ELSEVIER

International Journal of Industrial Ergonomics 31 (2003) 411–423

International Journal of Industrial Ergonomics

www.elsevier.com/locate/ergon

VDT versus paper-based text: reply to Mayes, Sims and Koonce

J.M. Noyes[a,*], K.J. Garland[b]

[a] Department of Experimental Psychology, 8 Woodland Road, University of Bristol, Bristol BS8 1TN, UK
[b] Department of Psychology, University of Plymouth, UK

Received 29 April 2002; received in revised form 21 October 2002; accepted 5 February 2003

Abstract

Research into reading performance during the 1980s suggested that reading from video display terminals (VDTs) was slower and less accurate than reading from printed material. A recent study by Mayes et al. (Int. J. Ind. Ergon. 28 (2001) 367) reported in this journal also indicated that reading from a VDT took significantly longer than reading from paper. In reply to this, we report a study that examined directly comparable text in the two media in terms of study and reading times, number of correct answers and a memory retrieval measure. Neither study or reading times, nor the level of correct answers differed between the two forms of presentation, but the manner in which learned information was recalled did show a significant difference. It is suggested our replicated findings indicate that when material is adequately matched across media, reading times and number of correct answers do not differ, but differences in cognitive processing associated with memory assimilation do occur. This has major implications for the use of VDT presentation of knowledge. Further, the findings suggest that more traditional forms of performance measures, in particular reading speed, may be poor indicators of the amount and quality of information obtained from reading from VDTs in comparison to hard copy.

Relevance to industry

In the 1980s, research indicated reading performance was less for computers than paper. It is suggested this is not the case today, although the two media do have different effects on cognitive processing as measured by memory processes. This has important implications for using computers in learning and training activities.

Keywords: HCI performance issues; VDTs; Paper; Reading speeds; Memory processes

1. Introduction

Research into reading performance using different media has tended to indicate that reading from video display terminals (VDTs) is slower and less accurate than reading from printed material. For example, Muter et al. (1982) found that the speed of (silent) reading from computer screens was slower than that from paper. This finding was supported by Creed et al. (1987), Gould et al. (1987a), Gould and Grischkowsky (1984), Heppner et al. (1985), Kruk and Muter (1984), Weldon

*Corresponding author. Tel.: +44-1179-288-560; fax: +44-1179-288-588.

E-mail address: j.noyes@bristol.ac.uk (J.M. Noyes).

0169-8141/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0169-8141(03)00027-1

et al. (1985), Wilkinson and Robinshaw (1987), and Wright and Lickorish (1983). Dillon (1992) in a review of performance comparisons for reading from computers and paper suggested that although the evidence was inconclusive, there appeared to be a decrement for screen reading. In his later authoritative text on this topic, he suggested that this difference was in the region of some 20–30% (Dillon, 1994). Gould et al. (1987a) found a similar difference when comparing two different, printed typefaces in a proof-reading task; a reading rate of some 20% faster was found for the clearer, 'high-quality' paper version. In a follow-up study, Gould et al. (1987b) also found reading from paper to be faster than reading from computer screen. More recently, this finding has also been indicated by Pickering (1997), who found that participants' proof-reading and preference levels for a paper version of a musical score were higher than those for a static, comparable, computer-based presentation. However, some studies reported insignificant differences in readability between the two presentational formats (e.g. Askwall, 1985; Cushman, 1986; Keenan, 1984; Muter and Maurutto, 1991; Oborne and Holton, 1988), while some have indicated only minimal differences (Kak, 1981; Switchenko, 1984).

The only study that we have found that has indicated that it is beneficial to read from a computer screen rather than from paper was by Newsted (1985) in a study using online questionnaires. The computer presented questionnaire was found to be completed more accurately than the pen and paper version. However, those using the computer version chose to do so, while those using hard copy were given no alternative. Therefore, it could have been anticipated that the computer group would show superior performance since they were able to exercise their preference.

In contrast to Newsted's (1985) finding, reading from computer screens has been found to be less accurate. For example, Gould et al. (1987a) found that proof-reading for horizontal and vertical page orientation versions of printed material produced similar levels of speed, and both were significantly 'better' than those obtained for the same task presented on a computer screen (in a horizontal orientation). Both Creed et al. (1987) and Muter et al. (1982) found accuracy, as measured by proof-reading, to be less for computer-based text. However, Gould et al. (1987b) reported no reliable differences in accuracy.

Another approach has been to consider the degree of comprehension exhibited by readers when working from computer screens and equivalent hard copy. Measurement of comprehension often takes the form of asking participants questions about the content of the material read with the level of comprehension being derived from the number of correct responses. Findings from previous studies indicate that there is little difference between levels of comprehension for screen and printed presentations (Cushman, 1986; Muter et al., 1982; Muter and Maurutto, 1991; Oborne and Holton, 1988). However, Belmore (1985) found slower reading speeds (by 12%) and poorer comprehension (by 67%) for computers compared to paper-based presentation, but only when the computer-based text trials were given to participants before the paper-based text. This suggests reading from paper first facilitates subsequent computer-based performance, but not vice versa. However, there was no evidence in this study of gradual improvement across individual trials, which might have been expected due to practice effects.

