Effect of Presentation on Reading Behaviour

Leana Copeland
Research School of Computer Science
Australian National University
Canberra, Australia
leana.copeland@anu.edu.au

Tom Gedeon
Research School of Computer Science
Australian National University
Canberra, Australia
tom.gedeon@anu.edu.au

ABSTRACT
Eye tracking is a useful tool for investigating how people read and the attention that they give to certain words and phrases. Eye tracking is used to investigate how different presentation formats of the same learning material affect learning performance, eye movements, and reading behaviour. We show that different presentation formats induce different eye movements and that reading behaviour is subject to the goals placed on the reader. We also observe that the presentation format affects not only their learning performance but also how they perceive their performance. Finally, we show that different formats and question types can induce specific reading behaviour such as thorough reading. This can be used to influence how students interact with the learning environment as well as how they learn the material. The purpose of this investigation is to be able to make informative decisions about designing adaptive eLearning environments.

Author Keywords
Eye gaze; eye tracking; adaptive eLearning; presentation format

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Educational material is being offered through online mediums more frequently in part due to increased accessibility and availability of computer technologies. This is especially true for tertiary education, where face-to-face education is now heavily supplemented with material that is available through online learning environments, such as Moodle and Blackboard. It has become common for universities to offer online/off-campus degrees where students may have little or no face-to-face interaction with their instructors or other students. Eye tracking has been shown to be an effective way of analysing human behaviour, particularly reading (Rayner, 1998). Given that a primary form of educational material is text, this raises the question of how eye tracking can be used to make the learning process more effective when an instructor has little or no direct interaction with students.

OzCHI '14 , December 02 - 05 2014, Sydney, NSW, Australia
Copyright 2014 ACM 978-1-4503-0653-9/14/12..$15.00
http://dx.doi.org/10.1145/2686612.2686648

The question of how electronic text (eText) can be used to support learning and comprehension is long standing (Anderson-Inman and Horney, 2007). There is much potential in providing support to readers and much research on how to do so in particular situations, but no consensus on how to do so in a generalised way and for large variety of readers and situations. In particular, research has often tended to focus on readers with difficulties or disabilities (Anderson-Inman and Horney, 2007). Our investigation contributes to the wider research on how to produce specific eText supports in an eLearning environment context, with the use of eye tracking to predict when support is required. The support investigated is the use of quiz questions and their relation to the presented text. This is an evaluation support, whereby materials and prompts are designed to assess student learning from the text.

To do so, the presentation of educational reading material is investigated to assess how eye movements, answer-seeking behaviour, and learning performance are affected. A user study was conducted to compare four presentation formats. These presentation formats are manipulations of the order of which text and quiz questions are shown to a student. The hypotheses for this investigation are that the presentation format of the text and comprehension questions will: 1) have an effect on students' performance, in terms of time and quiz score, and perceived understanding of the text; 2) cause differences in eye movements; and 3) induce different reading behaviour.

This paper is organized into the following sections: background information; user study methodology; results and discussion; recommendations based on the results; finishing with conclusions and further work.

BACKGROUND

Electronic Text (eText)

Electronic text, eText, is the general term for digital presentation and storage of text. eText is read through some kind of computer device, such as a computer screen, tablet, phone, eReader, etc. The advent of smartphones, tablet devices, and eReaders has meant that eText is becoming more prevalent.

Initially a large amount of research went into comparing reading digital to paper based texts (see review by Dillon (1992)). Many studies are centred on the outcomes of reading, such as speed and comprehension. In general, the

literature has shown that there is little evidence to support claims that one method of display is better than another in terms of improving comprehension (Dillon and Gabbard, 1998).

Paper offers advantages over digital presentation, which has been studied to provide design suggestions for better reading technologies (O'Hara and Sellen, 1997). These include supporting annotation, quick and easy navigation as well as control of spatial layout. Meanwhile, eText does itself have advantages over paper that include increased accessibility, easy storage and retrieval, ubiquity, and flexibility. It is the flexibility that eText provides that gives rise to the potential to support learning. Flexibility refers to the ability to dynamically change the way text is read. Changes can be simple, such as font size, colour, or typeface or they can be more complex, such as verbalisations of the text, embedded definitions and links to background information (Anderson-Inman and Horney, 2007). The reader controls such simple changes; conversely, the eText can be transformed to promote learning and comprehension. Horney and Anderson-Inman (1999) produced a typology of resources for supported eText. These include presentational, navigational, translational, explanatory, illustrative, summarising, enrichment, instructional, notational, collaborative, and evaluation resources. The typology is a list of ways in which eText can be supported; they vary vastly in method and purpose. Perhaps for this reason there is no consensus which supports should be provided (Anderson-Inman and Horney, 2007).

Many studies have considered navigation through eTexts as it is considered a non-trivial text to accomplish in electronic form (Dillon, 2004). Studies have investigated navigation in ebooks (McKay, 2011) and periodicals (Marshall and Bly, 2005) as well as the impact of screen size on document triage (Loizides and Buchanan, 2010).

