# Extended essay marking on screen: Does marking mode influence marking outcomes and processes?

**Hannah Shiell, Martin Johnson, Rebecca Hopkin, Rita Nádas and John Bell** Research Division

## Introduction

Technological developments are impacting upon UK assessment practices in many ways. For awarding bodies, a key example of such impact is the ongoing shift towards examiners marking digitally scanned copies of examination scripts on screen rather than the original paper documents. This digital shift affords opportunities to manage and distribute information in ways that are not possible in paper-based marking systems, and this has important quality assurance benefits.

At the same time, however, the shift towards marking scripts on screen has prompted questions about whether the mode of marking might influence the outcomes of the marking process, particularly in relation to essay responses.

Research into comparisons between how people read texts on paper and computer screen suggests that the medium in which a text is read might influence the way that a reader comprehends that text. This is because some of the reading behaviours that support comprehension building, such as seamless navigation and annotation of text, are not easily replicated on screen (Dillon, 1994; Marshall and Bly, 2005; O'Hara and Sellen, 1997; Piolat, Roussey and Thunin, 1997; Rose, 2010).

Additional research also suggests that reading long texts can be more cognitively demanding on screen (Wästlund, Reinikka, Norlander and Archer, 2005), and that this extra demand can have a detrimental effect on how readers comprehend longer texts (Just and Carpenter,1987; Mayes, Sims and Koonce, 2001). In the context of examination marking, there might be concerns that such a mode-related effect might lead to essays being marked less accurately when marked on screen compared with when they are marked on paper.

The theoretical basis for concerns about mode-related influences on essay marking can be summarised by the model presented in Figure 1. This model outlines the potential relationships that are involved when an examiner reads an essay in order to mark it. In summary, literature underpinning the model infers that the shift from marking essays on paper to marking them on screen might be expected to impact upon examiners' manual and cognitive marking processes. This could, in turn, result in examiners having a weaker comprehension of essays when marking them on screen and this might be reflected in the final marking outcome.

Research in this area is therefore a principal concern for awarding bodies and stakeholders, posing potential implications in terms of both the defensibility of assessment outcomes and public trust in the assessment system.

In response to these concerns, researchers at Cambridge Assessment and elsewhere have been investigating how transition from paper-based to screen-based essay marking might influence examiners' marking behaviours and their marking accuracy. Four recent studies have investigated how mode might affect essay marking (Johnson and Nádas, 2009; Coniam, 2009; Fowles, 2008; Shaw and Imam, 2008). These studies,

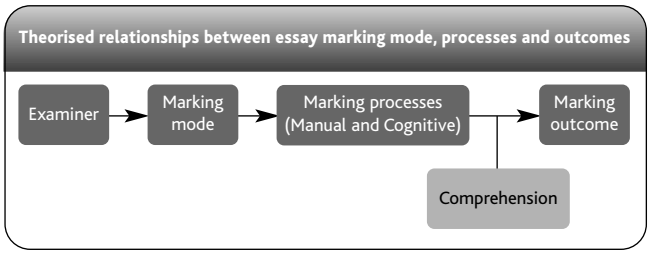


**Figure 1 : Theorised relationships between essay marking mode, processes and outcomes**

which consider essays of 150 to 600 words, report a negligible mode-related effect on marking accuracy; suggesting little cause for concern as the marking of digital essay images on screen replaces the marking of hard-copy paper essays.

Among the four studies, Johnson and Nádas (2009) is noteworthy as it employs a wider variety of quantitative and qualitative methods. The aim of the project was to broaden investigation beyond the singular consideration of the effects of mode on marking accuracy; to also explore mode-related influences on recognition of essay quality and examiners' marking processes.

As reported in Issue 8 of this journal, the findings of Johnson and Nádas (2009) showed that marking GCSE English Literature essays on screen had no significant effect on marker accuracy when compared with how they were marked on paper, although the examiners did exhibit different marking behaviours when marking in each mode.

The examiners in the Johnson and Nádas (2009) project also experienced significantly heightened cognitive workload levels while they marked on screen. The authors concluded that the examiners may have attained similar levels of accuracy across modes because they had sufficient spare cognitive capacity to accommodate the additional cognitive workload exacted by the screen marking task. Based on this conclusion, the authors suggested that the marking accuracy findings may not generalise to extended essays, therefore recommending that further research should explore mode-related effects in the marking of essays with lengths greater than those which were the focus of the earlier studies.