1.1. Implications from these research studies

The findings from these studies are largely inconclusive, but tend towards VDTs eliciting poorer user performance. The inconsistency in the findings appears primarily due to variations in the methodologies employed: different designs make comparative interpretation difficult, especially where the computer-based learning format is used as an adjunct to paper-based instruction, or where study time is not matched. More importantly, in the studies mentioned, it was often not possible to attain directly comparable media due to the computer and printing technological limitations of the era in which the research was taking place. For example, with reference to text density, book pages typically display about one-third more words than a VDT (Muter et al., 1982), and Kolers et al. (1981) demonstrated that text presented with 80 characters per line was read 17% faster than a 40 character display. Further, many factors relating to screen characteristics that might have influenced performance were examined in isolation, and possible interactions were not investigated. A particular example of this is the comprehensive 1987 study carried out by Gould et al. (1987b). They attempted to isolate a single variable (from user experience, display orientation, character size, font size, and polarity) that might explain why reading from a VDT was slower, and although they manipulated these variables in turn, they could not find any one variable, which would account for the differences. Hence, they concluded that a combination of variables might be affecting the reading speeds.

Research comparing VDTs and paper was largely carried out in the 1980s, and the findings from this work are still cited (e.g. Martin and Platt, 2001). Given the advances in computer technology, which have occurred over the last 10–15 years, the situation may have now changed. As an example, Muter et al. (1982) required their participants to read continuous text from a television screen or a book. These media are unlikely to have been matched in terms of their physical characteristics. We know that the quality of the image whether on the screen or paper is likely to affect the ease and efficiency of the interaction (see, for example, Gould et al., 1987a). The resolution of both formats is important for quality, and has been found to affect speed and accuracy in visual search (Harpster et al., 1989) and proof-reading tasks (Ziefle, 1998).

1.2. The Mayes et al. study

Performance between the two forms of media using current technology can be assessed from the study reported by Mayes et al. (2001) who carried out a study comparing VDTs and paper in terms of comprehension and workload measures. In their study, they attempted to ensure that the information presented on the VDT and the paper was made as similar as possible in terms of resolution, character size, colour, and visual angle. They timed participants reading an article presented on either a VDT or paper, and then assessed their level of comprehension by asking a number of declarative questions. They found that the VDT group took significantly longer to finish reading the article; however, comprehension scores were found not to differ significantly between the two media. These findings appear to support earlier results that reading from a VDT takes longer. It could be argued that this is surprising given that technological advances would allow more directly comparable material for both media, and as already stated, Mayes et al. attempted to control for this. It was not possible to control for the practical difference in 'page-turning', but as Mayes et al. argued, this factor was unlikely to offer an explanation for the differences found. The inherent characteristics of the VDT (contrast, luminance and refresh rates levels) cannot be effectively controlled across the two media. They are factors that offer a possible explanation, and we argue that these interfere with cognitive processes, thus reducing performance from VDTs.

An extension of the Mayes et al. study was to look at workload and the demands on working memory. The rationale for this was that slower reading speeds would be associated with higher cognitive workloads (as measured by the NASA-Task Load Index) due to increased demands on working memory. In order to investigate this, a second study was conducted by them where participants had to hold a list of letters in working memory. In contrast to the first study, no significant differences in reading times were found. However, when academic performance as measured by grade point average (GPA) was partialed out, the participants using paper were found to take significantly longer to read the article than the VDT group. They also reported that GPA was the main predictor of comprehension scores. This finding is not surprising given the link between working memory capacity and 'fluid' intelligence (see, Engle et al., 1999). Individuals who have greater working memory capacities have been shown to demonstrate higher levels of fluid intelligence, and hence, better academic performance. This suggests if VDTs and paper are going to be compared using working memory tests that the academic background of the participants needs to be taken into account. The findings of Mayes et al. suggest that reading from a VDT reduces the capacity of working memory, and we would argue that this is due to the visual inputs specific to cathode-ray-tube (CRT) monitors.

The evidence reported above implies that some basic performance differences still exist between VDT and paper-based formats. Mayes et al. used current VDT technology and yet they were able to replicate the earlier work that indicated reading from screen was slower than reading from paper. It is suggested, however, that reading speeds do not provide enough insight into what participants are doing in tasks where performance is being measured using comprehension scores. Hence, the first aim of the study being reported here is to compare current VDT technology and paper-based text in terms of both reading and study times. This distinction was made due to inconsistency in the instructions given to participants in earlier reading speed studies. In some instances task demands emphasised speed, in others accuracy, while some papers failed to report fully procedural details, making it impossible to determine what was being measured by 'reading speed'. Mayes et al. gave participants 25 min to read the material before answering questions. However, it was not specified whether they were told to read as quickly as possible, or whether they were forewarned that they would be tested on their comprehension of the material, and accordingly should read at a rate more conducive for these purposes. In addition, it is not known whether once reading had finished, if participants were allowed to continue learning the material up until the maximum time limit. People vary in their style of studying material; it may be that some will read slowly and in depth for a single read, while others prefer to read at a faster rate, but then re-read to consolidate the learning process. Variations in total time exposed to the learning material could be expected to affect performance outcome. It may be this was a factor that influenced the findings of Mayes et al.