Dillon (1996) proposed a framework for building electronic texts so that the issues relevant to their design are considered. The framework, known as TIMS, was designed to be a representation of the cognitive behaviours humans exhibit whilst reading documents.

We focus on small pieces of text material that are presented to the reader in a structured fashion. Within the limits of the eLearning environment, the reader is free to choose where in the quiz they would like to read but the structure of the tutorial is such that there is a linear progression through the content. In this way the eText can be considered supported eText whereby the addition of an evaluation resource and the format in which it is presented are intended to influence learning.

Eye Movements and Reading Comprehension

Tracking a reader's eye has looked has long been used to measure the reading process (see review by Rayner (1998)). During the process of reading the eye moves in well-studied ways that can be broadly characterised as fixations and saccades. A fixation is where the eye remains relatively still to take in visual information. A saccade is a rapid movement that transports the eye to another fixation. This behaviour is due to the anatomy of the eye. At the centre of the retina is a special part of the eye called the fovea that sees in fine detail. The foveal region of the eye is very small, being only about 0.2mm in diameter. Around the point of fixation, visual acuity extends about 2° (Rayner, 1998). Humans see very little in detail at any fixation so the eye must move around rapidly so that it can compose a more detailed view of the environment.

When reading English fixation duration ranges anywhere between 60-500 milliseconds and is generally about 250 milliseconds (Liversedge and Findlay, 2000). Saccadic movement is between 1 and 15 characters with an average of 7-9 characters. The majority of saccades are to transport the eye forward in the text when reading English; however, a proficient reader exhibits backward saccades to previously read words or lines about 10-15% of the time. Backward saccades are termed regressions. Short regressions can occur within words or a few words back and may be due to problems in processing the currently fixated word, overshoots in saccades, or oculomotor errors. However, longer regressions occur due to comprehension difficulties, as the reader tends to send their eyes back to the part of the text that caused the difficulty (Frazier and Rayner, 1982).

Eye movements can be used to understand the ongoing cognitive processes that occur during reading (Liversedge and Findlay, 2000). Comprehension of text can have significant effects on eye movements (Rayner et al., 2006). A number of studies have shown there are many variables based upon comprehension functions that can influence eye movements during reading. These variables include: semantic relationships between words, anaphora and co-reference, lexical ambiguity, phonological ambiguity, discourse factors, stylistic conventions, and syntactic disambiguation (see review by Rayner (1998)). These variables have different effects on eye movement, causing them to deviate from the default reading process. For example, syntactically ambiguous sentences induce regressions to resolve the comprehension problems (Frazier and Rayner, 1982). Eye movements have been shown to reflect global text difficulty as well as inconsistencies within text (Rayner et al., 2006). More difficult text causes more fixations, more regressions, and longer fixation duration time. Eye movement has also been shown to indicate reading comprehension and reading skill (Underwood et al., 1990).

Whilst eye movements are a good way of measuring the observable part of the reading process it is important to note that the limitation in the context of HCI research is that the researcher cannot tell what the reader is thinking or doing at the time of reading (Dillon, 2004).

Eye Tracking in Adaptive eLearning

As mentioned previously, eText can be modified to support learning. One way of supporting learning is the use of eye tracking. Eye gaze patterns can be used to detect what kind of task the participant is performing (Iqbal and Bailey, 2004) or whether a person is reading or not (Campbell and Maglio, 2001) as well as if they are reading or skimming (Buscher et al., 2008).

The use of eye tracking in adaptive learning systems (ALS) is not novel and has been approached in a number of ways. An example is iDict, a reading aid designed to help readers of a foreign language (Hyrskykari et al., 2000). iDict uses eye gaze to predict when a reader is having comprehension difficulties. If the user hesitates whilst reading a word then a translation of the word is provided with a dictionary meaning. Another example is The Reading Assistant (Sibert et al., 2000) that uses eye gaze to predict failure to recognise a word. The Reading Assistant then provides auditory pronunciation of the word to aid reading. These applications work on the assumption that the user pauses on a problematic word, and then the system provides feedback about that word. They do not look at overall text comprehension or provide feedback about the overall comprehension of that text.

Separate from direct applications of eye tracking in ALSs are the investigations of eye movements within learning systems such as analysing how multiple choice questions are answered (Nugrahaningsih et al., 2013; Tsai et al., 2012), or using eye movements to predict student comprehension of physics concepts when presented as text or images (Chen et al., 2014).

Prior Work

This study presents further analysis and a follow-up to a previous user study (Copeland and Gedeon, 2013). In the initial user study participants' eye gaze was recorded as they completed a tutorial and quiz. The tutorial was composed of 9 screens of textual content each covering a specific area about the main topic of the tutorial (Web Search). After each text screen, participants were required to answer two questions to measure their comprehension. Whilst answering the questions, participants were given the opportunity for a second read-through of the content to assist answering the questions.