## Research questions and research design

To investigate further the potential links between marking mode and the outcomes and processes of extended essay marking, the current project replicated the Johnson and Nádas (2009) project, replacing GCSE essays with longer Advanced GCE essays.

The current project considered six research questions in three broad areas of enquiry, exploring mode-related influences on (i) marking outcomes, (ii) manual marking processes and (iii) cognitive marking processes. The six questions are displayed in Figure 2.

**Research Questions**

Mode-related influences on marking outcomes were considered through two research questions (RQs):

**RQ1:** Is examiner marking accuracy influenced by marking mode?

**RQ2:** Is examiner recognition of essay quality influenced by marking mode?

Mode-related influences on manual marking processes were considered through three research questions:

**RQ3:** Is examiner manual interaction (i.e. physical contact) with essays influenced by marking mode?

**RQ4:** Is examiner essay navigation influenced by marking mode?

**RQ5:** Is examiner annotation practice influenced by marking mode?

Mode-related influences on cognitive marking processes were considered through one research question:

**RQ6:** Is examiner cognitive workload influenced by marking mode?

**Figure 2 : Research questions**

This project used an essay question with a maximum of 60 marks available from an Advanced GCE History unit. One hundred and eighty essays from the June 2009 examination session were selected and split into two samples of 90 essays which were broadly similar in terms of mean marks (from the live session) and mark distributions. Table 1 shows the sample features of the essays used in the current project, compared to the sample used in the Johnson and Nádas (2009) project, which used GCSE English Literature essays.

**Table 1: GCE History and GCSE English Literature essay sample features**

| | *N* | *Written A4 pages* | *Written lines* | *Estimated word count* |
|---|---|---|---|---|
| | | *Mean* | *Mean* | *Mean* |
| GCE History project | 180 | 5.3 | 123.5 | 900 |
| GCSE English Literature project | 180 | 3.4 | 75.8 | 573 |

The 180 essays were blind marked on paper by the examination's Principal Examiner (PE) to establish a project reference mark for each essay. A sample of 12 Advanced GCE examiners participated in the project. The examiners were all relatively experienced, holding between 6 and 31 total years' experience (mean 16.8 years) of marking for large-scale educational assessment agencies in the UK. Five of the examiners had some previous experience of marking essays on screen.

The 12 examiners marked one of the two samples on paper and the other sample on screen. To control for essay sample and for marking order, a crossover research design was used and the examiners were randomly allocated to one of four examiner marking groups. Table 2 shows the crossover research design used.

**Table 2: Examiner marking groups and essay allocation design**

| *Examiner marking group* | *1st marking* | *2nd marking* |
|---|---|---|
| 1 | Sample 1 – Paper | Sample 2 – Screen |
| 2 | Sample 2 – Paper | Sample 1 – Screen |
| 3 | Sample 1 – Screen | Sample 2 – Paper |
| 4 | Sample 2 – Screen | Sample 1 – Paper |

Prior to marking, all 12 examiners attended a two day meeting to be trained in using the marking software and to standardise their marking in both paper and screen modes. Semi-structured interviews were carried out with each examiner after the marking period had finished, to allow the researchers to probe and check their understanding of the data.

# Findings

## Mode-related influences on marking outcomes

*RQ1: Is examiner marking accuracy influenced by marking mode?*

Marking accuracy was defined as the extent of agreement between the examiner marks and the corresponding PE reference marks. Marking accuracy was investigated by considering the differences between the examiners' marks and the reference marks awarded for each essay. These analyses considered two distinct measures of marking accuracy: *absolute*[1] and *actual*[2] mark differences. These measures give an indication of the magnitude and direction of marking accuracy differences between the examiners and the PE for each essay. Descriptive and general linear modelling statistical analyses were then used to investigate whether examiners' marking accuracy was influenced by marking mode.