1.3. CRT characteristics

The characteristics of VDTs seem to be a crucial element when comparing user performance, computer screens and hard copy. It is thought there are certain characteristics of CRTs that may affect our reading abilities (for a comprehensive list, see, Mills and Weldon, 1987). To maintain an image on a CRT monitor, the electron beam repeatedly scans (refreshes) the phosphor surface. Montegut et al. (1997) showed refresh rates of 60 MHz (creating a gap of some 16 ms between each screen image) reduced reading speed by some 3% compared to refresh rates of 500 MHz (a gap of some 2 ms), which could, perhaps, be equated to a decrement of approximately 5% in comparison to a static image. Flicker, the visible movement of the screen image caused by refreshing, can be evident if this refresh rate is below a certain threshold and/or the persistence level of the phosphor is low. Flicker is known to be annoying and this may have a detrimental effect on reading skills (see, Lyskov et al., 1998).

Other possible physical influences include luminance and contrast. Luminance is higher for CRT monitors than paper, and its level fluctuates over time (Blanco and Leirøs, 2000). Contrast is a further attribute that can vary between the two media. When expressed in ratio terms (foreground to background), screens have a contrast of 2:1 to 30:1, while measurements from paper vary between 1:5 and 1:10 (Grandjean, 1980). These factors and screen flicker have been demonstrated to affect brain activity differentially. For example, Lyskov et al. (1998) found significantly higher visually evoked potentials for low luminance, and for a refresh rate of 60 Hz compared to 72 Hz, extending beyond the peripheral areas of the visual system to the central brain regions. While luminance was also found to affect visual evoked potential latencies (Kammer et al., 1999), with evidence suggesting this effect influenced processing beyond visual input. Goodyear and Menon (1998) found increases in activity in the visual cortex as levels of contrast increased, and changes in activity due to interactions of luminance, contrast and refresh rates have also been found (Wollman and Palmer, 1995). Further, the same characteristics have been demonstrated to affect performance on various visual tasks (e.g. Kennedy and Baccino, 1995).

A further point concerns the novelty of using VDTs for activities normally associated with paper-based text. Muter et al. (1982) when trying to explain their findings suggested that a lack of familiarity with the display screen might have hindered performance and/or the novelty of reading from a screen may have slowed down the reader. This may have been the situation 20 years ago, but is unlikely to be the case today as we are exposed to a fast increasing range of display technology from mobile telephones and personal organisers, to cash point machines and ordering systems in retail outlets. If familiarity can be eliminated as an influence on performance, then a stronger case can be made for any differences found being due to the monitor characteristics.

1.4. Measurement of learning performance

Measurement of learning can take a number of forms, including score improvement between pre- and post-tests and final achievement; however, these can be confounded by variations in academic ability or prior experience. Therefore, an additional measure in our study, namely memory awareness ratings, was employed in conjunction with comprehension scores.

Memory awareness measures, as a means of gauging the nature of recall (and hence learning processes), have been widely used by psychologists. They are based on the work of Tulving (1985) who described two main types of retrieval response, namely 'Remember' and 'Know'. Knowledge that is 'remembered' is characterised as being recalled in association with related information about the learning episode or personal details. Whereas 'known' knowledge is recalled without any such conscious recollection and thus, is information which is simply based on a certain sense of just knowing or familiarity, e.g. one's date of birth or age. This has become known as the Remember–Know learning paradigm. Tulving argued that as time passes, memory for most specific events 'fades', or reduces in contextual details and, therefore, there is a transition from Remember to Know responses. This change in memory representation is consistent with the schema theory of memory (e.g. Schank and Abelson, 1977), in that memories are reconstructed to represent the essence or essential elements of the original experience, with the amount of contextual detail remaining being dependent on the importance of the experience. Although memories can shift states, material can be assimilated into the semantic memory system (characterised by a Know response) almost immediately, dependent on the type of knowledge and existing memory structures (for a review see, Gardiner and Richardson-Klavehn, 2000). Findings suggest that knowledge in the Know state is more readily applied and is indicative of 'better' learning (Conway et al., 1997; Herbert and Burt, 2001). Neuropsychological findings (e.g. Düzel et al., 1997; Henson et al., 1999), and evidence of dissociations from experimental research for independent and participant variables (for a review see, Gardiner and Richardson-Klavehn, 2000), additionally suggest that Remember and Know responses may represent distinctions between processes and behaviour, as well as different recollective experiences. Further, patterns of activation for episodic and semantic memory retrieval have been found to differ (for a review see, Cabeza and Nyberg, 2000).

An extension of the Remember–Know paradigm when measuring awareness of knowledge is the addition of a Guess response category. This is one way to reduce the amount of guessing that otherwise contributes to Know responses, as a guess would be without contextual memory detail and therefore, not recorded as a Remember response. In addition, Conway et al. (1997) suggested a fourth category arising from their experimental work. They reported results where participants indicated that besides Remember, Know and Guess responses, there was another form of recall, linked to forms of familiarity that could describe the manner in which memories were retrieved. Accordingly, Conway et al. used four memory awareness categories in their main study: Remember, Know (to represent 'just know'), Familiar (to represent more familiar than others) and Guess. The use of these categories has been validated in a study by the authors (see, Garland and Noyes, 2000; Herbert and Burt, 2001), and this approach is adopted in the current study.

