GCSE subject examinations showed that there was:

- exact agreement between markers and the quality control items of between 79% and 99%;
- agreement to within one mark of between 91% and over 99%.

2.6 **National evidence** – there is evidence in the UK to support the improvement of marking quality at a national level which, if implemented more widely nationally or internationally could show significant quality improvements and act as an important driver of increased confidence in the examinations and assessment systems globally. Comparisons between examinations marked on paper in one year and then electronically the following year show that there was a statistically significant reduction in the proportion of grades changed following results queries and appeals.

3. Segmentation as an aid to quality improvement

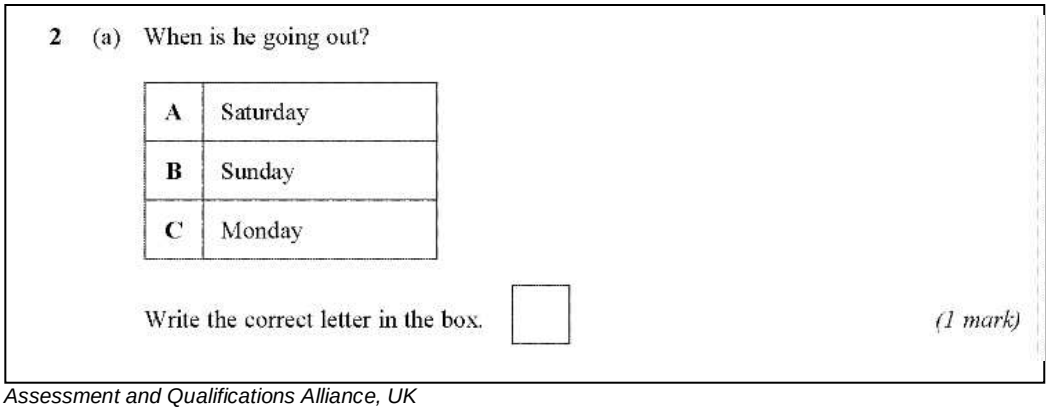
3.1 Segmentation is the key to improving marking quality with electronic marking. Images of candidates' scripts are divided up into individual questions or parts of questions (known as items) which can then be moved independently to different marking routes. This provides:

- the freedom to send the images of candidates' answers to markers simultaneously in order that double marking can take place. This approach to quality control is recognised as an effective way of improving marking quality by bringing more than one marker's marking opinion to bear on a piece of work;
- the ability to send items of different marking difficulty to specific marker types. This provides the means to allow markers with specialist expertise to be focused on marking the questions which require particular subject knowledge to interpret, leaving the more straightforward questions to well-trained general markers;
- a means of enabling each answer from a candidate to be marked by a different marker. This is a key plank in reducing marking bias which can result in either severe or lenient marking or, more problematic, erratic marking.

3.2 Segmentation of a complete script can be carried out on a range of item types and is particularly suited to papers where candidates write their answers on the question papers. Essay questions and longer-answers can be segmented, but this requires more careful management – which will be discussed later. A range of different item types are shown below to illustrate the range that can be covered.

Computer Assisted Expert Marking (auto marking)

Example Item 1: A Simple Auto Mark Question



2 (a) When is he going out?

A	Saturday
B	Sunday
C	Monday

Write the correct letter in the box. ☐ *(1 mark)*

Assessment and Qualifications Alliance, UK

3.3 By keying the images of the candidates' answers, these can be imported into e-Marker® where all unique instances of the answers are ranked according to the number of answers for

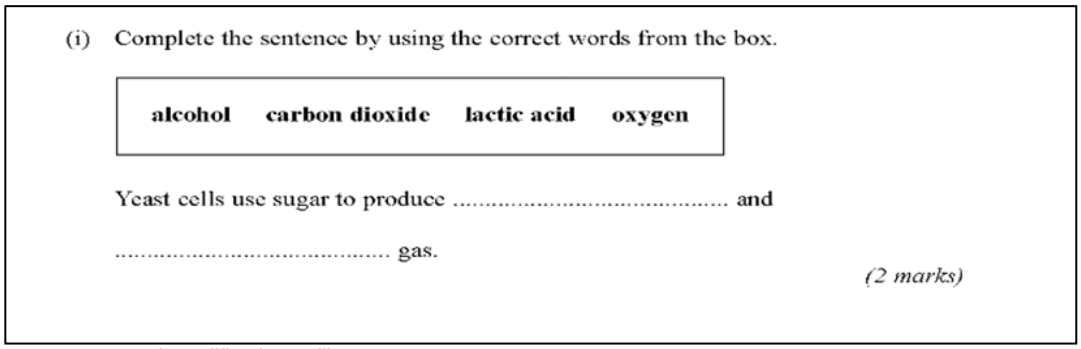
[1] National Foundation for Educational Research, Slough, UK

each. Each unique answer is awarded credit by a senior marker. Once this has been done, all subsequent answers imported into e-Marker® that match those awarded credit previously are marked automatically. This approach enables thousands of items that are 'not quite' multiple choice to be marked by one person.

3.4 In addition, it provides a record of every answer – not just the correct ones. It is therefore a rich mine of information available for feedback to students, schools, centres, examination boards and item authors on trends and of course the quality of the item itself.

General Marking

Example Item 2: A Simple General, Complete the Sentence Question



(i) Complete the sentence by using the correct words from the box.

alcohol	carbon dioxide	lactic acid	oxygen

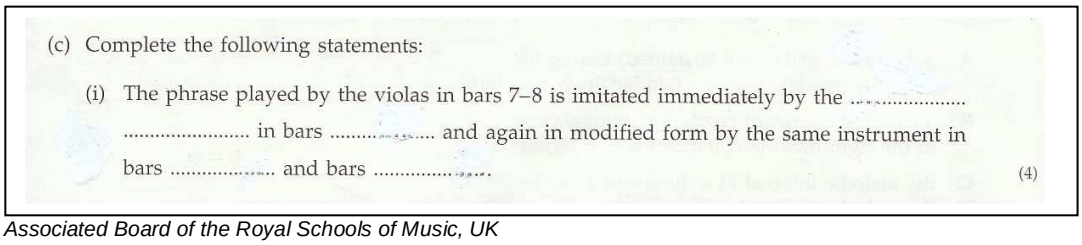
Yeast cells use sugar to produce ..... and

..... gas.