Table 3 shows descriptive statistics of absolute and actual mark differences between examiner and PE marks by marking mode. Descriptive analyses of absolute mark differences revealed that in both marking modes half of all examiner marks were awarded within five marks of the corresponding PE reference mark. Given the 60-mark range available for the essays, this suggests close equivalence in the overall magnitude of marking accuracy on paper and on screen. Furthermore, a disparity of just 0.08 marks between mean absolute mark differences was identified across modes. Descriptive analyses of actual mark differences add greater depth to this picture. On paper the overall median absolute mark difference was 0 and mean absolute mark difference 0.02, indicating a balance of leniency and severity in marking. In contrast, on screen the overall median absolute mark difference was 1 and mean absolute mark difference 0.47, indicating a very slight tendency towards more lenient marking on screen.

**Table 3: Absolute and actual mark differences between examiner and PE marks by marking mode**

| | *Marking mode* | |
|---|---|---|
| | *Paper* | *Screen* |
| *N* | *1080* | *1067* |
| **Absolute mark difference** | | |
| Mean | 5.82 | 5.74 |
| Standard Deviation | 4.86 | 4.45 |
| Median | 4.5 | 5 |
| **Actual mark difference** | | |
| Mean | 0.02 | 0.47 |
| Standard Deviation | 7.59 | 7.25 |
| Median | 0 | 1 |

1. The absolute difference between an examiner mark and the corresponding PE reference mark. This measure assigns all differences a positive value, regardless of their direction. Absolute mark differences therefore provide a clear indicator of the *magnitude* of marking accuracy: smaller absolute mark differences represent greater marking accuracy.
2. The actual difference between an examiner mark and the corresponding PE reference mark. This measure assigns a negative value to marks below the reference mark and a positive value to marks above the reference mark. Actual mark differences therefore provide a useful indicator of the *direction* of marking accuracy: negative actual mark differences represent severe marking and positive actual mark differences represent lenient marking.

To enhance the descriptive outcomes, general linear modelling was used to test the statistical significance of any association between marking mode and marking accuracy (Table 4). No statistically significant association between absolute mark differences and marking mode was identified. This reiterated the findings of the descriptive analyses, confirming that there was no statistically significant mode-related difference in the overall magnitude of marking accuracy.

Analyses of actual mark differences suggested a significant association between marking mode and the direction of marking accuracy. Compared to the reference marks, essays marked on screen tended to be marked slightly more leniently than on paper, with screen-marked essays being awarded an average of 0.44 marks more than paper-marked essays. This small difference was statistically significant at the 5% level. Nevertheless, the effect size of this result, another statistical indication of the estimated strength of the relationship, was almost negligible (partial eta squared = 0.002), highlighting an extremely weak association.

Overall, the general linear models found no significant association between marking mode and the magnitude of marking accuracy, and only a small and extremely weak association between marking mode and the direction of marking accuracy. The findings therefore suggest that the examiners were marking with similar accuracy in both marking modes.

**Table 4: Results for general linear models of absolute and actual mark differences between examiner and PE marks**

| | | *ANCOVA table (N = 2147)* | | | | | |
|---|---|---|---|---|---|---|---|
| *Variable* | *DF* | *Model 1.1: Absolute mark difference* | | | *Model 1.2: Actual mark difference* | | |
| | | *Type III SS* | *F* | *p* | *Type III SS* | *F* | *p* |
| Marking mode | 1 | 4.23 | 0.26 | 0.61 | 106.10 | 4.14 | < 0.05 |
| Examiner | 11 | 789.19 | 4.34 | < 0.01 | 10481.91 | 37.20 | < 0.01 |
| Essay sample | 1 | 61.07 | 3.70 | 0.05 | 3002.49 | 117.20 | < 0.01 |
| Individual essay (nested in essay sample) | 1 | 13453.51 | 4.57 | < 0.01 | 54497.48 | 11.95 | < 0.01 |
| Error | 1955 | 32308.83 | | | 50083.57 | | |

ANCOVA, analysis of covariance; DF, degrees of freedom; SS, sum of squares

*RQ2: Is examiner recognition of essay quality influenced by marking mode?*

To investigate this question the features which the PE felt were contributing to essay quality were elicited using a modified Kelly's Repertory Grid method (Kelly, 1955). The PE then rated each of the sample essays against each of these essay features to generate a measure of quality for each essay. Finally, these measures were added to the marking accuracy general linear models to investigate whether examiner recognition of essay quality was equal across modes.