The application of the Remember–Know learning paradigm in the comparison of VDTs and paper comprises a new approach. The findings of Garland and Noyes (Submitted b) showed that the Remember/Know relationships significantly differed between presentational formats. Similar levels of the two types of retrieval were found for paper, while lower Know levels were found in the computer condition. Hence, it was suggested that a significant difference in the frequency of Remember and Know responses would again be observed between those learning from the two media. It was predicted that reading and study speeds would be slower when using a VDT in keeping with the findings of Mayes et al., and previous research relating to performance variations due to inherent monitor characteristics. Since there appears to be a difference between the two media, this suggests that there was a difference in the cognitive processing taking place when studying information presented on a VDT and paper.

A second aim of the Mayes et al. study was to look more closely at the cognitive processes that take place when reading material from VDTs and paper. They found that working memory demands disrupted the VDT task more, although they did have the confounding variable resulting from the strong association between academic performance and memory abilities. In order to avoid this, the current study introduces a memory awareness measure rather than a memory measure in order to find out if different learning processes are occurring between the two media.

2. Method

2.1. Design and participants

A between-subjects design was employed for the form of presentation of the learning material, namely, VDT or paper-based (these conditions will be referred to as Computer and Paper, respectively). A within-subjects design was employed for the content of the material and the responses required. The dependent variables were the number of correct answers, and the frequencies for the memory awareness responses. In addition, the total time taken to read through the study material was recorded, and a separate test to gauge reading speed was also completed.

A total of 50 students from the University of Bristol carried out the experiment. Three of these were paid volunteers (postgraduates) whilst the others were unpaid undergraduates who participated as part of a course requirement. (In addition, two, paid postgraduates piloted the study.) The 9 men and 41 women had an age range of 18–24 years ($M = 20.00, SD = 2.06$). None were known or observed to have any visual or other impairment that may have affected their ability to complete the task. All participants had similar levels of academic ability as indicated by comparable course entry requirements. Further, all indicated they had received no formal teaching in Economics and could be considered naïve in terms of the learning material presented. Allocation of equal numbers of participants to each of the two conditions was randomised.

2.2. Materials

The Computer group received their learning material via a 15 in, SVGA monitor (65 Hz refresh rate) powered by a Memax 'x86 Family Model' personal computer. The study material comprised an introductory Economics course adapted from a package developed by a consortium of UK universities. It was presented as a document that comprised a single image on each page. The Paper group was provided with a spiral-bound booklet containing the same learning material, matched as closely as possible for size, colour and resolution. The material comprised 22 screens/pages incorporating images/diagrams with between one and three paragraphs per screen/page.

A page of text on the topic of Ying and Yang (295 words in four paragraphs) was used to find reading times. This was matched in terms of colour, polarity, typeface, font size and layout across the computer and paper presentations.

2.3. Procedure

Participants were seated, either in front of a computer or at a desk and asked to read the instruction sheet. Those in the computer condition were given the opportunity to familiarise themselves with the presentation and its navigation. The instructions were then repeated verbally to ensure that participants fully understood the task requirements. Each participant began the experiment by studying the material provided indicating to the experimenter when they had done this. A total of 20 min was allowed for this; pilot work had indicated that this amount of time was more than ample for participants to study the material. They were instructed to use all the time available, re-reading any sections they wished once they had completed the initial reading, but were asked to indicate when they had finished reading all of the pages for the first time. Participants were told they would receive 20 questions, with multiple-choice responses, on any aspect of the material after the study period and that they would also be required to indicate the manner in which they retrieved their answer. The latter, the memory awareness ratings, would comprise four options. Additional explanations about these, over and above that given in the written instructions, were given and the participants had the opportunity to ask questions about the instructions. Participants were instructed to select the memory awareness rating that best reflected their initial recall of the answer, as they might be expected to receive more than one type of memory in order to formulate their answer. They were told that there was no time limit for answering, and where they were unsure of an answer they were asked to make a guess. The instruction sheet, with definitions and examples of the four memory awareness ratings was retained by participants during the test period to act as a guide, if needed.

On completion of the questions, participants were tested on their reading speed. They were presented with a page of text, on computer screen or paper, determined by their condition allocation in the main study, and were asked to read the text as quickly as possible while still being able to comprehend the contents. Timing commenced when any key was pressed to activate the presentation on the screen, or when the experimenter turned over the printed page. As soon as they had finished they pressed any key on the computer, or the stop button on a stopwatch provided. Participants were thanked and debriefed.

3. Results

3.1. Study and reading times

The mean time to complete a single reading studying the material was 872 s (SD = 143.19). The Computer group was slightly faster ($M = 868.84$, SD = 148.22) than the Paper group ($M = 875.28$, SD = 140.95) for study times: however, this difference was not significant. The mean time taken to read the single page of text was 46.18 s (SD = 10.70), with the Computer group ($M = 45.95$, SD = 10.77) slightly (but not significantly) faster than the Paper group ($M = 46.41$, SD = 10.84). A Pearson's correlation was carried out between the response times for the two data sets. A very low correlation was found ($r = 0.014$) and this was not significant.

Pearson's correlations between study times and number of correct scores, and reading times and correct scores were $r = 0.171$ and $r = -0.004$, respectively. Separate Pearson's correlations were completed on the data for the two groups. The correct scores for the Computer group did not significantly correlate with either study or reading times ($r = 0.141$ and $r = -0.001$, respectively), nor did the two measures ($r = -0.198$). Likewise, no significant correlations were found for the Paper group data. Correct score correlations for study and reading were $r = 0.297$ and $r = -0.106$ ($p = 0.149$, two-tailed), respectively, and between the two measures, $r = 0.235$.