(2 marks)

Assessment and Qualifications Alliance, UK

Example Item 3: Another Simple General, Complete the Sentence Question



(c) Complete the following statements:

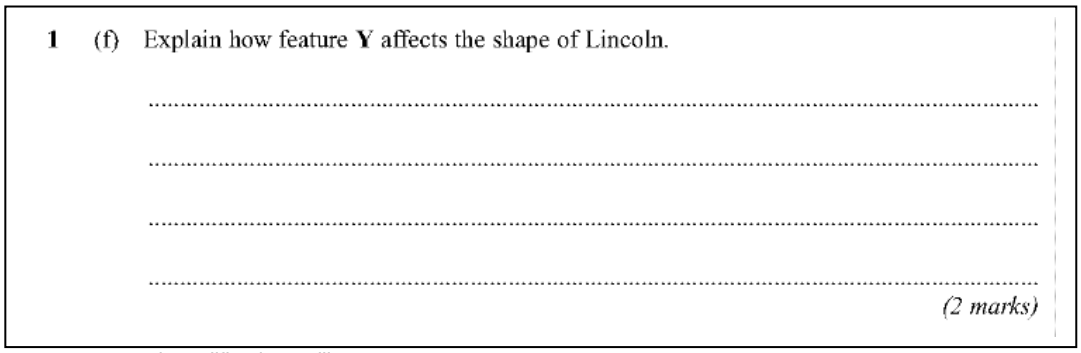
(i) The phrase played by the violas in bars 7–8 is imitated immediately by the .....
..... in bars ..... and again in modified form by the same instrument in bars ..... and bars ..... (4)

Associated Board of the Royal Schools of Music, UK

3.5 General markers are not necessarily subject experts, but can be trained to apply a well-structured marking scheme very consistently. It could be possible to mark Example Items 2 and 3 using Auto Marking, but DRS has found that the potential possible permutations makes awarding credit to unique answers a long exercise.

Expert marking

Example Item 4: A Short Sentence Free-Form Question



1 (f) Explain how feature **Y** affects the shape of Lincoln.

.....

.....

.....

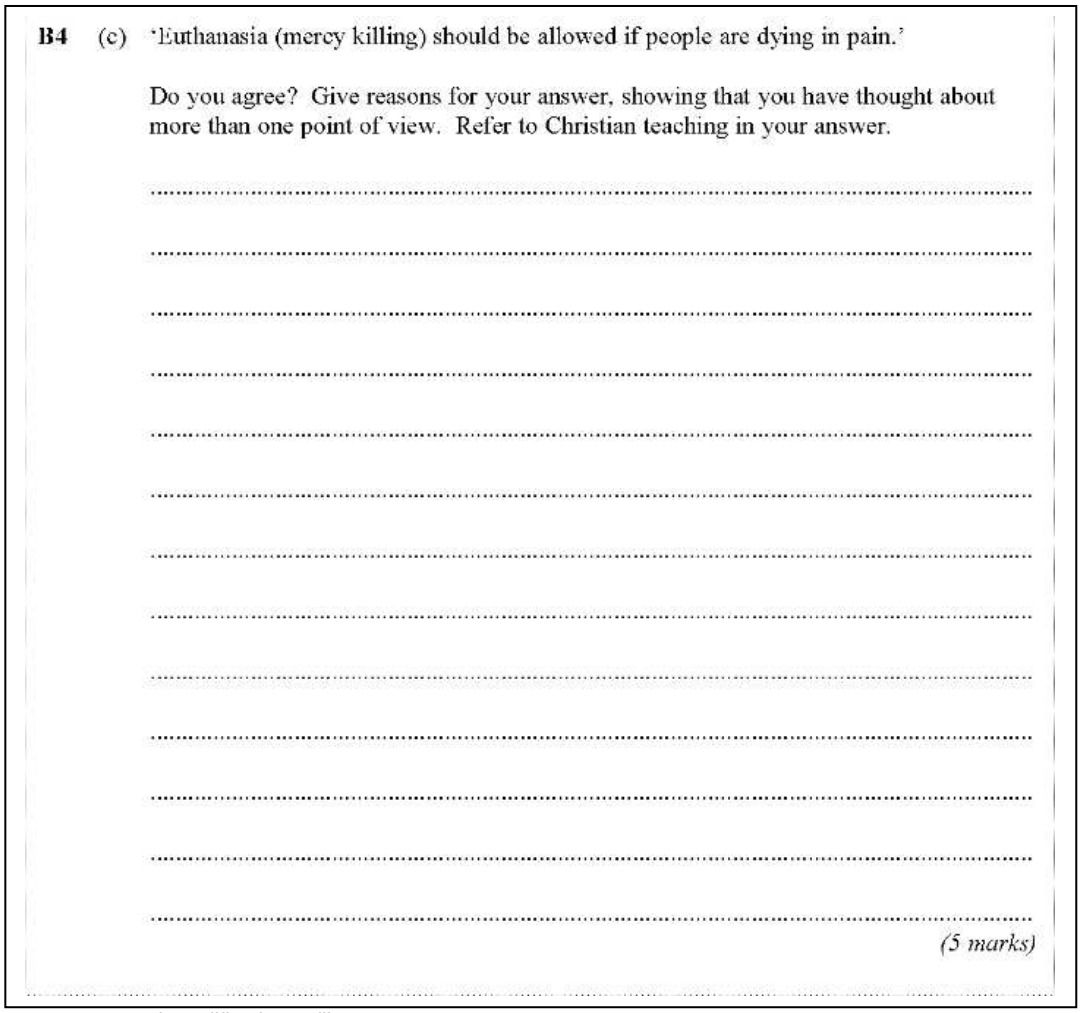
.....