Importantly, the study showed that there were distinctively different reading behaviours observed when the text was shown alone compared to when it was shown with the questions. When the text was shown alone, 'normal' reading behaviour is observed. When text was shown with the questions then a behaviour we term *answer-seeking* is observed; this is a targeted search, read, and confirm behaviour. Given the discrepancies in observed behaviours we postulate that by manipulating the way in which the text and the questions are presented will alter not only the reading behaviour observed but also learning outcomes. The sequences the text and questions are presented in are therefore strategically manipulated in this our study to assess this hypothesis.

METHOD

Design

A user study was conducted to collect participants' eye gaze as they read a tutorial and completed a quiz based on the tutorial's content. The tutorial and quiz are coursework from a first year Computer Science course taken at the Australian National University. The tutorial and quiz is composed of 9 screens of textual content, each covering a specific area about the main topic of the tutorial (Web Search). Each screen is 400 words long and has an average Flesch Kincaid Grade readability level of 11.5. This indicates that participants need around a 12th grade education level. As the slides are targeted at a first year university students this is a suitable readability level. For each screen there were two comprehension questions; one of the questions was multiple-choice and the other was cloze (fill-in-the-blanks). These two types of questions were used because they can be used to assess different forms of comprehension (Fletcher, 2006).

The tutorial and quiz was presented to participants in four formats to measure the effect of presentation on participants' eye gaze and answering behaviour. The presentation formats are based on when the quiz questions are presented to the participants in relation to the text. These presentation formats are described below:

Format A: The tutorial text slide *(T)* Figure 1 is first shown to participants followed by a slide with both questions and the tutorial text *(Q&T)* Figure 2. Since there are 9 topics there are 18 slides in total displayed in this part of the study.

In this format participants are required to read the text before being able to read the questions relating to it. Whether the participants understand it or not they have knowledge about the concepts in each of the paragraphs. When they reach the questions, participants can either answer them straight away or search the text to look for the answers, i.e. answer-seeking behaviour.

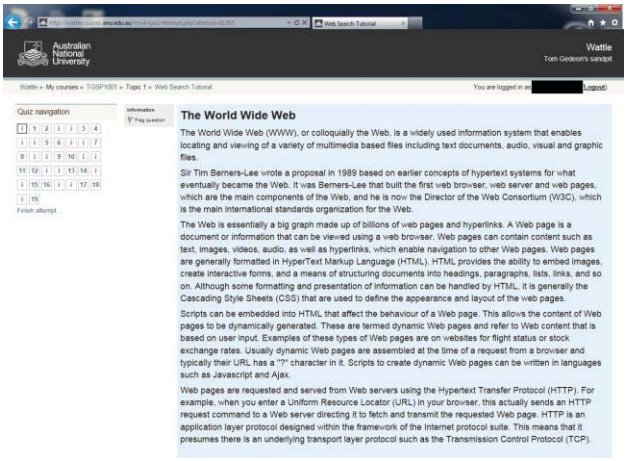

Figure 1: Example of tutorial text only slide

Format B: The questions and tutorial text slide *(Q&T)* is shown to participants. An example of this is seen in Figure 2. Since there are 9 topics there are 9 slides in total displayed in this part of the study.

In this format participants are no longer required to read the text before they see the questions. Our question is, will participants read the text completely or will answer seeking behaviour be observed? Additionally, is there a difference in quiz performance when participants can immediately answer the questions without reading the text?

Format C: The tutorial slide *(T)*, shown in Figure 1, is first shown to participants followed by the questions slide *(Q)* but no access to the content again, see Figure 3. Since

there are 9 topics there are 18 slides in total displayed in this part of the study. This format can be considered to be a control presentation method. In this format the reference text is removed from the questions slide so the participants are forced to answer the questions from understanding and memory. We expect that the worst comprehension scores will be observed for this format.