The marking accuracy findings from RQ1 indicated that, on average, the examiners marked essays with similar accuracy on screen as on paper. It was not possible to know, however, whether the examiners' recognition of essay quality was also similar across modes (for example, were the examiners better on screen at marking lower quality essays but worse at marking higher quality essays?). When a measure of essay quality was added to the marking accuracy models, analyses showed that examiner recognition of essay quality was not influenced by marking mode. In other words, the examiners marked high and low quality essays with equal accuracy on paper and on screen.

Together, the findings of RQs 1 and 2 support the conclusion that the accuracy of the examiners' extended essay marks and their recognition of essay quality are not influenced by marking mode, and that accurate and valid marking of extended essays is feasible on screen.

## Mode-related influences on manual marking processes

*RQ3: Is examiner physical interaction with essays influenced by marking mode?*

Data about how examiners tangibly interacted with the essays in both modes (e.g. how they physically touched the essays) were gathered through direct observation of one examiner from each of the four marking groups and augmented by interview evidence from all 12 examiners. The observed behaviours were:

- Tagging – physically holding a position in a text while looking at another text to relate two things;
- Overlapping pages in the line of vision;
- Dynamic Tracking – horizontal physical movement with a finger, pencil or mouse during reading;
- Static Tracking – vertical physical movement with a finger, pencil or mouse during reading;
- Pointing/Circling with a focus on one particular aspect (for example, a word) in the text.

The behaviour profiles gathered for the four observed examiners varied in terms of the number and variety of physical interactions that they used on paper and on screen, suggesting that these behaviours reflect highly personalised reading styles.

Overall, the four observed examiners physically interacted less with the essays on screen. Observation evidence suggested that examiners demonstrated fewer focused attention behaviours (i.e. indications that the examiner was attending to a particular word or piece of information; static and dynamic tracking and pointing/circling) on screen, whilst comparative referencing behaviours (i.e. indications that they were attending to more than one piece of information simultaneously; overlapping and tagging) did not alter across modes.

Some evidence from the examiner research interviews suggested that the increased tendency to interact physically with paper was because it was physically and mentally easier to do so in that mode.

*RQ4: Is examiner essay navigation influenced by marking mode?*

Data for this area of enquiry were also gathered via direct observation of the four examiners and interview evidence from all 12 examiners. The observations captured data about examiners' navigating behaviours while reading essays in both modes, specifically identifying the number of backward reading movements and movements of focus to other documents, such as mark schemes, question papers and other marked essays. Figure 3 shows the mean number of navigating behaviours per observed page by marking mode.

The observation evidence shown in Figure 3 suggests that examiners attended to the mark scheme, question paper or to other marked scripts relatively infrequently whilst marking, with no notable mode-related differences.

In contrast to the observation evidence, however, in the interviews six examiners suggested that they tended not to return to previously marked

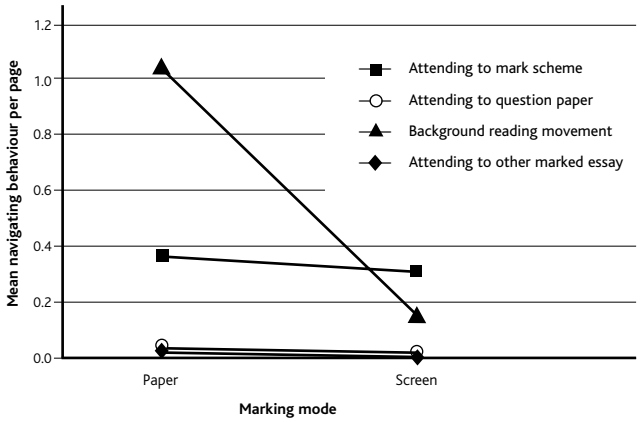


**Figure 3 : Mean number of navigating behaviours per observed page by marking mode**

essays as readily on screen. Examiners felt that this difference was due to such activity being more difficult to carry out on screen, for example:

> *"Well, I suppose I felt frustrated because it's so difficult…if you wanted to go back three scripts…I thought, 'Oh, can I be bothered with all this clicking and faffing and navigating it, and re-reading it and all this?', and I thought, 'No, I can't'."* (Examiner 8 interview)

Observation evidence also showed that examiners tended to read in a more linear fashion when marking on screen, with fewer iterative or backward reading movements. Examiners suggested in interviews that this was due to the relative difficulty of navigating around essays in this mode:

> *"It's an easier act physically just to turn the page over than to scroll back."* (Examiner 2 interview)

*RQ5: Is examiner annotation practice influenced by marking mode?*

Thirty essays from essay Sample 1 and 30 matched essays from essay Sample 2 were selected for annotation analysis. The 60 selected essays had each been marked by all 12 examiners and by nature of the research design, each examiner had marked 30 of the selected essays on screen and 30 of the selected essays on paper. Evidence of annotating behaviours was gathered through coded analyses of the marked essays. Again, these data were augmented by interview data from all 12 examiners.