(2 marks)

Assessment and Qualifications Alliance, UK

3.5 This type of question would be marked by an expert marker as it requires interpretation of the answer given by a candidate.

Example Item 5: A longer Free-Form Question



B4 (c) 'Euthanasia (mercy killing) should be allowed if people are dying in pain.'

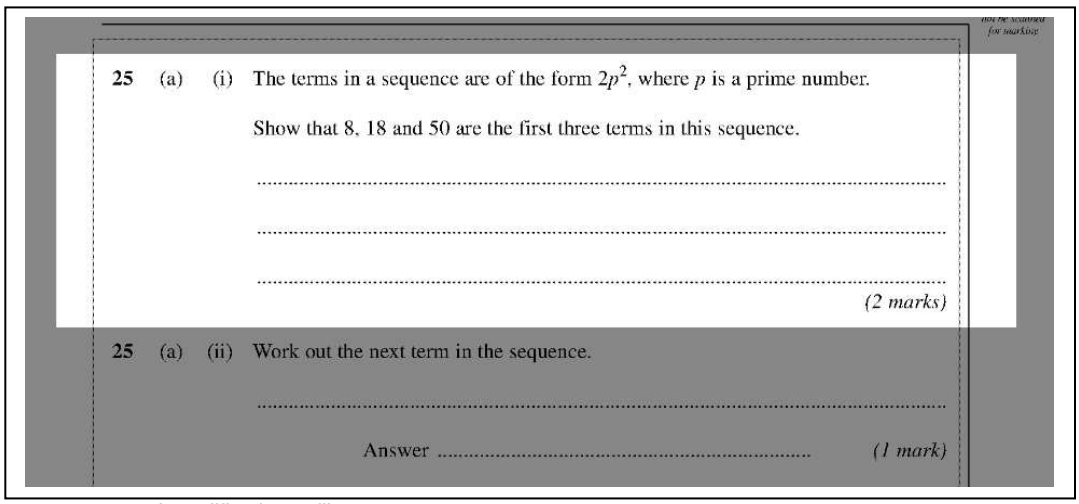
Do you agree? Give reasons for your answer, showing that you have thought about more than one point of view. Refer to Christian teaching in your answer.

(5 marks)

Assessment and Qualifications Alliance, UK

3.6 Again, this type of answer would be marked by an expert marker.

Example Item 6: A Related Part Question



25 (a) (i) The terms in a sequence are of the form $2p^2$, where p is a prime number.

Show that 8, 18 and 50 are the first three terms in this sequence.

(2 marks)

25 (a) (ii) Work out the next term in the sequence.

Answer ..... (1 mark)

Assessment and Qualifications Alliance, UK

3.7 This example illustrates how a marker is able to mark two parts of a question which are related. The marks for the individual item are captured, but the marker is able to see both parts as the marking decision for the second part is dependent on the answer given in the first part.

3.8 All these examples are taken from examination papers set for candidates in the UK. The paper will discuss the application of this approach more widely in Section 5.

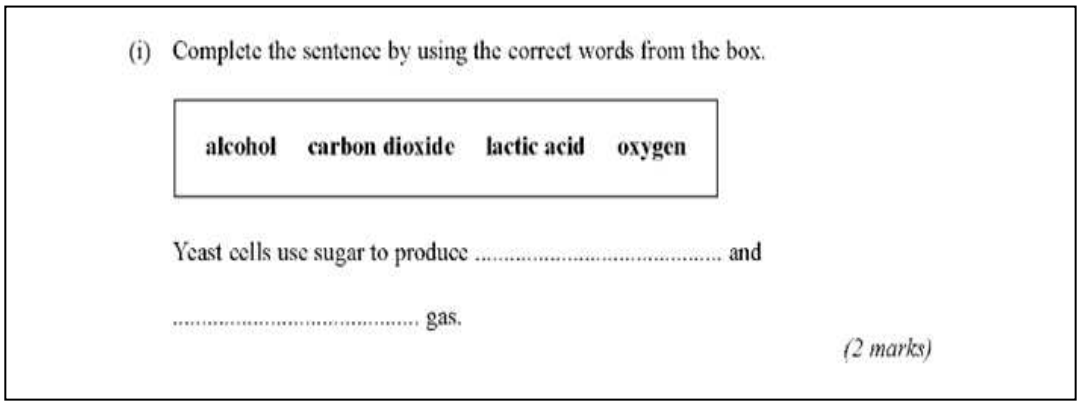
4. Changes made to improve usage of quality control measures and productivity

4.1 DRS has worked with a number of assessment bodies on the application of electronic marking to a variety of question types. Many question papers can be scanned and segmented without change, but DRS has found that small changes improve the way in which candidates answer the questions and the ease with which the items can be scanned and marked. This section will discuss some of the changes which have been made to improve processing in this way.

4.2 Some examples of items are given below that have been altered slightly to enable the full benefits of electronic marking to be realised.

General Marking

Example 1: General Item Prior to Change


(i) Complete the sentence by using the correct words from the box.

alcohol	carbon dioxide	lactic acid	oxygen

Yeast cells use sugar to produce ..... and

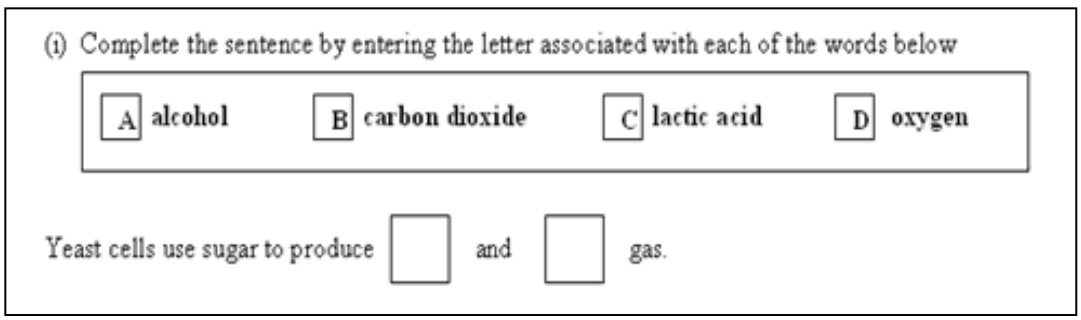
..... gas.

