

Essay Marking on Screen

Essay marking accuracy is reliable across modes

Educational assessment agencies in the UK have shifted towards assessors marking digitally scanned copies of exam scripts rather than the original paper documents. This shift has prompted questions about whether marking accuracy differs by marking mode. Cambridge Assessment research into examiners' marking practices on screen and on paper found that examiners can mark essays as accurately on screen as they can on paper.

Does mode affect examiner accuracy?

Literature suggests that reader comprehension might be weaker when longer texts are read on screen rather than on paper. This has important implications for assessment, implying a need to explore whether the mode in which an essay is marked might influence examiners' judgements. The aim of the research was to investigate whether examiners could mark a set of GCSE English Literature essays with equal accuracy on screen as they could in the traditional paper marking mode.

The project involved examiners marking two matched samples of essays, one on paper and the other on screen. To explore whether there were any mode-related influences on the marks awarded, examiners' marks were statistically compared against a reference mark for each essay. These reference marks were generated through a blind marking process involving two principal examiners.

Table 1 presents a descriptive analysis of the overall actual mark differences for all examiners, by marking mode. Actual differences between examiner marks and reference marks for the same essays give a measure of the direction of marking accuracy, with negative differences showing severe marking and positive differences showing lenient marking. Analysis suggested that there was no mode-related effect with mean and median actual mark differences very closely matched in both modes.

Table 1: Actual mark differences between examiner and PE reference marks by marking mode

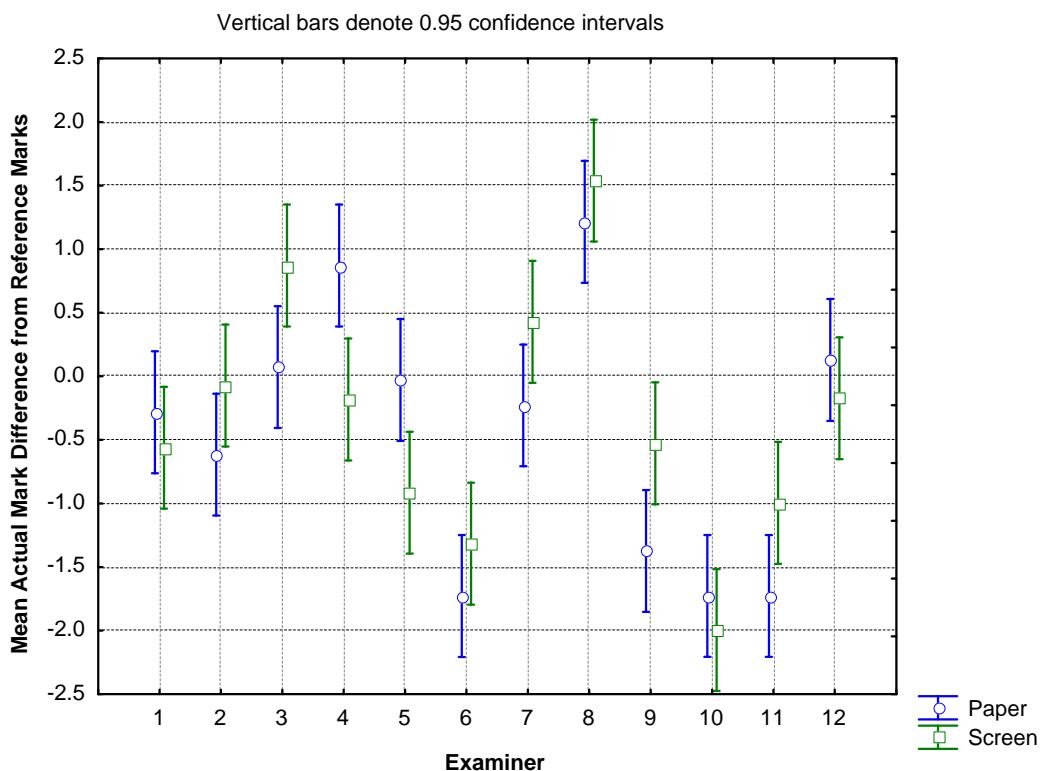
	Marking mode	
	Paper	Screen
N	2160	2160
<i>Actual mark difference</i>		
Mean	-0.45	-0.34
Median	-0.5	-0.5

To explore whether there were any particular mode-related effects on individual examiners, further statistical analyses were carried out using general linear modelling. This method was able to take into account different aspects of the research design, for example, whether examiners marked on screen first or second.

Figure 1 (overleaf) shows the results of this statistical modelling and presents a visual representation of the relationships between individual examiners and marking mode. Again, negative values indicate severe marking and positive values indicate lenient marking.

The confidence intervals overlap for all examiners except for examiner 4, suggesting no significant mode-related marking difference for 11 examiners. With the single exception of examiner 4, where an examiner was severe or lenient in one mode they were similarly severe or lenient in the other mode.

Figure 1: Least square means for actual mark difference by examiner and mode



These findings show that the examiners were able to mark extended essays on screen with similar accuracy levels to those essays that were marked on paper. This suggests that the reading comprehension levels of the examiners were not affected by the mode in which the essays were read.

This research has important implications for assessment. At a time when the role and impact of technology in education generates continuous debate it is imperative that e-assessment solutions are robust, reliable and valid. Therefore, educational research such as this plays a vital role in the continuous improvement of education and assessment policies and practices.

Context of research

A number of recent studies have considered whether the mode in which essays are marked influences the reliability or accuracy of the marking of those responses. These screen marking studies represent an important step forward in helping to develop greater understanding of how mode might affect examiners' marking practices.

In this project 180 GCSE English Literature essay scripts were selected and divided into two samples of 90, broadly matched by mark distribution. The essays were blind marked by the principal and assistant principal examiners to establish a reference mark for each essay. Twelve examiners with no prior experience of marking on screen then marked one sample of essays on paper and the other on screen. Quantitative and qualitative methods were used to gather and analyse the research data.

The findings of this study have been published and presented widely and have provided a valuable contribution to knowledge in the field of on screen marking. Further research is ongoing to investigate whether the findings of this study also apply to extended essays at Advanced GCE level.

Further information

Full details of the essay marking on screen work by Cambridge Assessment are available at: www.cambridgeassessment.org.uk

Contact: Martin Johnson
Research Division
Cambridge Assessment
1 Regent Street
Cambridge
CB2 1GG

Tel: 01223 553843
Email: martin.johnson@cambridgeassessment.org.uk