RPA_08_RT_TR_019

THE IMPACT OF E-MARKING ON ENQUIRIES AFTER RESULTS 2006/2007

SUMMARY

Analysis of the enquiries after results (EAR) data from 2005/2006 suggested that the introduction of e-marking in June 2006 for 47 GCSE written components had positively impacted upon quality of marking (Taylor, 2007). This analysis was extended to incorporate the EAR data from 2007, to investigate how the introduction of e-marking had impacted upon quality of marking for 12 components marked electronically for the first time in 2007, and in particular, for those components that contained the longest responses that are currently marked on screen (GCSE Religious Studies A). Overall, the findings suggest that the introduction of e-marking in 2007 had no effect upon quality of marking; whilst there was a slight increase in the number of mark changes following a re-mark, this was statistically non-significant, suggesting it had occurred by chance. Analysis of the two Religious Studies components that contained longer response items produced similar findings; there was a small, but non-significant increase in the number of mark changes. This paper also considered the trial of electronic re-marking, and the findings suggest that electronic re-marking is associated with more mark changes than paper based re-marking. It is recommended that quality of electronic marking is monitored over time, although the introduction of re-marking electronically may make the continued use of EAR data invalid. This highlights the need to find other evaluative methods for assessing any impact of e-marking upon quality of marking.

BACKGROUND

Previous analysis of the enquiries after results (EAR) data from June 2005 and 2006 was undertaken to explore how the introduction of e-marking has impacted upon quality of marking for GCSE written components. The most important finding from this analysis was that components marked on paper in 2005 and e-marked in 2006 had significantly fewer mark changes following a re-mark when they were marked electronically, than when they were marked on paper. This suggests that e-marking had positively impacted upon quality of marking for these components. Further, there tended to be an increase in the number of re-mark requests over time, regardless of how a component was marked. This increase was not necessarily associated with an increase in the number of mark changes, suggesting quality of marking might not be the only factor that initiates a re-mark request.

To continue to monitor the relationship between e-marking, EAR and quality of marking, the analysis undertaken on the 2005/2006 EAR data was extended to compare data from June 2006 and 2007. Additionally, the impact of re-marking electronically was also considered, following a trial in June 2007 where 7 GCSE components were re-marked on screen. This report explores i) how the introduction of e-marking has impacted upon quality of marking for components e-marked for the first time in June 2007; ii) in particular, how the introduction of e-marking has impacted upon quality of marking for components that contain the longest responses currently marked on screen (GCSE Religious Studies A); and iii) how electronic re-

marking has impacted upon the number and size of resultant mark changes, for components included in the electronic re-marker trial in 2007.

COMPONENTS MARKED ON PAPER IN 2006 AND ON SREEN IN 2007

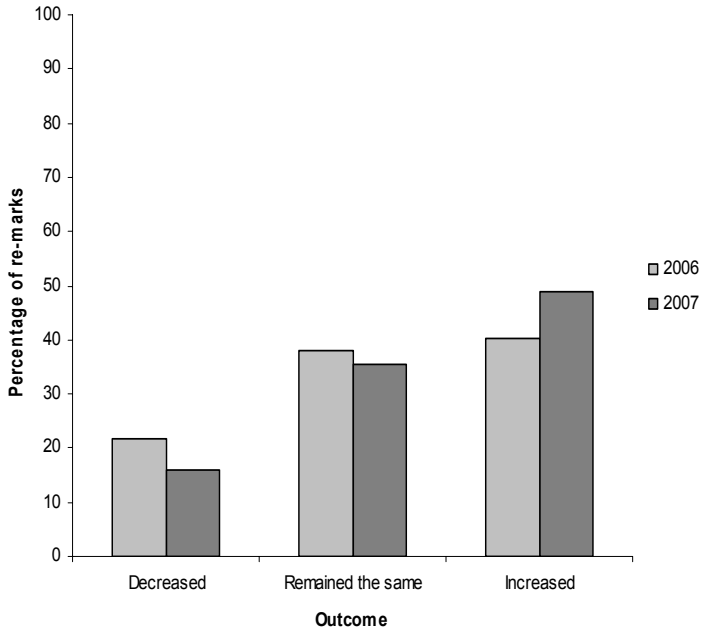
There were twelve GCSE written components that were e-marked for the first time in June 2007[1]. For each component the number of re-mark requests (as a percentage of that component's total entry), and the number of resultant mark changes (as a percentage of the total number of re-marks for that component) was calculated for June 2006 and 2007. None of these components were included in the electronic re-marker trial, meaning re-marks for these components were carried out on paper, at whole script level as in previous years.

The average percentage of re-mark requests for these components increased from 0.77 in 2006 to 0.85 in 2007. This is perhaps not surprising however, considering that the number of EAR tends to be increasing yearly, regardless of how a component is marked. Previous findings have also suggested that concerns over quality of marking might not be the only factor that initiates a re-mark request (Taylor, 2007); hence an increase in re-mark requests does not necessarily indicate a decline in quality of marking. It is also important to note that whilst the number of re-mark requests has increased in 2007, the average number of re-mark requests for these components remains relatively low, at less than 1 *per cent* of a component's total entry.