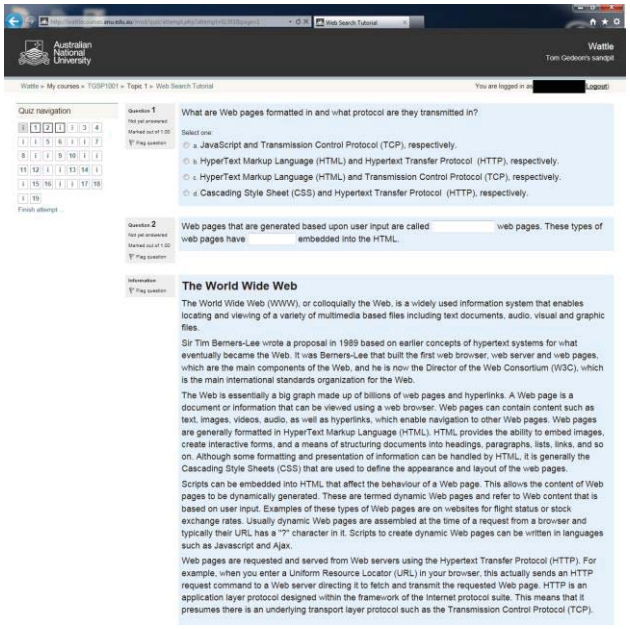

Figure 2: Example of comprehension questions and text

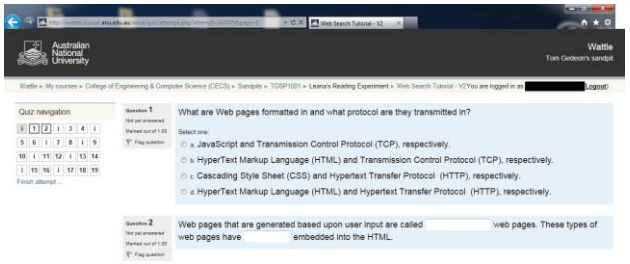

Figure 3: Example of the questions only slide

Format D: The last presentation consisted of displaying a slide with only the questions *(Q)* on it, as seen in Figure 3, followed by the tutorial text slide *(T)* Figure 1, and then again presenting them with the questions slide *(Q)* as in Figure 3. Since there are 9 topics there are 27 slides in total displayed in this part of the study. The reasoning for this format is to mimic a situation where the participants knew what the comprehension questions are but no access to them as they read. The hypothesis is that participants will read the text differently than for formats A and C.

Experiment Setup

The tutorial text was accessible via the online learning environment used at ANU, called Wattle (a Moodle variant). The study was displayed on a 1280x1024 pixel Dell monitor. Eye gaze data was recorded at 60Hz using Seeing Machines FaceLAB 5 infrared cameras mounted at the base of the monitor. The study involved a 9-point calibration prior to data collection for each participant. As the data recorded is a series of gaze points, EyeWorks Analyze was used to pre-process the data to give fixation points. The parameters used for this were a minimum duration of 60milliseconds and a threshold of 5 pixels.

Demographic Information

The study used a total of 33 participants that were divided into the four presentation categories. The experiment used a between-subjects design so that each participant was presented with only one presentation type.

The initial study used format A and can be split into two demographics, COMP1710 students and others. In this analysis only the COMP1710 student subgroup is consider so that a comparison to the follow-up studies can be made, as the remaining presentation formats only used COMP1710 students. The choice of participants is based on the target user group of the eventual online learning environment, which are university students. Further, the COMP1710 students had been exposed to these tutorials throughout the duration of the course and therefore have some experience and familiarity not only using the online learning environment but also with this type of tutorial.

Format A: There were 9 participants (2 female, 7 male) in this group with an average age of 21.3 years (standard deviation 4.7 years, range 17-31 years). English was not their first language for 4 of the participants.

Format B: There were 8 participants (1 female, 7 male) in this group with an average age of 21.8 years (standard deviation 7.9 years), age range 18-41 years. English was not the first language for 3 of the participants.

Format C: There were 9 participants (2 female, 7 male) in this group with an average age of 22.8 years (standard deviation 6.4 years, range 18-37 years). English was not the first language for 5 of the participants.

Format D: There were 7 participants (1 female, 6 male) in this with an average age of 20.1 years (standard deviation 2.8 years, range 17-24 years). English was not the first language for 3 of the participants.

Data Pre-processing

The raw eye gaze data consists of x,y-coordinates recorded at equal time samples (60Hz). Fixation and saccade identification was performed on the eye gaze data. From this point many other eye movement measures are derived. The measures used in this analysis are:

Number of fixations: The sum of fixations recorded for each tutorial page. The number of fixations can be affected by the reading behaviour, text difficulty, and reading skill (Rayner, 1998).

Maximum fixation duration (seconds): The maximum duration of the longest fixation recorded for a tutorial page. Longer fixations can be an indicator of difficulties in processing particular words or due to linguistic and/or comprehension difficulties (Rayner, 1998).

Average fixation duration (seconds): The sum of the duration of all fixations on a paragraph divided by the number of fixations on that paragraph. This measure has been used to predict reading comprehension (Underwood et al., 1990).

Total fixation duration (seconds): The sum of all fixations on complete text. This measure is useful in global text processing analysis (Hyona et al., 2003) because this measures immediate as well as delayed effects of comprehension.

Number of regressions and regression ratio: The number of regressions divided by the total number of saccades on a paragraph. There is evidence that when reading more difficult text more regressions are observed (Rayner et al., 2006).

Reading analysis: Using our combination of two reading detection algorithms (Buscher et al., 2008, Campbell and Maglio, 2001), this is the percentage of saccades classified as being part of reading (read ratio), skimming (skim ratio), and scanning/searching (scan ratio).

RESULTS AND DISCUSSION

This section outlines the analysis performed on the collected eye gaze data. After the gaze points have been converted to fixations and saccades, a number of eye movement measures were calculated and used for this analysis.

First, a comparison of the performance (in terms of quiz results and time taken) observed between the formats is presented. This is in relation to the first hypothesis, that the presentation format will have an effect on these performance measures. The second two hypotheses are dealt with in following subsections. Finally, the effect of different questions on reading behaviour is shown.

Effect of Presentation on Overall Performance

The first aspect analysed is how the participants performed on the tutorial and quiz in each format. The hypothesis is that the different presentation formats will affect how well the participants score, how long it takes them, and also will affect how they perceive their performance. The averages these performance measures are presented in Table 1.