3.2. Correct responses

The data showed that correct scores for the Paper group ($M = 10.68$, SD = 3.13) were slightly higher than those for the Computer group ($M = 10.04$, SD = 2.11). However, a one-way ANOVA confirmed that this difference was not significant.

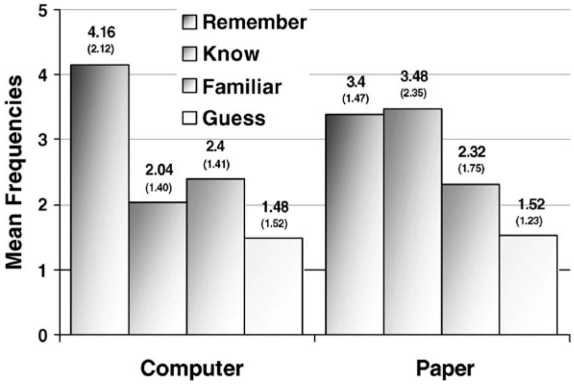

Fig. 1. Mean frequencies (SDs) for memory awareness categories.

3.3. Memory awareness ratings

The frequency of Remember responses was approximately twice that for Know responses in the Computer condition. By comparison, levels of Remember and Know responses were similar in the Paper condition. Familiar and Guess ratings showed similar frequencies in the two conditions. The mean frequencies (SDs in parenthesis) are given in Fig. 1.

A one-way ANOVA found a significant effect of awareness, $F(3, 150) = 31.312$, $p < 0.001$ ($\eta^2 = 0.385$). Chi-square tests of independence (χ^2), using the cross-tabulation procedure, were completed on the frequencies for all memory awareness ratings, between conditions (two-tailed significance levels are reported). A significant effect for condition × awareness was found, $\chi^2(3) = 11.763$, Cramer's $V = 0.152$, $p = 0.008$. In order to determine whether the Remember/Know relationships differed across conditions, a planned comparison was conducted. Significant differences were found for the Remember and Know pairing between the two groups ($p = 0.001$) showing that the relationship for Remember/Know for the Computer condition (Remember means higher than Know) differed from the Remember/Know relationship for the Paper condition (where Know levels were similar to Remember frequencies).

Pearson's correlations were carried out, individually, for the number of correct scores associated with Remember and Know responses for both reading and study times. None of these correlations were significant. Study times × Remember and Know correct answers were, $r = 0.156$ and $r = -0.058$, respectively. Reading times × Remember and Know correct answers correlations were, $r = -0.247$ and $r = 0.013$, respectively.

4. Discussion

4.1. Reading speeds

Our results indicated that mean reading times for a single page of text did not differ between computer and paper-based presentational formats. While this is inconsistent with some earlier findings (see, Dillon, 1994; Gould and Grischkowsky, 1984) and Mayes et al. (2001), it is consistent with other results (e.g. Askwall, 1985; Cushman, 1986; Keenan, 1984; Muter and Maurutto, 1991; Oborne and Holton, 1988) that found no differences between computer and screen presentation. The findings of our study suggest that if the material used is matched, for typeface, font size, polarity and general clarity (resolution), then reading speed (as measured by time) does not differ significantly between these two media.

The inconsistencies of earlier findings may be partly due to differences in the comparability of the two forms of the media presented in the studies, and variations in the measurement of reading speed. However, the former point does not explain however why Mayes et al. (p. 367) found that "those who read from a VDT took significantly longer than those reading from paper". Mayes et al. matched the resolution of the two images and made the VDT and paper versions of their reading material as similar as possible. This suggests that their measure of reading speed might have been applied in a different way than in our study. Mayes et al. stated in their paper that participants stopped reading either after the time limit of 25 min had been reached or upon completion of the article. The precise instructions given to participants may have been an important factor here. If, for example, participants were told to read through the article in order to prepare for a comprehension test, they may have found this easier to do (i.e. involving less cognitive effort) with a paper-based version than a computer screen. Although there were no significant differences between the mean NASA-Task Load Index workloads for the two conditions in the Mayes et al. study, there was a trend towards more cognitive effort being expended when using the VDT.

4.2. Studying versus reading

In our experiment, we differentiated between studying and reading. Time taken to study the learning material, as measured by a single, complete reading, was also found to be comparable between the two conditions. Nevertheless, with a confidence interval of 95%, it is possible that a small loss in reading speed from CRT monitors compared to paper (or an alternative, non-flickering format) would not be identifiable from the design employed. For example, the loss of some 3–5% as suggested by Montegut et al. (1997) due to the screen refreshing.

The measurement of study time is problematic, since in an experimental setting, it is necessary to provide participants with some guide as to how long they might expect to spend undertaking a particular task. The instructions need also to ensure that participants are not placed under undue pressure to complete the task either very quickly or very slowly. Our instructions did this, but Mayes et al. do not provide any information about their instructions. The length (and complexity) of the material is important here. Mayes et al. gave the length of their reading material as 19 paragraphs; it is not possible to compare this with the amount of text in our study, as we do not know the length of their paragraphs.

