Essay Marking on Screen
## Cognitive workload is greater on screen

Electronic marking of digitally scanned images of exam scripts is now used by all major UK educational assessment agencies. Although it is not yet replacing traditional paper-based marking in all contexts, marking on screen represents a significant development. Measures of examiner cognitive workload in this project suggested that examiners with no prior experience of marking on screen had to work harder to achieve the same results when marking on screen as on paper. It was also found that marking on screen was more frustrating than traditional paper based marking, particularly in the earliest stages of marking for the examiners in this study.

Is cognitive workload greater for marking on screen?

Although previous research has investigated the effects of mode on the reliability and accuracy of essay marking, none has investigated the effects of mode on examiners' cognitive workload during marking. The project found that marking mode had no systematic effect on examiners' marking accuracy. However, it was considered important to investigate the toll it takes on examiners to mark to the same high standards on screen as they do on paper. The effects of mode on examiners' cognitive workload were therefore measured, recognising that rising levels of cognitive workload could have detrimental long-term consequences in terms of examiners' efficacy and satisfaction.

Examiners' cognitive workload was measured using an adapted version of the National Aeronautics and Space Administration Task Load Index (NASA-TLX). This tool gathered information about six dimensions of cognitive workload: mental demand; physical demand; temporal demand; performance; fatigue; and frustration.

Figure 1: Mode-related differences in examiners' weighted ratings on the NASA-TLX individual dimensions and overall cognitive workload measure
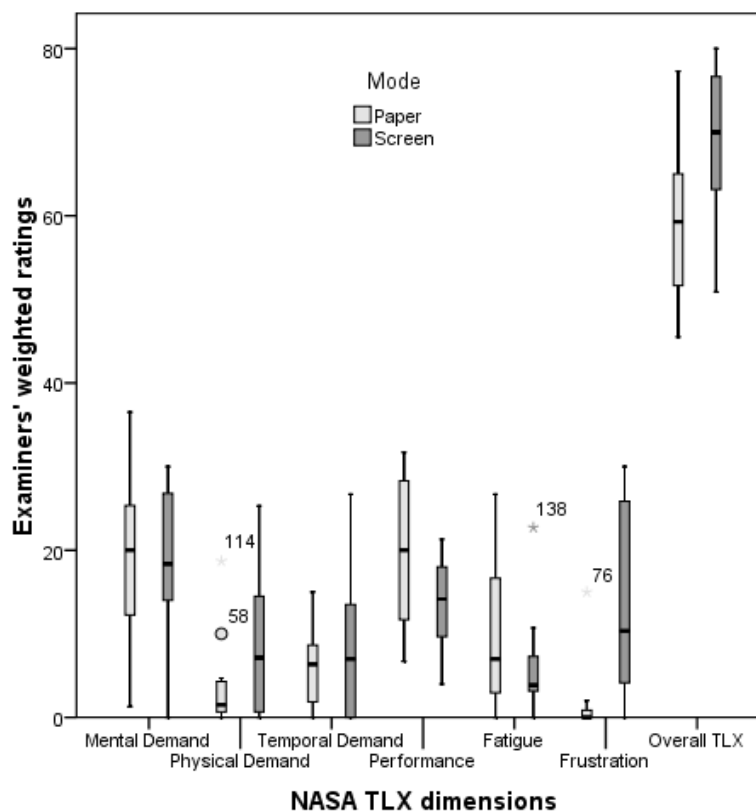
Examiners completed the NASA-TLX survey midway through their paper marking and again midway through their screen marking, enabling comparison of cognitive workload levels across both marking modes. Figure 1 presents the overall cognitive workload data alongside the examiners' ratings for the individual workload dimensions. This analysis found that overall cognitive workload was significantly greater for marking on screen compared with marking on paper.

The six dimensions contributed differently to overall cognitive workload levels and each dimension was influenced by mode to a differing degree. The *frustration* subscale showed a large and statistically significant mode-related difference, with examiners reporting much greater levels during marking on screen. All other dimensions mode had no significant effect on cognitive workload.

To explore in more depth which aspects of marking contributed to the findings from the NASA-TLX survey the examiners were interviewed about which specific features of marking practice contributed to their experiences of each cognitive workload dimension.

Mental demand: examiners reported a number of sources of mental demand that were inherent to any marking processes, e.g. the responsibility for reaching the correct judgement.

Physical demand: this included the constraints of the physical environment which contributed to physical strain. In both marking modes these constraints appeared to originate from examiners' working environments.

Temporal demand: was generally reported to be very low in both marking modes, perhaps reflecting the generous project marking deadlines compared with 'live' marking sessions.

Performance: generally speaking it seemed that examiners perceived two types of performance; the satisfaction of task completion in a new marking environment, and the professional accomplishment of performing high quality work.

Fatigue: a variety of elements contributed to fatigue, e.g. novelty and initial struggles with technology; physical strain and looming deadlines; and mental fatigue. Others felt energised by some particular aspects of marking on screen, e.g. the ability to read poor handwriting; the lack of administrative requirements; and 'seeing the scripts off by a click'.

Frustration: most examiners mentioned the novelty of marking on screen or specific elements of the software environment as causes for initial frustration. However, once technical problems were resolved, examiners generally grew more comfortable with screen marking and frustration levels receded.

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