(2 marks)

Assessment and Qualifications Alliance, UK

4.3 This question would be marked by a general marker in this form. Small changes, which do not affect the item difficulty or validity can be made to enable it to be auto marked – realising the benefits of improved marking consistency and capture of all the answer types that candidates have given, and giving the feedback potential to stakeholders I mentioned earlier.

Example 1: General Example Changed to an Auto Mark Format


(i) Complete the sentence by entering the letter associated with each of the words below

A alcohol	B carbon dioxide	C lactic acid	D oxygen

Yeast cells use sugar to produce ☐ and ☐ gas.

Example 2: Expert Examiner Example Prior to Change

1 (b) (ii) Label **A** and label **C** are from ready-made chicken dinners.

Which dinner would a 'healthy eater' choose?

Explain your choice.

Label .....

Reason .....

.....

Assessment and Qualifications Alliance, UK

4.4 This is an example that shows that small changes to constrain the way in which candidates answer questions can enable a question to be marked more easily or, as in this case, to be marked by a general marker instead of an expert marker.

Example 2: Expert Examiner Example After Change to a General Format

1 (b) (ii) Label **A** and label **C** are from ready-made chicken dinners.

Which dinner would a 'healthy eater' choose?

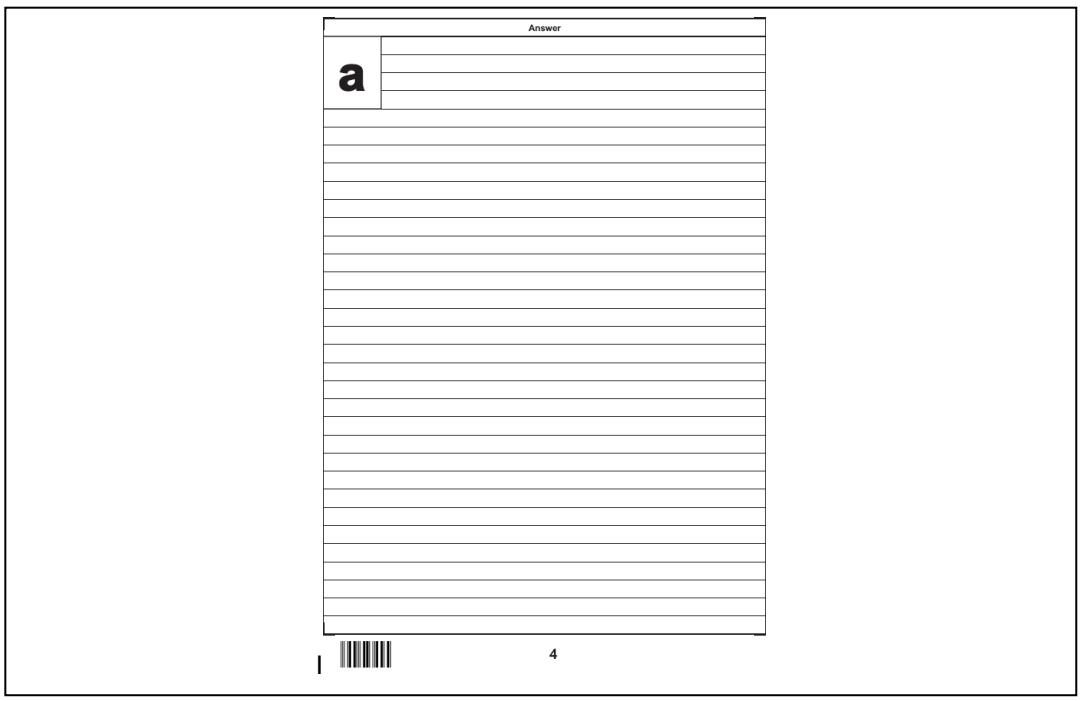
Explain your choice.

Label ☐

Reason .....

(2 marks)

Example 3: Change of Answer Booklet Format



4.5 In this example, the format of the answer booklet has been changed so that candidates answer a specific part on a specific page. This enables the segmentation of scripts to be supported more easily.

4.6 As well as improving quality control, changes such as these add to the benefits of speed and turnaround already provided by electronic marking. Some typical metrics from a recent large-scale examination processed in the UK illustrate this point:

Total number of scripts to mark	2,390,000
Total number of markers active	2,700
Total number of items to mark`	79,000,000
Total number of auto items to mark	17,500,000
Total number of general items to mark	11,700,000
Total number of expert items to mark	49,800,000