We found there was no significant correlation between reading and study times; the two performances were totally unrelated. This suggests that the way in which a short passage of text may be read, perhaps when done so in haste, bears no relationship to the speed with which the same reader will study material for learning purposes. Thus, variations in experimental design and materials, instructions to participants, and the way in which reading speed is determined may be crucial and offer an explanation for the differences found by some and not by others. Further, other factors may influence study reading speed, e.g. internal performance shaping factors such as learning style, motivation, and prior knowledge, but not significantly the mode of presentation of the material, provided it is adequately matched in content and appearance.

4.3. Learning measures

In terms of comprehension scores, no difference in the number of correct scores was found between the VDT and paper-based materials. This suggests that learning from CRT monitor-based material can be quantitatively similar to that from comparable paper-based material. Although our participants were not matched across conditions for general academic ability or prior knowledge of the topic, it was hypothesised they had similar levels of academic ability as measured by course entry requirements. Further, they professed no background in Economics. Notwithstanding the random allocation of participants to the Computer and Paper groups, variations in these factors might have differentially influenced the levels of correct scores obtained (as indicated by Mayes et al.).

A significant effect for awareness frequencies was found in our study. In the Computer group, Remember frequencies were almost twice that of Know frequencies, while Remember and Know response levels were similar in the Paper group. The differences in the Remember/Know relationship in the two groups supported our hypothesis and replicated earlier findings. Additional research by the authors (Garland and Noyes, unpublished) found access to memory characterised by Know responses was reliably faster than access to Remember responses. These and the current findings provide support for the argument that the two retrieval states reflect qualitatively different memory sources and that the former characterises knowledge that is more readily applied, as was suggested by Conway et al. (1997) and Herbert and Burt (2001).

An explanation that could be offered to explain our findings is that they reflect a change in decision criteria levels. In signal-detection models of retrieval (e.g. Donaldson, 1996) Remember and Know are, respectively, characterised as stronger and weaker memory traces accessed with different response criteria, which are, in effect, on a continuum. Donaldson's model has been found to model effectively most Remember–Know response findings (Gardiner and Richardson-Klavehn, 2000). However, there appears to be only limited theoretical explanation to account for, first, the differing effects of variables on the criteria and secondly, how both Remember and Know can be obtained from the same source merely by adjusting the criteria. The signal-detection model predicts that derived strength from Remember data will be the same as that from using combined Remember and Know data. However, Gardiner and Gregg (1997) demonstrated that this was not the case. Further, when a Guess response option was included, the resulting memory strength was also not equivalent to that derived from Remember responses alone. In addition, the model is inconsistent with findings that suggest Know responses reflect a qualitatively different memory source from that indicated by Remember responses (e.g. Gardiner and Gregg, 1997; Gardiner and Java, 1990), nor can it account for the shift from Remember to Know over time (Conway et al., 1997; Herbert and Burt, 2001), for in this case Remember would, in effect, become a weaker memory trace than Know.

It appears that signal-detection models cannot offer an explanation for the differences found in memory retrieval states. We consider that a possible alternative explanation is that characteristics of the computer screen (refresh rate, high levels of contrast and fluctuating luminance) interfere with cognitive processing for long-term memory, as well as working memory (Mayes et al., 2001). These characteristics have been found differentially to affect brain activity, while Remember and Know responses show functional and anatomical variations in neuronal activity (e.g. Henson et al., 1999).

These findings have important implications for workers carrying out computer-based tasks. Since knowledge is more readily retrieved when presented in paper format, there may be occasions, e.g., in safety-critical systems, when vital information that needs to be assimilated quickly should not be presented via a computer screen.

4.4. Cognitive loads

In their first study, Mayes et al. found that reading from VDTs was significantly slower. This replicates previous findings and could be due to the additional effort needed to process information from a display screen. The various characteristics of VDTs, e.g. the refresh rate, luminance, contrast, etc., interfere with our processing of material leading to slower performance as measured by reading speed and comprehension scores. This was particularly prevalent with the older VDT technology, but the effect can be reduced today, and even eliminated as we demonstrated, if the materials used in the two conditions are closely matched.

Mayes et al. found no significant differences for workload or comprehension scores across the two conditions. This may seem surprising since it could be hypothesised that increased cognitive processing might be associated with increased workload. Indeed, a significant negative relationship was found between workload and comprehension scores. This somewhat counter-intuitive finding is one that has been found in a visual search task where making errors was associated with increased workload as measured by the NASA-Task Load Index (Leggatt and Noyes, 1996). Hence, it could be suggested that lower comprehension scores equate to more mistakes, and subsequently, a greater effort in terms of cognitive workload.

In their second study, Mayes et al. introduced a distraction task to limit working memory capacity. They now found that there was no difference between performance on VDTs and paper. It is suggested that this task was very successful at limiting working memory, and did this to such an extent, that no performance differences were found between VDTs and paper. Thus, the interference caused by the working memory task was far greater than the interference effects caused by the VDT characteristics. Hence, the more subtle effects emanating from the VDT demonstrated in the first study were lost in the performance data generated in the second study. An alternative explanation is that CRTs interfere with working memory, as when the distraction task was introduced, little further decrement was experienced. This was in contrast with the paper condition, where there was no initial interference, and therefore, more room for a decrement to occur from the intervening task.