Format	Time (minutes)	Quiz Score	Subjective Score
A *(T-Q&T)*	38.4±9.1	16.2±1.4	7.7±1.8
B *(Q&T)*	26.0±8.0	16.1±1.6	7.9±1.0
C *(T-Q)*	28.2±10.9	11.9±3.6	6.8±2.9
D *(Q-T-Q)*	23.9±7.8	14.1±2.2	7.3±0.95

Table 1: Comparison of means of time taken, quiz score and subjective scores for each format.

Format A took significantly longer to complete than formats B, C and D ($p=0.009<0.05$, $p=0.046<0.05$ and $p=0.004<0.05$, respectively, all two-sided, unpaired t-test). In format A, participants were requested to read the tutorial text and then move on to answering the questions where they also had the option to re-read the content. If a participant did not read the text when it is first shown to them and rather waited for the questions to read through the text to find the answers, then the time taken to complete the quiz should be similar to that of format B. Conversely, if participants read the text before seeing the questions then they should have little need to re-read the text when re-presented with it, instead they would just have to re-read to confirm that they have the correct answers. This is roughly the same as thinking about what the answer is and fully considering it to be correct which is the situation in format C. It would then be expected that the time taken should be similar to that of format C. Given that both formats B and C took significantly less time on average to complete than format A these assumptions do not stand. Instead the participants are re-reading the text so that it increases the total time taken to complete the quiz. There is no difference in time taken to complete the quiz for formats B, C or D.

There is no significant difference between the quiz scores obtained from format A and B. This indicates that reading the tutorial text before being presented with the questions did not improve comprehension scores. Format A has significantly higher quiz scores compared to formats C and D ($p=0.005<0.05$ and $p=0.043<0.05$, respectively). This is also true for format B, where the quiz scores from format B are higher than those from formats C and D ($p=0.009<p=0.05$ and $p=0.073<p=0.1$, respectively). In formats C and D, the participants did not have access to the content as they answered the questions. Participants therefore had to rely on short-term memory and their understanding of the material. This could account for the discrepancy in results. There is no significant difference between the quiz scores obtained from formats C and D. Interestingly, the knowledge of the questions before reading the text did not significantly improve their quiz score or perceived comprehension.

The Pearson's correlation coefficients (r) between the performance measures are shown in Table 2. Notably, in format C there is a very strong positive correlation between the quiz scores and the subjective comprehension scores ($r=0.9$). This indicates that in this presentation format participants are more accurately able to estimate their comprehension level compared to other formats. This relationship is closely followed by the medium positive correlation ($r=0.63$) between quiz scores and subjective scores for format A. When considering formats B and D the participants seem unable to interpret their own comprehension levels and in the case of presentation format B there is even a small negative correlation ($r=-0.15$) between the quiz scores and the subjective comprehension scores. Students should be able to interpret their knowledge level rather than over or underestimate their comprehension level. Over estimation of comprehension could lead to students not learning the material properly as they think they know the answers. In format C the questions are asked after the participants have read the content and they cannot refer back to the material. The participants can seemingly gauge whether they know the answers or not. This effect is still observed to some extent for format A even though participants are able to re-read the tutorial text to answer the questions. The participants are still able to interpret that they have a decreased level of comprehension. In format B the participants only have access to the text whilst answering the questions; this may induce "laziness" in students

whereby they do not fully read the content and thus fail to find key concepts in the text. In format D participants read the questions before reading the tutorial text but then still need to answer the questions without access to the text. In both cases the participants have a false sense of confidence in their answers.

Format	Time & Quiz Score	Time & Subjective score	Quiz Score & Subjective Score
A *(T-Q&T)*	-0.65	-0.46	0.63
B *(Q&T)*	-0.63	0.45	-0.15
C *(T-Q)*	-0.43	-0.25	0.90
D *(Q-T-Q)*	-0.16	-0.29	0.22

Table 2: Correlation analysis (Pearson's R) between performance measures for each format.

There are medium negative correlations between times taken to complete the quiz and actual quiz scores for all formats except D. This correlation is strongest for format A and B. These presentation formats show the questions with the content. The longer the participants spend completing the quiz the lower their quiz score is observed to be. This indicates that the more uncertain a student is the longer they will spend looking for the answers to the questions and therefore may not understand the material well enough to identify the correct answer. In format C participants are presented with the material and then with the questions so participants spending longer trying to answer the questions could account for the correlation between time taken and their quiz score.

There is a medium positive correlation (r=0.45) between times taken and subjective scores for format B. The longer that the participants took the more they thought they understood the material. Yet there is a negative correlation between times taken and quiz score for this format so in fact the opposite is true.

For the rest of the formats there was a negative correlation between the time taken and the subjective score. These results are consistent with the correlations between times taken and quiz score for these formats. The longer that the participants took, the poorer they understood the material but were also unaware of their lack of understanding unlike, for format B.