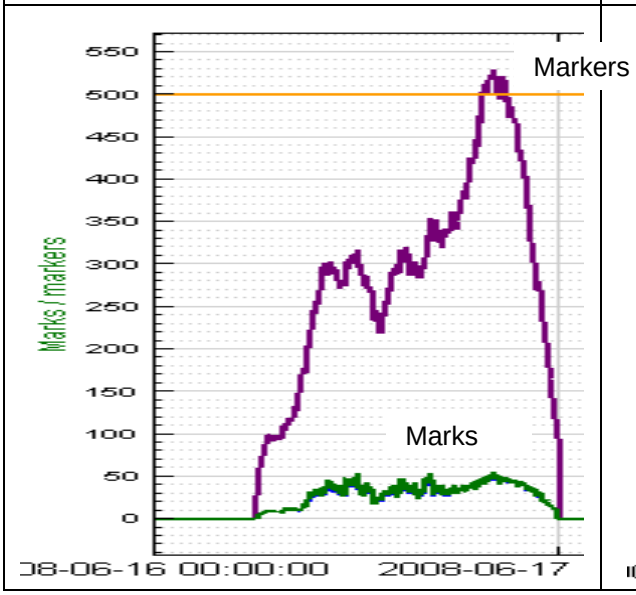
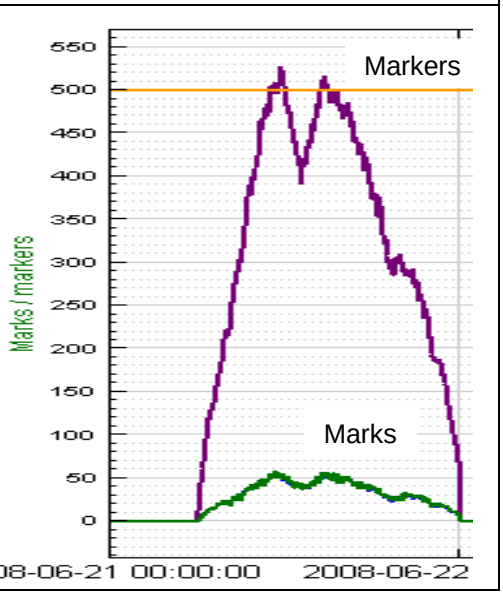
General markers have marked at rates up to **3,300 marks per hour** – or an average of nearly one mark per second.

Marks have been submitted at a rate of between **50 and 60 per second**, but the system can cope with over ten times that number.

In this examination series, total items marked in any one day (for general and expert) have exceeded **2 million**.

In one week, a total of nearly **17 million items** of all types have been marked.

4.7 In addition, information that allows better future planning of work and system capacity can be derived, such as marker marking profiles. The graph below shows two typical marking profiles – one for a weekday and one for a weekend day. The **purple line** represents markers active on the system and the **green line** represents the number of marks per second being submitted.

Typical Weekday Profile	Typical Weekend Day Profile
	

4.8 The benefits of having access to such information are clear:

- management information about marking rates and progress;

- information that is vital in planning capacity for future work;
- resources targeted to the right areas;
- increased confidence in the system – key performance metrics are constantly available.

5. How might this approach be employed more widely?

5.1 DRS has compiled survey information of examinations, assessments and tests from a number of European countries with a view to establishing how widely applicable this approach to marking items by segmentation may be and what scale of consequential benefits might result.

5.2 Countries have been chosen that have examination, assessment and testing processes and outcomes similar to those used in the UK, with question and item types of a format that could be marked electronically. Examples have been provided with proposals for how changes to the format could improve marking quality and marking rate.

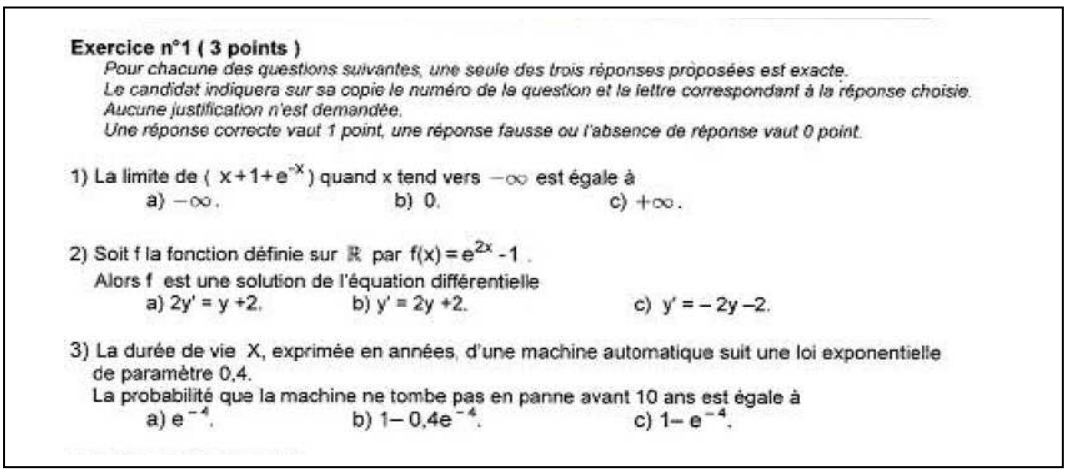
5.3 The sample of countries which have been included in this analysis are given in **Table 1** below. All the examinations listed have a key role to play in the education system of each country and require prompt and accurate marking. Cohort and population sizes have been included as we shall discuss later the potential impact of implementing electronic marking across countries such as these.

Table 1 – Information About National Examinations

Country	Cohort Size (k)	Examination Type	Estimated number of papers to mark
Country A	32,300	Leaving	129,200
Country B	36,600	Leaving	219,600
Country C	143,000	Leaving	715,000
Country D	53,000	Leaving	265,000
Country E	244,000	Leaving	1,464,000
Country F	71,000	Leaving	284,000
Country G	105,600	Leaving	633,600

5.4 Examples of questions that could be considered are given below. These are taken from examinations administered nationally and which are set at school leaving age. The questions as set are given with examples of how they could be adjusted to make them more easily marked within and electronic marking environment.

Example 1 from Tunisia – Before Adjustment



Exercice n°1 (3 points)

Pour chacune des questions suivantes, une seule des trois réponses proposées est exacte.
Le candidat indiquera sur sa copie le numéro de la question et la lettre correspondant à la réponse choisie.
Aucune justification n'est demandée.
Une réponse correcte vaut 1 point, une réponse fausse ou l'absence de réponse vaut 0 point.

1) La limite de $(x+1+e^{-x})$ quand x tend vers $-\infty$ est égale à
a) $-\infty$. b) 0. c) $+\infty$.