The finding from Mayes et al. of increased working memory load fits well with our results that suggest differences in long-term memory organisation, i.e. additional visual input reduces working memory capacity, and this may reduce the amount of rehearsal processing possible, thus affecting the consolidation processes in long-term memory.

5. Conclusions

In our study, no significant differences between matched computer and paper-based text in terms of the time taken to study and to read material, and the level of learning achieved, were found. In contrast, significant differences were found for the memory awareness patterns of responses between the two conditions. The memory awareness measures provide an insight into the memory processes and by studying retrieval, it is possible to infer what is happening at encoding. Hence, it is suggested that there are differences in cognitive processing taking place when learning from VDTs and paper. These emanate from the physical characteristics of each, and will not always be found when using macro-measures such as reading speed. The work of Mayes et al. supports this.

At this stage, we do not have a full understanding of how the various attributes of VDTs influence and interact together to affect our cognitive processing. Further research is needed to explain how these characteristics relate to the type of material being studied. As Mayes et al. suggested, physiological measures might be most appropriate in helping us understand what exactly is happening when participants attempt to learn from VDTs and paper. To reiterate an earlier point, it is important that research moves away from comparing VDTs and paper using only superficial measures such as reading times and comprehension scores. In order to begin to assess the real differences between the two media, there is a need to find out about the individual's cognitive processing and see how this relates to performance.

References

Askwall, S., 1985. Computer supported reading vs reading text on paper: a comparison of two reading situations. International Journal of Man–Machine Studies 22 (4), 425–439.

Belmore, S.M., 1985. Reading computer-presented text. Bulletin of the Psychonomic Society 23 (1), 12–14.

Blanco, M.J., Leirøs, L.I., 2000. Temporal variation in the luminance level of stimuli displayed on a cathode-ray tube monitor: effects on performance on a visual vigilance task. Ergonomics 43 (2), 239–251.

Cabeza, R., Nyberg, L., 2000. Imaging Cognition II: an empirical review of 275 PET and fMRI studies. Journal of Cognitive Neuroscience 12 (1), 1–47.

Conway, M.A., Gardiner, J.M., Perfect, T.J., Anderson, S.J., Cohen, G., 1997. Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates. Journal of Experimental Psychology: General 126 (4), 393–413.

Creed, A., Dennis, I., Newstead, S., 1987. Proof-reading on VDUs. Behaviour and Information Technology 6 (1), 3–13.

Cushman, W.H., 1986. Reading from microfiche, VDT and the printed page: subjective fatigue and performance. Human Factors 28 (1), 63–73.

Dillon, A., 1992. Reading from paper versus screens: a critical review of the empirical literature. Ergonomics 35 (10), 1297–1326.

Dillon, A., 1994. Designing Usable Electronic Text: Ergonomic Aspects of Human Information Usage. Taylor & Francis, London.

Donaldson, W., 1996. The role of decision processes in remembering and knowing. Memory and Cognition 24 (4), 523–533.

Düzel, E., Yonelinas, A.P., Mangun, G.R., Heinze, H.-J., Tulving, E., 1997. Event-related brain potential correlates of two states of conscious awareness in memory. Proceedings of the National Academy of Sciences of the United States of America 94 (11), 5973–5978.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., Conway, A.R.A., 1999. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. Journal of Experimental Psychology: General 128 (3), 309–331.

Gardiner, J.M., Gregg, V.H., 1997. Recognition memory with little or no remembering: implications for a detection model. Psychonomic Bulletin and Review 4 (4), 474–479.

Gardiner, J.M., Java, R.I., 1990. Recollective experience in word and nonword recognition. Memory and Cognition 18 (1), 23–30.

Gardiner, J.M., Richardson-Klavehn, A., 2000. Remembering and knowing. In: Tulving, E., Craik, F.I.M. (Eds.), The Oxford Handbook of Memory. Oxford University Press, Oxford.

Garland, K.J., Noyes, J.M., 2000. 'Remember–know' dissociation for learning between computer screen presentation and printed material. Paper presented at the British Psychological Society (BPsS) London Conference. BPsS, Leicester.

Goodyear, B.G., Menon, R.S., 1998. Effect of luminance contrast on BOLD fMRI response in human primary visual areas. Journal of Neurophysiology 79 (4), 2204–2207.

Gould, J.D., Grischkowsky, N., 1984. Doing the same work with hard copy and cathode-ray tube (CRT) computer terminals. Human Factors 26 (3), 323–337.

Gould, J.D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N., Minuto, A., 1987a. Reading is slower from CRT displays than from paper: attempts to isolate a single-variable explanation. Human Factors 29 (3), 269–299.

Gould, J.D., Alfaro, L., Finn, R., Haupt, B., Minuto, A., 1987b. Reading from CRT displays can be as fast as reading from paper. Human Factors 29 (5), 497–517.

Grandjean, E., 1980. Ergonomics of VDUs: review of present knowledge. In: Grandjean, E., Vigliana, E. (Eds.), Ergonomic Aspects of Visual Display Terminals. Taylor & Francis, London, pp. 1–12.

Harpster, J.L., Freivalds, A., Shulman, G.L., Leibowitz, H.W., 1989. Visual performance on CRT screens and hard-copy displays. Human Factors 31 (3), 247–257.