The examiners were able to use a wider variety of annotations on paper than on screen. The screen environment allowed 17 annotation types, including a highlight/underline function. These annotations were built into the marking software following consultation with the examination's PE. For analyses purposes these annotations were termed the 'restricted' annotation palette. Any additional annotations used by examiners when marking on paper were termed the 'extended' annotation palette.

Figure 4 shows the differences in the use of annotations by mode and also by annotation palette. Comparing both the extended and restricted annotation palettes, examiners used an average of 35 annotations per essay on paper and 6 per essay on screen. A Wilcoxon Signed Rank test confirmed that this large mode-related difference was statistically significant ($z = -3.06$, $p < 0.01$, $r = 0.62$). Perhaps this finding is not surprising given that the examiners had access to a limited number of annotation types in the screen marking environment.

When analyses compared only the restricted palette annotations that were available in both marking modes it was found that examiners still annotated less on screen, with a Wilcoxon Signed Rank test confirming this difference to be statistically significant ($z = -2.82$, $p < 0.01$, $r = 0.58$). However, analysis at individual annotation level found that this difference was based on examiners using significantly more underline, sideline and tick annotations on paper. Therefore, when these three annotations were excluded from the overall analysis, there was no significant difference in examiners' use of the remaining restricted palette annotations on paper and on screen.

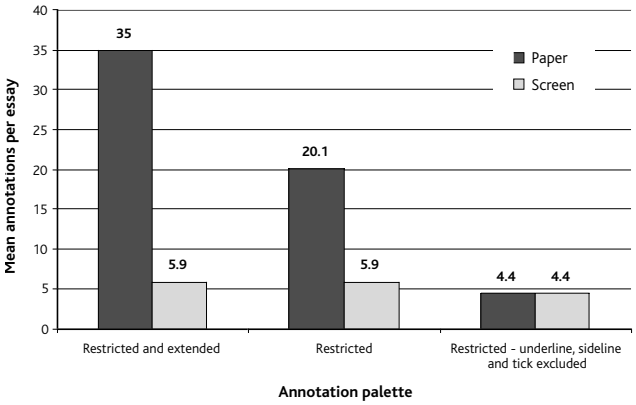


**Figure 4 : Mean annotations per essay by marking mode and annotation palette**

Examiner interview data were used to help explore the reasons for these mode-related differences. In interviews examiners suggested that they annotated less on screen because the process of using annotations was more difficult and that this might be related to issues of technical usability and their individual levels of proficiency at using the software. Reasons for more limited annotation on screen were also due, in part, to the way that the screen annotation palette sometimes lacked relevance for examiners.

Overall it was evident that physical marking processes were to a large degree idiosyncratic to individual marking behaviours. There was also a clear indication that mode influenced many aspects of examiners' manual marking processes. The physical interaction, navigation, and annotation behaviours that examiners employed for paper-based marking were more difficult for them to replicate when marking on screen.

## Mode-related influences on cognitive marking processes

*RQ6: Is examiner cognitive workload influenced by marking mode?*

Quantitative data about the levels of cognitive workload experienced in each marking mode were gathered using a modified version of the National Aeronautics and Space Administration Task Load Index (NASA TLX) (Hart and Staveland, 1988). The NASA TLX is a self-report survey designed to elicit subjective estimates of the cognitive workload experienced by an individual while performing a specific task. It is underpinned by the assumption that cognitive workload may be represented by a combination of six underlying factors: '*mental demand*', '*physical demand*', '*temporal demand*', '*performance*', '*effort*', and '*frustration*'. The NASA TLX survey was completed twice by 11 of the 12 examiners, midway through their marking sessions in each mode. The survey data enabled a statistical comparison of the cognitive workload

experienced by each examiner across modes to explore whether screen marking was more demanding than paper marking.