2) Soit f la fonction définie sur $\mathbb{R}$ par $f(x)=e^{2x}-1$.
Alors f est une solution de l'équation différentielle
a) $2y' = y+2$. b) $y' = 2y+2$. c) $y' = -2y-2$.

3) La durée de vie X, exprimée en années, d'une machine automatique suit une loi exponentielle de paramètre 0,4.
La probabilité que la machine ne tombe pas en panne avant 10 ans est égale à
a) e^{-4}. b) $1-0{,}4e^{-4}$. c) $1-e^{-4}$.

Example 1 from Tunisia – After Adjustment

Example #1 (Tunisia) – After Adjustment

Exercice n°1 (3 points)

Pour chacune des questions suivantes, une seule des trois réponses proposées est exacte.
Le candidat indiquera sur sa copie le numéro de la question et la lettre correspondant à la réponse choisie.
Aucune justification n'est demandée.
Une réponse correcte vaut 1 point, une réponse fausse ou l'absence de réponse vaut 0 point.

Réponse

1) La limite de ($x+1+e^{-x}$) quand x tend vers $-\infty$ est égale à
a) $-\infty$. b) 0. c) $+\infty$. ☐

2) Soit f la fonction définie sur $\mathbb{R}$ par $f(x)=e^{2x}-1$.
Alors f est une solution de l'équation différentielle
a) $2y' = y+2$. b) $y' = 2y+2$. c) $y' = -2y-2$. ☐

3) La durée de vie X, exprimée en années, d'une machine automatique suit une loi exponentielle de paramètre 0,4.
La probabilité que la machine ne tombe pas en panne avant 10 ans est égale à
a) e^{-4}. b) $1-0{,}4e^{-4}$. c) $1-e^{-4}$. ☐

5.5 The change enables the item to be marked more easily using the auto marking process.

Example 2 from Czech Republic – Before Adjustment

14. Každému členu pětičlenné party navrhl mistr odměnu 3 640 Kč. O kolik korun by mohl mistr zvýšit odměnu pro jednoho člena party, kdyby tato parta byla jen čtyřčlenná a mohl jí rozdělit stejnou částku peněz?

Example 2 from Czech Republic – After Adjustment

14. Každému členu pětičlenné party navrhl mistr odměnu 3 640 Kč. O kolik korun by mohl mistr zvýšit odměnu pro jednoho člena party, kdyby tato parta byla jen čtyřčlenná a mohl jí rozdělit stejnou částku peněz?

.....

.....

Odpověď: .....

5.6 This change enables the answer to be given on the question paper and enables it potentially to be marked by a general marker.

Example 3 from Portugal – Before Adjustment

2. Match the words/expressions in A with the corresponding definitions in B. Refer only to the numbers and the letters.

A	
1.	accent
2.	call centre
3.	miscommunication
4.	globalisation
5.	outsourcing
6.	proficiency
7.	asset

B	
a.	failure to express ideas or intentions clearly
b.	place where manual work is done, especially manufacturing or repairing
c.	to use labour from outside the company or business
d.	distinctive way of pronouncing words in a language
e.	something that gives you an advantage
f.	the established language usage of educated native speakers
g.	process of integration of economic, cultural, political and social systems worldwide
h.	level of skill or competence
i.	a place where people ask customers questions and/or receive their questions by telephone
j.	the usage of vocabulary that is characteristic of a specific group of people

Example 3 from Portugal – After Adjustment

2. Match the words/expressions in A with the corresponding definitions in B. Refer only to the numbers and the letters.

A		definition
1.	accent	
2.	call centre	
3.	miscommunication	
4.	globalisation	
5.	outsourcing	
6.	proficiency	
7.	asset	

B	
a.	failure to express ideas or intentions clearly
b.	place where manual work is done, especially manufacturing or repairing
c.	to use labour from outside the company or business
d.	distinctive way of pronouncing words in a language
e.	something that gives you an advantage
f.	the established language usage of educated native speakers
g.	process of integration of economic, cultural, political and social systems worldwide
h.	level of skill or competence
i.	a place where people ask customers questions and/or receive their questions by telephone
j.	the usage of vocabulary that is characteristic of a specific group of people

5.7 The addition of a column in which to place the appropriate letters will enable this item to be marked using auto marking.

Example 4 from Greece – Before Adjustment

ΟΜΑΔΑ Α

*Για τις προτάσεις από **Α.1** μέχρι και **Α.5** να γράψετε στο τετράδιό σας τον αριθμό της καθεμιάς και δίπλα σε κάθε αριθμό τη λέξη **Σωστό**, αν η πρόταση είναι σωστή, και **Λάθος**, αν η πρόταση είναι λανθασμένη.*

Α.1 Ένα από τα βασικά πλεονεκτήματα του καταμερισμού των έργων είναι ότι η μεγάλη εξειδίκευση οδηγεί και σε διάφορες βελτιώσεις του τρόπου με τον οποίο γίνεται η παραγωγή, δηλαδή σε διάφορες εφευρέσεις, και αυτό έχει ως αποτέλεσμα την αύξηση της παραγωγής.

Μονάδες 3

Α.2 Η εισοδηματική ελαστικότητα των κατώτερων αγαθών είναι αρνητική.

Μονάδες 3

Α.3 Η ζήτηση ενός αγαθού μεταβάλλεται προς την ίδια κατεύθυνση με τη μεταβολή της τιμής ενός συμπληρωματικού αγαθού (ceteris paribus).

Μονάδες 3

Α.4 Η καμπύλη ζήτησης με ελαστικότητα ζήτησης ίση με το μηδέν σε όλα τα σημεία της είναι ευθεία παράλληλη προς τον άξονα των ποσοτήτων.

Μονάδες 3

Α.5 Στη φάση της καθόδου του οικονομικού κύκλου παρατηρούνται μείωση της κατανάλωσης, στασιμότητα ή μείωση των επενδύνσεων, μείωση του εισοδήματος και της απασχόλησης.