Under-estimation of understanding can lead to students wasting time on material already understood instead of using the time to learn more material. On the other hand, overestimation of understanding will results in students not learning what they need to and not realising their lack of understanding. In this respect, presentation formats A and C are the optimal format methods for learning.

Additionally, the results from this study show that the formats A and B have the highest quiz scores. For these formats participants were able to consult the text whilst answering the questions to ensure that they understood it before answering. The results show that participants learnt the most from the material in these formats compared to C and D. Since a key aim of any educational material is to promote learning, formats A and B are optimal in this respect.

In conclusion, the analysis of performance measures shows that the initial hypothesis was correct. The different presentation formats have an effect on students' performance, in terms of time and quiz scores, as well as their perceived understanding of the text. The presentation format can therefore be manipulated to optimise the performance outcomes of students, thereby increasing their understanding.

Effect of Presentation on Eye Movements and Reading Behaviour

In this part of the investigation the two central differences in presentation are analysed separately. That is, the tutorial text when shown without the questions versus the tutorial text when shown with the questions. This addresses the final two hypotheses, that different eye movements will be observed for each format, and that the formats will induce different reading behaviours.

Text Page Only

The eye movement measures observed for reading the tutorial text slide are shown in Table 3. Formats A, C and D present the tutorial text slide on its own and these are compared to look at the reading and eye movement behaviour in this scenario.

Format	Number of fixations	Total fixation duration (s)	Number of regressions
A *(T-Q&T)*	278.4±107.3	51.1±32.3	85.8±29.5
C *(T-Q)*	299.7±122.9	68.21±33.1	92.8±42.7
D *(Q-T-Q)*	190.2±98.8	45.88±37.4	67.3±28.8

Table 3: Comparison of eye movement measures

Two types of behaviour are hypothesised; the first is that participants presented with format C will take more care reading the text, as they know they cannot refer to it again whilst answering the comprehension questions. In format D, participants have already seen the questions that they will need to answer. The second hypothesis is that for format D, participants will not (as) thoroughly read the text but rather skim the text to find the paragraphs where they believe the answers are located and read only those paragraphs thoroughly.

The average number of fixations and total fixation time for format C are higher than those from both formats A and D. The difference in number of fixations is statistically significant between formats A and D as well as formats C and D (both $p=0<0.05$). Meanwhile, the difference in total fixation duration time is statistically significant between formats A and C ($p=0.001\leq0.05$) and formats C and D ($p=0.0002<0.05$). Although the participants are showing no difference in the number of fixations whilst reading the text in formats A and C they

are spending longer reading the text in format C according to the fixation time.

There is a significant difference in the average number of regressions observed between formats A and D ($p=0.0002<0.05$) as well as C and D ($p=0<0.05$). Once again, there is no difference in the average number of regressions between formats A and C. It was expected that for format C participants would read the text more thoroughly. Supportive of this, format C has the highest number of regressions, which is indicative of participants regressing back to re-read words or sentences. Given that participants show higher on average numbers of fixations and total fixation duration time for format C, the hypothesis is supported. Since the text is the same in all formats, participants intentionally reading the text more carefully can explain this increase. Furthermore, format D has the lowest number of fixations, lowest total fixation duration, and lowest number of regressions, which is indicative of skimming behaviour. There is no difference in the text between formats, so participants reading less thoroughly can explain the difference in eye movements.

In support of the above evaluation of eye movement measures, and to address the final hypothesis, the reading behaviour is compared between the three formats. The reading behaviour is determined using the reading detection algorithm described in (Buscher et al., 2008) but modified to include scanning behaviour as described by Campbell and Maglio (2001). The percentage of fixation transitions (i.e. saccades and regressions) classified as being part of reading, skimming and scanning are compared between the formats; results are shown in Table 4.

Format	Read Ratio	Skim Ratio	Scan Ratio
A *(T-Q&T)*	0.66±0.23	0.15±0.11	0.19±0.12
C *(T-Q)*	0.52±0.21	0.21±0.09	0.28±0.15
D *(Q-T-Q)*	0.36±0.27	0.24±0.13	0.40±0.24

Table 4: Comparison of reading behaviour of tutorial text

Format A has the highest on average reading ratio and lowest skim and scan ratios. These measures are significantly different from both format C and D (all $p=0<0.05$).

As stated above, the hypothesis was that participants will read the text more thoroughly when presented in format C. Format C has the highest number of regressions, which is indicative of re-reading parts of the text, which supports this conclusion. Longer regressions may be part of skimming and scanning behaviour to find the appropriate location from which to re-read. These results align with those above where participants show less skimming and scanning in format A; they are only reading the text not thoroughly studying it.

As expected, format D has significantly lower reading ratios compared to format A and C ($p=0<0.05$ and $p=0.0001<0.05$, respectively). The skimming and scanning ratios are also significantly higher than those observed for format A (both $p=0<0.05$). The hypothesis was that participants would search the text in format D to find the answers, and only read those parts of the text. Furthermore, the scanning ratio is also significantly higher in format D than it is for C ($p=0.0001<0.05$. Although the skimming ratio is higher for format D than it is for C this is a weak difference ($p=0.06>0.05$).