Analyses of these data revealed that the examiners experienced greater overall cognitive workload while marking on screen. A Wilcoxon Signed Rank test statistically confirmed that overall cognitive workload was significantly greater on screen ($z = -2.85$, $p < 0.01$, $r = 0.61$). The primary underlying sources of this mode-related difference were identified as the *physical demand* and *fatigue* factors.

Evidence from interview data suggested that the heightened physical demand experienced by the examiners during screen marking was attributed to three key areas of demand: using fine motor skills to operate the computer; maintaining a suitable position at the workstation; and looking at the computer screen. The latter of these physical demands, looking at the computer screen, was highlighted as the most common cause of the fatigue experienced by examiners whilst marking on screen. However, examiner interview comments suggested that this reflected their lack of familiarity with the marking software and might be expected to diminish as their experience of the marking software grows.

## Discussion

This project sought to investigate the feasibility of marking extended essays on screen by exploring the potential links between marking mode, essay marking outcomes and marking processes in three broad areas of enquiry;

(i) marking outcomes,

(ii) manual marking processes, and

(iii) cognitive marking processes.

It should be noted that the generalisability of the project findings might be limited by several factors. As a marking simulation exercise, the project differed from a true live marking session in the following key ways:

- The outcomes of the marking exercise had no consequence for candidates, which may have affected examiners' sense of responsibility.
- The marking exercise afforded a comparatively generous time allowance.
- The total marking allocation of 180 essays was comparatively light.
- The previous marking experience of the participating examiners was relatively high.

### Marking outcomes

This investigation aimed to consider whether examiners awarded marks which were equally close to the 'true' essay marks in both marking modes. Findings from the statistical analyses suggested that there was no mode-related influence on the magnitude of examiner marking accuracy, but a significant association between marking mode and the direction of examiner marking accuracy was identified. Screen-marked essays were, on average, awarded 0.44 marks more than paper-marked essays. However, the effect size of this result indicated an extremely weak association, and in the context of a 60-mark range the importance of less than half a mark difference is certainly debatable. In light of these perspectives, the findings presented no substantial evidence to indicate that overall marking accuracy was influenced by marking mode.

The examiners' recognition of essay quality across marking modes was also explored. Findings from the statistical analyses suggested that there was no mode-related influence on examiner recognition of essay quality. The examiners attended equally to essay quality when they marked in both marking modes, and the marks awarded recognised that quality.

Together, the marking outcomes findings support the conclusion that the accuracy of the examiners' extended essay marks and their recognition of essay quality are not influenced by marking mode, and that accurate and valid marking of extended essays is feasible on screen.

### Manual marking processes

When analyses shifted from marking outcomes to manual marking processes, mode-related influences became more pronounced. The examiners' manual marking processes were broken down into three separate processes: physical interaction, navigation, and annotation. Mode appeared to have an influence on all three of these processes.

The findings show that overall, the examiners physically interacted with essays less on screen than on paper, demonstrating fewer focused attention behaviours when marking on screen. The data did suggest, however, that examiners' physical interaction behaviours were highly personalised, varying widely across individual examiners. Again, when looking at evidence about navigation both within and across essays there were pronounced mode-related tendencies. Evidence showed that the examiners tended to navigate less iteratively on screen and read the essays in a more linear fashion. The most commonly articulated explanation for this difference was the relative difficulty of carrying out traditional paper-based navigation processes on screen.

The examiners in this study also used fewer annotations when marking on screen, due in part to the limited annotation palette available to them on screen. Although the examiners were trained in the use of the software annotation tools it was clear that the examiners still felt that the process of using annotations for marking on screen was too burdensome.

Despite these mode-related differences, examiners were still able to mark extended essays on screen with similar accuracy levels to their paper marking. This implies that the changes in manual marking processes induced by the shift in marking mode did not influence their marking outcomes.

### Cognitive marking processes

The examiners experienced greater cognitive workload when marking on screen and this was due to two particular factors – physical demand and fatigue. The examiners attributed the heightened physical demand during on screen marking to the use of fine motor skills to operate the computer, maintaining a suitable position at the workstation or looking at the computer screen. Looking at the computer screen was also highlighted as a common cause of increased and more rapidly arising fatigue.