Μονάδες 3

Example 4 from Greece – After Adjustment

ΟΜΑΔΑ Α

*Για τις προτάσεις από **Α.1** μέχρι και **Α.5** να γράψετε στο τετράδιό σας τον αριθμό της καθεμιάς και δίπλα σε κάθε αριθμό τη λέξη **Σωστό**, αν η πρόταση είναι σωστή, και **Λάθος**, αν η πρόταση είναι λανθασμένη.*

Σωστό* η *Λάθος

Α.1 Ένα από τα βασικά πλεονεκτήματα του καταμερισμού των έργων είναι ότι η μεγάλη εξειδίκευση οδηγεί και σε διάφορες βελτιώσεις του τρόπου με τον οποίο γίνεται η παραγωγή, δηλαδή σε διάφορες εφευρέσεις, και αυτό έχει ως αποτέλεσμα την αύξηση της παραγωγής.

Μονάδες 3

Α.2 Η εισοδηματική ελαστικότητα των κατώτερων αγαθών είναι αρνητική.

Μονάδες 3

Α.3 Η ζήτηση ενός αγαθού μεταβάλλεται προς την ίδια κατεύθυνση με τη μεταβολή της τιμής ενός συμπληρωματικού αγαθού (ceteris paribus).

Μονάδες 3

Α.4 Η καμπύλη ζήτησης με ελαστικότητα ζήτησης ίση με το μηδέν σε όλα τα σημεία της είναι ευθεία παράλληλη προς τον άξονα των ποσοτήτων.

Μονάδες 3

Α.5 Στη φάση της καθόδου του οικονομικού κύκλου παρατηρούνται μείωση της κατανάλωσης, στασιμότητα ή μείωση των επενδύνσεων, μείωση του εισοδήματος και της απασχόλησης.

Μονάδες 3

5.8 Again, the addition of boxes will enable this item to be marked using auto marking.

5.9 Analysis of these examination papers is given below, showing how the present papers could be divided up across electronic marking types. Clearly, in this analysis, we have had to make some assumptions that would need to be validated by discussion with the examination experts in each country, but we trust that this forms an interesting starting point from which potential benefits could be calculated – see Section 5.10. The split is shown as MCQ (Multiple Choice Questions), Auto Marking, General Marking and Expert Marking.

Country A

Examination paper	Total number of questions	% MCQ	% Auto	% General	% Expert
Mathematics	30	-	43%	0%	57%
Language and Literature	37	-	27%	24%	49%

Country B

Examination paper	Total number of questions	% MCQ	% Auto	% General	% Expert
Mathematics	40	75%	-	10%	15%
English	71	-	87.5%	11%	1.5%

Country C

Examination paper	Total number of questions	% MCQ	% Auto	% General	% Expert
Mathematics	30	-	0%	27%	73%
English	79	-	76%	22%	2%

Country D

Examination paper	Total number of questions	% MCQ	% Auto	% General	% Expert
Mathematics	17	-	29%	0%	71%
Language and Literature	32	-	31%	25%	44%

Country E

Examination paper	Total number of questions	% MCQ	% Auto	% General	% Expert
Mathematics	11	-	0%	0%	100%
Geography	66	-	29%	15%	56%

Country F

Examination paper	Total number of questions	% MCQ	% Auto	% General	% Expert
Mathematics (Testing 9)	30	-	67%	27%	6%
Language and Literature	64	-	55%	42%	3%

Country G

Examination paper	Total number of questions	% MCQ	% Auto	% General	% Expert
Mathematics	28	-	11%	0%	89%
French	8	-	0%	12.5%	87.5%

5.10 These datasets have been used to illustrate what might be achieved if electronic marking were introduced measured by number of person days saved in marking the total cohort of scripts. **Tables 2 and 3** below show the potential benefits that could be gained if electronic marking was carried out based upon the country analyses in Section 5.9. Benefits delivered will be dependent upon a number of factors, but conservative estimates have been used in this analysis.

5.11 The tables show the possible savings (in days marking effort) that could arise from the use of e-Marker® imaging and segmentation.

5.12 Cohort sizes have been taken from Section 5.3 above and an estimate of the potential number of auto mark type questions has been made. Some MCQ items have been deemed to be auto mark items for this purpose.

5.13 In order to estimate how long these questions would take to mark in total now, a marking time of **5 seconds per item** has been assumed. This has been converted into the number of total marking days required.

5.14 If the items were to be marked using auto marking, a time of **one hour per item** has been assumed for the senior marker to review the list of potential items. The analysis in **Table 2** is the result.

Table 2 – Potential Savings from Auto Marking

Country	Auto mark questions (marking days)			
	Now	Using e-Marker®	Potential Saving	Resource Saving (%)
Country A	344	29	315	91.6%
Country B	No change – currently use OMR			
Country C	4,965	25	4,940	99.5%
Country D	460	17	443	96.3%
Country E	No change – currently use OMR			
Country F	No change – currently use OMR			
Country G	220	9	212	96.4%

5.15 Clearly, this analysis would have to be tested in each case to ensure that the assumptions were appropriate. However, it serves to illustrate the order of potential savings that could result.

5.16 There are other, potential savings that could results from using electronic marking. General markers can be assigned items (as discussed above) that were previously marked by expert markers. This has been quantified by identifying the possible time of expert markers that could be made available for other work or for cost savings. A marking time of **15 seconds per item** has been assumed. This is shown in **Table 3**.

5.17 Also shown in **Table 3** are potential savings as a result of the removal of any clerical checking of the scripts after markers have completed marking. A checking time of **one minute per script** has been assumed.

Table 3 – Potential Savings from General Marking and Removal of Clerical Checking

Country	General Marker Savings (Marking Days)	Script-Checking Changes (Marking Days)
Country A	404	674
Country B	915	1,455
Country C	6,207	6,926
Country D	737	1,179
Country E	5,084	3,210
Country F	3,452	2,545
Country G	220	1,971

5.18 Again, the assumptions would need to be tested in each case, but if thought to be reasonable, considerable savings could be made as a result.