Henson, R.N.A., Rugg, M.D., Shallice, T., Josephs, O., Dolan, R.J., 1999. Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. Journal of Neuroscience 19 (10), 3962–3972.

Heppner, F., Anderson, J., Farstrup, A., Weiderman, N., 1985. Reading performance on a standardised test is better from print than from computer display. Journal of Reading 28, 321–325.

Herbert, D.M.B., Burt, J.S., 2001. Memory awareness and schematization: learning in the university context. Applied Cognitive Psychology 15, 617–637.

Kak, A.V., 1981. Relationship between readability of printed and CRT-displayed text. In: Sugarman, R.C. (Ed.), Proceedings of the 25th Annual Meeting of the Human Factors Society. Human Factors Society, Santa Monica, CA, pp. 137–140.

Kammer, T., Lehr, L., Kirschfeld, K., 1999. Cortical visual processing is temporally dispersed by luminance in human subjects. Neuroscience Letters 263 (2–3), 133–136.

Keenan, S.A., 1984. Effects of chunking and line length on reading efficiency. Visible Language 18, 61–80.

Kennedy, A., Baccino, T., 1995. The effects of screen refresh rate on editing operations using a computer mouse pointing device. Quarterly Journal of Experimental Psychology Section A—Human Experimental Psychology 48 (1), 55–71.

Kolers, P.A., Duchnicky, R.L., Ferguson, D.C., 1981. Eye movement measurement of CRT displays. Human Factors 23, 517–527.

Kruk, R.S., Muter, P., 1984. Reading of continuous text on video screens. Human Factors 26, 339–345.

Leggatt, A.P., Noyes, J.M., 1996. Workload under complex task conditions. In: Robertson, S.A. (Ed.), Contemporary Ergonomics. Taylor & Francis, London, pp. 281–286.

Lyskov, E., Ponomarev, V., Sandstrom, M., Mild, K.H., Medvedev, S., 1998. Steady-state visual evoked potentials to computer monitor flicker. International Journal of Psychophysiology 28 (3), 285–290.

Martin, L.A., Platt, M.W., 2001. Printing and screen reading in the medical school curriculum: Guttenberg vs. the cathode ray tube. Behaviour and Information Technology 20 (3), 143–148.

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.

Mills, C.B., Weldon, L.J., 1987. Reading text from computer screens. ACM Computing Surveys 19 (4), 329–358.

Montegut, M.J., Bridgeman, B., Sykes, J., 1997. High refresh rate and oculomotor adaptation facilitate reading from video displays. Spatial Vision 10 (4), 305–322.

Muter, P., Maurutto, P., 1991. Reading and skimming from computer screens and books: the paperless office revisited? Behaviour and Information Technology 10 (4), 257–266.

Muter, P., Latrémoullie, S.A., Treurniet, W.C., Beam, P., 1982. Extended reading of continuous text on television screens. Human Factors 24 (5), 501–508.

Newsted, P.R., 1985. Paper versus online presentations of subjective questionnaires. International Journal of Man–Machine Studies 23, 231–247.

Oborne, D.J., Holton, D., 1988. Reading from screen versus paper: there is no difference. International Journal of Man–Machine Studies 28 (1), 1–9.

Pickering, R., 1997. Reading music from screens vs paper. Behaviour and Information Technology 16 (2), 72–78.

Schank, R.C., Abelson, R.P., 1977. Scripts, Plans, Goals and Understanding. Lawrence Erlbaum Associates Inc, Hillsdale, NJ.

Switchenko, D.M., 1984. Reading from CRT versus paper: the CRT-disadvantage hypothesis re-examined. In: Alluisi, M.J., de Groot, S., Alluisi, E.A. (Eds.), Proceedings of the 28th Annual Meeting of the Human Factors Society. Human Factors Society, Santa Monica, CA, pp. 429–430.

Tulving, E., 1985. Memory and consciousness. Canadian Psychology 26 (1), 1–12.

Weldon, L., Mills, C., Koved, L., Shneiderman, B., 1985. The structure of information in online and paper technical manuals. In: Proceedings of 29th Annual Meeting of Human Factors Society. Human Factors Society, Santa Monica, CA, pp. 1110–1113.

Wilkinson, R.T., Robinshaw, H.M., 1987. Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. Behaviour and Information Technology 6 (2), 125–133.

Wollman, D.E., Palmer, L.A., 1995. Phase-locking of neuronal responses to the vertical refresh of computer display monitors in cat lateral geniculate-nucleus and striate cortex. Journal of Neuroscience Methods 60 (1–2), 107–113.

Wright, P., Lickorish, A., 1983. Proof-reading texts on screen and paper. Behaviour and Information Technology 2, 227–235.

Ziefle, M., 1998. Effects of display resolution on visual performance. Human Factors 40 (4), 554–568.

Jan Noyes is a Reader in the Department of Experimental Psychology at the University of Bristol. She graduated from Loughborough University of Technology in 1981 with a Ph.D. in Computer Ergonomics following a B.Sc. in Human Sciences.

Kate Garland is a Lecturer in the Department of Psychology at the University of Plymouth having completed her Ph.D. in Cognitive Psychology at the University of Bristol. She has a Master's degree in Research Methods in Psychology, and an Honours degree in Psychology.