It is possible that there is an inherent cognitive workload needed when long-held working practices are changed and individuals have to accommodate new ones. The screen marking software influenced examiners' marking processes and these changes could have been initially challenging for the examiners, requiring greater effort. Some of the heightened workload experienced by the examiners could be attributed to their lack of familiarity with the screen marking software, and therefore it is possible that the difference between cognitive workload levels reported across modes might be reduced as examiners' screen marking experience increases.

## Conclusion

Returning to the theorised links between extended essay marking mode, processes and outcomes (Figure 1), it appears that mode does have an important influence on some examiner marking processes, but that this does not necessarily influence their marking outcomes. The key practical implication of the findings of this project is that extended essays can be marked on screen without necessarily compromising accuracy. This project supports the conclusions of the Johnson and Nádas (2009) project, and quantitatively demonstrates that the marking of extended essays on screen is feasible. The finding that mode did not present a systematic influence on essay marking outcomes can help to reinforce the defensibility of those marking outcomes and contributes in some way to the maintenance of levels of trust in the assessment system. These findings are of great importance to educational assessment agencies and their stakeholders, and potentially opens the way to the expansion of screen marking to high stakes assessments involving extended essays.

**References**


Coniam, D. (2009). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. *Educational Research and Evaluation*, **15**, 3, 243–263.

Dillon, A. (1994). *Designing usable electronic text*. London: Taylor & Francis.

Fowles, D. (2008). *Does marking images of essays on screen retain marker confidence and reliability?* Paper presented at the International Association for Educational Assessment Annual Conference, 7–12 September, Cambridge, UK.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North Holland Press, 239–250.

Johnson, M. & Nádas, R. (2009) An investigation into marker reliability and some qualitative aspects of on screen marking. *Research Matters: A Cambridge Assessment Publication*, **8**, 2–7.

Just, M.A. & Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn and Bacon.

Kelly, G.A. (1955). *The Psychology of Personal Constructs*. New York: Norton.

Marshall, C.C. & Bly, S. (2005). *Turning the page on navigation*. Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, 7–11 June 2005, Denver, Colorado, USA, 225–234.

Mayes, D.K., Sims, V.K. & Koonce, J.M. (2001). Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics*, **28**, 367–378.

O'Hara, K. & Sellen, A. (1997). *A comparison of reading paper and on-line documents*. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, Atlanta, Georgia. ACM Press, New York, 335–342.

Piolat, A., Roussey, J-Y. & Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human-Computer Studies*, **47**, 565–589.

Rose, E. (2010). The phenomenology of on-screen reading: university students' lived experiences of digitised text. *British Journal of Education Technology*. DOI: 10.1111/j.1467–8535.2009.01043.x

Shaw, S. & Imam, H. (2008). *On-screen essay marking reliability: towards an understanding of marker assessment behaviour*. Paper presented at the International Association for Educational Assessment Annual Conference, 7–12 September, Cambridge, UK.

Wästlund, E., Reinikka, H., Norlander, T. & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: psychological and physiological factors. *Computers in Human Behavior*, **21**, 377–394.


EXAMINATIONS RESEARCH

# The effects of GCSE modularisation: a comparison between modular and linear examinations in secondary education


**Carmen L. Vidal Rodeiro and Rita Nádas** Research Division



*In this article, a summary of some key aspects and findings from a research project carried out to investigate the effects of modular assessment at GCSE level is presented. The research is described in depth in Vidal Rodeiro and Nádas (2010).*


## Introduction

GCSEs (General Certificates of Secondary Education) are the qualifications taken by the largest number of students in England. Over five million GCSEs were awarded in 2009, across a range of more than 40 subjects.

As part of the reform of 14–19 education, the national regulator in England revised the subject criteria for GCSEs in collaboration with teachers, awarding bodies, subject associations, higher education organisations and other interested parties. One of the main changes to these qualifications was the increase in the number of unitised or modular specifications.

Up to 2008, modular GCSE specifications were mainly confined to English, ICT, mathematics and science subjects, but since September 2009 almost all specifications are modular in structure, meaning that GCSEs are more in line with A levels, which have been modular since 2000.

A modular specification is one in which the content is divided into a number of units or modules, each of which is examined separately. Module examinations may be taken in different sessions (e.g. January, March, June) and any or all modules may be retaken if the student wishes, with the highest mark for each module retained. However, GCSE