To assess how e-marking has impacted upon quality of marking a more informative measure is the outcome of re-mark requests, as a change of mark is likely to indicate concerns with the quality of the original marking. Figure 1 illustrates the outcomes of all re-mark requests for these components in 2006 and 2007, as percentages of the total number of re-marks in each year. As shown, these components had more mark changes (+2.63%) following a re-mark in 2007 when they were e-marked (64.68%), than in 2006 when they were marked on paper (62.05%). Despite this, a chi-square test[2] revealed that there was no significant difference in the outcomes of re-mark requests for these components between 2006 and 2007 ($\chi^2(1, N = 1484) = 1.09$, *ns,* $w = 0.03$, power > 0.995), suggesting the increase in mark changes in 2007 had occurred by chance. Therefore, it would appear that e-marking has had no statistically significant impact upon quality of marking for these components.

[1] There were 8 additional components marked electronically for the first time in 2007; these components were not available in 2006 however, so are not included in any analysis.
[2] For all chi-square tests re-marks resulting in an increase or decrease of mark were combined into one category (i.e. change of mark).

Figure 1: Outcomes of re-mark requests in 2006 and 2007 for components marked on paper in 2006 and marked electronically in 2007



The average absolute mark difference (AMD) following a re-mark for each component was also calculated for 2006 and 2007 (Table 1)[3]. Table 1 appears to show no consistent pattern of changes in the average absolute mark differences between 2006 and 2007 for these components; while some components have an increase in the average AMD, others have a decrease. Most of the differences are small however, suggesting they are likely to be due to chance fluctuations between years.

Table 1: Average AMD for components marked on paper in 2006 and marked electronically in 2007

Component	AMD 2006	AMD 2007	Difference
Religious Studies A Paper 1C	1.78	2.01	0.23
Religious Studies A Paper 2B	1.20	1.60	0.40
Applied Science Unit 2 Tier F	0.74	0.50	-0.24
Applied Science Unit 2 Tier H	0.92	0.60	-0.32
D&T: Electronic Products Paper F	1.30	1.41	0.11
D&T: Electronic Products Paper H	2.50	3.70	1.20
D&T: Electronic Products (SC) Paper F	1.00	*	*
D&T: Electronic Products (SC) Paper H	9.00**	1.00	-8.00
Mathematics A Paper 1 Tier F	0.44	0.22	-0.22
Mathematics A Paper 2 Tier F	0.45	0.59	0.14
Statistics Paper Tier F	0.98	0.85	-0.13
Statistics Paper Tier H	2.65	1.00	-1.65

* There were no re-marks for this component in 2007.
** There was only one re-mark for this component in 2006, which resulted in a change of 9 marks.

[3] It is important to note that whilst the components themselves had the same maximum marks for both years, these differ between components, meaning comparisons between components are not valid. Whilst it is possible to convert the mark changes to a percentage of a components maximum mark, a change of one mark would appear large for components with a low maximum mark, compared to components with a high maximum mark.

To summarise, there has been a small but non-significant increase in the number of mark changes for components e-marked for the first time in June 2007, compared to when they were marked on paper. It would appear that there have been no effects of the introduction of e-marking on quality of marking for these components.

COMPONENTS MARKED ON SCREEN CONTAINING LONG RESPONSE ITEMS

The GCSE components that contain the longest written responses currently marked on screen are from GCSE Religious Studies A; the written responses for items on these components are worth up to 8 marks. To explore any possible impact e-marking has had on these components, given the length of item response, the EAR data for the components was explored in more detail.

GCSE RELIGIOUS STUDIES A

Two GCSE Religious Studies A components were e-marked for the first time in June 2007 (3061/6/C and 3061/6/2B) and contain items worth up to 8 marks. Table 2 shows the number of EAR (as a percentage of a components total entry) and the number of resultant mark changes (as a percentage of the total number of re-marks for that component), for the two components in 2006 and 2007. Both these components had an increase in the number of re-mark requests between 2006 and 2007 (+0.28 and +0.32, respectively), although this is probably not unexpected or of any concern. There was also an increase in the number of resultant mark changes for these components in 2007 (+5.37% and +5.04%, respectively). However, chi-square tests revealed this increase was not statistically significant for either of these components ((χ^2(1, N =320) = 1.22, *ns, w* = 0.06, power > 0.995) and (χ^2(1, N =278) = 0.941, *ns, w* = 0.06, power > 0.995)). Therefore, although there was a small increase in mark changes, this is likely to have occurred by chance, meaning there appears to have been no effects of the introduction of e-marking for these components.