6. New quality models for long-form answers

6.1 Current quality control mechanisms that are used in e-Marker® fall into two types.

Type 1 is a traditional double-marking approach. The images of candidates' answers are marked by a first marker and the by a second marker. Where there is agreement between the two markers within the tolerance set, the appropriate mark is accepted. Where there is disagreement within the tolerance set, the candidates' answer is submitted to a more senior marker for adjudication.

Type 2 is the use of 'seeded items'. 'Seed items' are used in two ways – first at the start of each marking day to check that marking quality is correct before marking of an item is allowed; second, pairs of seeds are introduced at regular points during the marking to check that marking consistency is being maintained.

A mark tolerance can be set that reflects the degree of agreement required between a marker's mark and the standard mark set for the 'seed item'. For small value items, this is usually zero – in other words, the marker has to give the same mark as the standard mark. **Table 4** summarises the way in which seeds are used.

Table 4 – Summary of the Use of Seeds

Type	Detail of usage
Qualification	A set number of seeded items is 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 to mark any further items that day. 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.
Marking	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: • 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. The parameters for setting the seed window values are expressed as a percentage, for example: • 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.

6.2 However, for longer answers that may extend to several pages and employ multiple marking criteria as part of the assessment, there are drawbacks for both of these methods.

Double-marking is a proven approach to improving marking quality. However, it does require marking every item twice, which increases marking costs.

Using 'seeded items' could mean that a considerable proportion of marking effort is deployed in marking the 'checking' items and not the candidates' scripts themselves.

6.3 As a result, DRS is exploring ways in which these approaches can be combined but regular sampling of examiners' work. A business rule driven approach would mean that all examiners' were being checked to the same degree and regularly. A sampling proportion could be used that matches current approaches used by awarding bodies and examining

authorities. However, work would be checked in smaller quantities but more regularly. Marking that is not meeting appropriate tolerances would be detected before large numbers of scripts have been completed and corrective action in 'real time' can be taken.

6.4 Factors to take into account in this approach are:

- the frequency of sampling;
- what view is taken of the 'correct mark' – is this the standard decided by the Chief Marker and cascaded to all markers or is this derived by applying multiple marking opinions by a peer group of markers. (Both approaches can be taken, but will depend upon the view of the awarding body or examining authority over which approach to take.);
- what action should be taken when a marker fails one of the business rules;
- how much marking can a marker undertake whilst work is being sampled and before a judgement on marking quality can be reached;
- how much work requires review and possible adjustment if marking quality is marginally out of tolerance.

6.5 The key drivers here have to be:

- a valid and statistically sound model from which to work;
- early detection of marking outside tolerance;
- ensuring the markers can continue to mark without any delay;
- applying corrective action that is appropriate and valid based on the sampling evidence collected.

6.6 However, such an approach has to be supported by a scanning and imaging process that will segment individual questions from all those that have been completed by a candidate. As the answers tend to be free-form, written in a unconstrained standard answer booklet, this provides some challenges. DRS has wide experience in image management of this kind and is currently developing a process that will detect where individual questions are on a page, what question it is, undertake the segmentation and provide the necessary electronic instructions to e-Marker® to manage and display each question correctly. DRS would be happy to provide more details of this approach to individuals or organisations that wish to find out more.

6.7 These are all potential fruitful areas of research that will enhance and add to the benefits that can be derived from electronic marking.

7. What barriers to change might there be?

7.1 DRS has gained considerable experience in helping awarding bodies and examining authorities to manage change and transition from one, traditional approach to marking to the electronic environment. Typical barriers to change that have been encountered are:

- funding for the transition and implementation of a new system;
- managing change from one set of processes to another – in particular, the move from marking some scripts as a whole to marking them by item;
- regulatory matters in relation to comparability of marking standards from one year to the next and from existing approach to the new one;
- risks (practical and political) in moving from a 'tried-and-tested' approach to a new one that involves technology.

7.2 These are manageable barriers and can be addressed through an application of the business benefits of the changes that would be realised.

7.3 The above analysis shows that time can be saved, a rich vein of diagnostic and feedback information is available to students and assessment authorities and quality improved by the implementation of electronic marking. Engagement of key stakeholders in the right way would help with the transitions required. Business and policy stakeholders can been shown

the benefits of improved management control and ease of redirecting resources. Those with the pedagogical interest can be shown how quality and the reliability of candidates' outcomes can be improved.

7.4 The benefits to the overall educational system can also be shown which, in the case of the UK, has led the Examinations Regulator to adapt the 'Codes of Practice' to account for electronic marking without undermining the strength of the framework being applied.

7.5 Sufficient, experienced suppliers are now implementing electronic marking to show that risks can be managed, with good planning and a proven infrastructure.

7.6 Given that these areas can be addressed with demonstrable experience, the question has to be asked 'Why hold back the benefits?'.

8. Proven implementation

8.1 Much time and effort has been spent on ensuring that the technology does not become the barrier in itself to successful implementation of change.

8.2 In the largest implementation this year, DRS will have processed:

- 8 million marks using electronic data capture methods – both electronic marksheets and scanning;
- over 2 million marks per day using the imaging approach alone has been achieved;
- marking rates in excess of one mark per second have been achieved by general markers;
- a total of just under 79m marked items will be completed.

8.3 The capacity available now would allow for scaling of such volumes several times and can enable awarding bodies and examining authorities to engage in piloting to build confidence and aid change management, without the need to invest in high overhead costs at the outset.

9. Further information

9.1 Further information about the ideas and issues presented in this paper can be obtained by contacting:

Graham Hudson
National Business Development Manager, Education
DRS Data Research Services Limited
1 Danbury Court
Linford Wood
Milton Keynes
MK14 6LR, UK

Tel: +44 (0)1908 666088
Email: graham.hudson@drs.co.uk