In conclusion, the eye movement and reading behaviours that are observed for the formats A, C and D are quite different. This difference reflects the participants' overall intentions in reading the text and the goals set for the participants. The purpose of this analysis was to assess the hypotheses that the different presentation formats would affect the eye movements observed and therefore the reading behaviours observed. These hypotheses have been shown to be true. The implications of these findings can be used to support design decisions for eLearning environments. That is, if the teacher wants to promote thorough reading, the goals placed on the reader should not be targeted at certain parts of the text as in format D. Instead, thorough reading is observed for format C where the goal was to understand the text overall.

Questions and Text

Now the slide with the questions and text is considered. Format A consists of two presentations of the text, first on its own and second with the questions. It is hypothesised that the first read through of the text in format A will help participants answer the question and less reference to the text will be needed compared to format B.

Format	Number of Fixations	Total Fixation Duration (s)	Number of Regressions
A *(T-Q&T)*	224.4±171.2	40.4±35.8	95.7±66.9
B *(Q&T)*	374.2±167.8	77.2±43.7	157.5±66.4

Table 5: Comparison of eye movement measures

As seen in Table 5, we observe the average numbers of fixations, number of regressions, and total fixation times are significantly lower for format A compared to format B (all $p=0<0.05$). This indicates that participants spend far less time reading the question and text *(Q&T)* compared to the participants in format B.

The reading behaviour is compared between the formats; shown in Table 6.

Format	Read Ratio	Skim Ratio	Scan Ratio
A *(T-Q&T)*	0.39±0.21	0.18±0.11	0.44±0.17
B *(Q&T)*	0.35±0.19	0.22±0.10	0.43±0.14

Table 6: Comparison of reading behaviour of questions and tutorial text

There is a significantly higher ratio of skimming behaviour observed for format B compared to format A ($p=0.01<0.05$). However, none of the other ratios are

significantly different. This is a surprising finding given the observations from Table 5. A higher read ratio for format B compared to format A was predicted. The reason for this may be that the participants have not read the material in format B, so it is expected that participants presented with format B would read the material more carefully than those in format A. Instead what is seen is the same reading, skimming and scanning behaviour for both the formats. The participants are most likely skimming and scanning through the text quite a lot to find where the answers to the questions may be and only reading the parts that they find most relevant as in format A.

Comparing the reading behaviour to format D (as shown in Table 4), there is also no difference between any of the ratios when comparing format B to D and no difference between the read and scan ratios when comparing formats A to D. There is statistical difference between the skim ratios for formats A and D ($p=0.0009<0.05$). Essentially, the reading behaviour observed for format D can be likened to that observed when the questions are present.

As stated, formats A and B provide participants with the ability to check the content whilst answering the questions. This behaviour is defined as answer seeking (Copeland and Gedeon, 2013). In format A, participants are requested to read the content before moving on to answer the comprehension questions. Participants should have some idea about the answers to the questions as well as have some idea where to find the answers in the text.

In summary, the eye movement behaviours that are observed for formats A and B are considerably different. This difference reflects the fact that participants are requested to read the material before moving on to answer the questions in format A, and participants do not have to do this in format B. Although being asked to read the tutorial text does not improve comprehension results (Table 1), reading the text before answering the questions reduces the effort, measured by number of fixations and total fixation time required for finding and answering the questions.

The analysis in this subsection was designed to investigate the final two broad hypotheses that different eye movements will be observed for each format and that the formats will induce different reading behaviours. There were several sub-hypotheses that were also addressed in this section that broke down the broad hypotheses. Overall both broad hypotheses are validated; the different formats affect the eye movements observed and therefore the reading and answer seeking behaviour observed.

RECOMMENDATIONS FOR DESIGNING ETEXT IN ELEARNING ENVIRONMENTS

The investigation so far has compared how different presentation formats of the text and comprehension questions affects performance, in terms of time and quiz score, and perceived understanding of the text, as well as affects eye movements, and reading behaviour. Conclusions have been made throughout the analysis in regards to these metrics, however, this section will summarise these conclusions to make recommendations for 1) educators designing courseware in an eLearning environment and 2) design considerations for developers of eLearning environments.

Foremost, the analysis has established that the presentation of text and evaluation resources, such as quiz questions, can have a great impact on the learning process and outcomes. The presentation format can be manipulated to optimise the performance outcomes of students, thereby increasing their understanding. Format A has been shown to be the optimal presentation format to do so. The reasoning for this is now summarised. Format A and C were shown to promote accurate self-assessments of comprehension, which in turn minimises both under- and over-estimation of knowledge. Formats A and B have the highest quiz scores. Given that the aims of any educational material is to promote learning, as well as accurate self assessment format A is thus optimal.