Table 2: Re-marks and resultant mark changes in 2006 and 2007 for Religious Studies A components

Component	EARs 2006	Resultant mark changes	EARs 2007	Resultant mark changes
3061/6/C	0.72	72.03	1.00	77.40
3061/6/2B	0.80	72.44	1.12	77.48

The average absolute mark change following a re-mark for the two Religious Studies components is shown above in Table 1. This shows that the average AMD following a re-mark increased slightly for both of these components in 2007; for 3061/6/C there was increase of +0.23 (from 1.78 to 2.01), and for 3061/6/2B there was an increase of +0.40 (from 1.20 to 1.60). Despite there being a slight increase in the average AMD in 2007, the average mark changes are still very small, given that the maximum mark available for both papers is 83. Furthermore, the changes in average AMD between 2006 and 2007 for both components are very small.

The findings for the Religious Studies components are broadly similar to those discussed above focussing on all components marked electronically for the first time; overall there was an increase in re-mark requests and an increase in the number of mark changes. However, the increases in mark changes were all non-significant, suggesting there have been no effects of e-marking on quality of marking. For the Religious Studies components there may also be other factors that have impacted upon quality of marking and influenced these findings. The

examiners marking Religious Studies online experienced considerable connectivity problems during the marking period, more so than most other subjects. This might have impacted upon their marking, due to the disruptions it would have caused.

ELECTRONIC RE-MARKER TRIAL

For the first time in June 2007 a selection of components marked electronically were also re-marked on screen; in previous years any re-marks of electronically marked components involved a printed copy of the paper being marked at whole paper level by one examiner. The examiner undertaking the electronic re-mark, usually the Principal Examiner or a Team Leader, is able to see the mark given by the previous examiner as usual, but marks at item level; consequently in larger components different examiners could potentially mark different items on one re-marked script. Seven GCSE components were re-marked electronically in June 2007; all of these components have been e-marked for at least one year previously, so did not contribute to the findings discussed above exploring how the introduction of e-marking in 2007 impacted upon EAR.

To explore how electronic re-marking has impacted upon the number and size of resultant mark changes, Table 3 summarises the distribution of absolute mark changes for these components in 2006 and 2007, as a percentage of the total number of re-marks in each year. As shown, there was a greater percentage of mark changes when components were re-marked electronically than when they were re-marked on paper, and this increase was found to be statistically significant ($\chi^2(1, N = 7044) = 22.15$, $p < 0.001$, $w = 0.06$)[4]. Note, however, that the size of the effect is small; indeed, it is the same size as the effect of e-marking on the number of mark changes for the Religious Studies components. The statistical significance of the effect is related to the large sample size.

Table 3: Distribution of absolute mark changes for components re-marked electronically in 2007

Absolute mark difference	0	1	2	3	4 and above (max = 14)
Percent 2006	48.46	33.01	12.79	3.78	1.97
Percent 2007	42.85	38.17	13.10	4.32	1.57
Change	-5.61	5.16	0.31	0.54	-0.40

The increase in mark changes when components are re-marked electronically is perhaps not surprising when it is considered that marking is done at item level, and particularly for larger components, different examiners may mark different items within the same script. Marking at item level increases the number of opportunities for examiners to award an additional mark, regardless of whether the same or a different examiner re-marks each item.

CONCLUSION

To summarise, the findings from the EAR 2006/2007 analysis suggest that the introduction of e-marking in 2007 for 12 GCSE written components has had no effect upon quality of marking.

[4] To conduct this analysis the outcomes of re-marks were grouped together into those that resulted in a change of mark, and those where the mark remained the same, despite data being available on the size of each mark change. This was deemed a more appropriate analysis given that comparisons of mark changes between components is difficult as the components have different maximum marks. Whilst it is possible to convert the mark changes to a percentage of a components maximum mark, a change of one mark would appear large for components with a low maximum mark, compared to components with a high maximum mark.

Although there was an increase in the number of mark changes in 2007 for these components overall, and for the GCSE Religious Studies components containing longer response items, these were all non-significant. This suggests the increases are likely to have occurred by chance.

Notably, however, the findings from this analysis contrast to those from the previous year. Components marked electronically for the first time in June 2006 had a significant decrease in the number of mark changes compared to when they were marked on paper, suggesting the introduction of e-marking had a positive impact upon quality of marking. Whilst the increase in mark changes in 2007 was non-significant, this change is in the opposite direction to that which is desired, and may relate to the connectivity problems some examiners experienced whilst marking. It is therefore recommended that quality of e-marking is monitored over time to continue exploring how e-marking impacts upon re-mark requests, resultant mark changes, and quality of marking. The introduction of electronic re-marking and other planned changes to the re-marking process may, however, make comparisons of EAR data over time invalid, making it necessary to find other evaluative methods.

REFERENCE

Taylor, R.C. (2007). *The impact of e-marking on enquiries after results*. AQA Research Committee paper, RPA_07_RT_TR_050

Rachel Taylor
April 2008
W:\Rachel\EARs\EARs 2007\EAR 2007 report_v3