The differences in eye movement measures and reading behaviours reflect the overall purpose and goals placed on the reader. If an educator wants to promote thorough reading, the goals placed on the reader should not be targeted with the use of quiz questions. In this case, students only read the parts of the text that they think contains the answers. However, not showing the text with the questions means that the students have to rely too heavily on short term memory and impacts on their quiz scores. The happy medium is format A where the students are requested to read the text and then move on to answer the compression questions. Of course this raises the question of how to actually have students read the text before moving on to the questions and text page. This is where eye tracking can be utilised. The eye tracker can be integrated into the learning environment so that it can monitor reading behaviour. Once the student has read the text then the learning environment will allow the student to move on to the questions.

As can be seen, the common denominator for optimal presentation formats is format A. Educators should consider using this method of presentation when designing text based learning material. In addition, the inclusion of eye tracking into eLearning environments would be beneficial in monitoring the reading behaviours of students.

CONCLUSIONS AND FURTHER WORK

Different presentation formats of text and comprehension questions were investigated to see the effect on students' learning performance, eye movements, and reading behaviour. The purpose is to support informative decisions about how adaptive eLearning environments should present text-based learning material in combination with evaluation resources such as quizzes, as well as to gain insight into how presentation formats affect eye movement and reading behaviour.

Three hypotheses were investigated; the presentation sequence of text and comprehension questions will: 1) have an effect on students' performance, in terms of time and quiz score, as well as their perceived understanding of the text; 2) cause different eye movements to be

observed; and 3) induce different reading behaviours. The hypotheses were validated through the investigation and the conclusions are summarised to give recommendations to educators and developers of eLearning environments.

The presentation format can affect a student's ability to self-identify their comprehension level. A learning environment should present material in such a way that students can accurately perceive their own comprehension so as to alleviate the potential of under or overestimation of their own knowledge.

Importantly, the investigations shows that different presentation formats induce different eye movements and reading behaviours. The differences are due to the goals placed on the reader. In particular, giving the participant the goal of understanding the text to the point that they subjectively believe they understand it causes the participant to read the text more thoroughly then when specific questions are asked of the reader about the contents of the text. Requesting the participants read the text with this intention and then asking them comprehension questions with access to the text again does promote thorough reading behaviour, as well as decrease the need to refer back to the text whilst answering the questions.

The overall conclusion is that the optimal presentation format is format A; presentation of the text on its own followed by presentation of the questions and the text for reference. This format optimised the learning outcomes in addition to promoting thorough reading behaviours.

A limitation of this study is that a relatively small sample size was used. It would be beneficial to run this experiment will a greater number of student participants so that the results can be more generalized. Another limitation is that the majority of the participants are male. This is because the participants were volunteers from a first year computer science course that is predominantly made up of male students. In eLearning environments it has been shown that there is no difference between the observations (Tsianos et al., 2009), further study should be performed to investigate whether this is true for this study. Liu and Huang (2008) showed that there is a significant difference in preference for reading digital text and printed text based on gender so could potentially impact results.

There are only two types of questions that are investigated in this analysis, being multiple choice and cloze questions. These are commonly used question type but not the only types generally available in eLearning environments, so further research should include to see what effect other question types have on the observed behaviour.

Further exploration on how presentation formats on mobile devices would be beneficial given the prevalence of this technology, as this study only considers reading from a computer screen in a University setting.

An area of interest is to investigate the relationship between eye movements and subjective comprehension. Although we touch on it in this investigation we want to investigate this relationship further. In a follow-up study we propose recording eye gaze from participants as they once again read through a tutorial and quiz. After each tutorial slide we can ask participants for their subjective comprehension score as well as other factors regarding how they read and perceived the text.

ACKNOWLEDGMENTS

Thank you our proofreaders and all the participants of the study. Thank you to the reviewers for your thoughtful comments and recommendations.

REFERENCES

Anderson-Inman, L. & Horney, M. A. Supported eText: Assistive Technology Through Text Transformations. Reading Research Quarterly, (2007) 42, 153-160.

Buscher, G., Dengel, A. and Elst, L. V. Eye Movements As Implicit Relevance Feedback. CHI '08 Extended Abstracts On Human Factors In Computing Systems. Florence, Italy: ACM. (2008), 2991-2996.

Campbell, C. S. and Maglio, P. P. A Robust Algorithm For Reading Detection. Proceedings Of The 2001 Workshop On Perceptive User Interfaces, ACM, (2001), 1-7.

Chen, S.-C., She, H.-C., Chuang, M.-H., Wu, J.-Y., Tsai, J.-L. and Jung, T.-P. Eye Movements Predict Students' Computer-Based Assessment Performance Of Physics Concepts In Different Presentation Modalities. Computers & Education, (2014), 74**,** 61-72.

Copeland, L. and Gedeon, T. Measuring Reading Comprehension Using Eye Movements. Cognitive Infocommunications (CogInfoCom), IEEE 4th International Conference On, 2013. IEEE, (2013), 791-796.

Dillon, A. Reading From Paper Versus Screens: A Critical Review Of The Empirical Literature. Ergonomics, (1992) 35, 1297-1326.

Dillon, A. TIMS: A Framework For The Design Of Usable Electronic Text. In: Van Oostendorp, H. & De Mul, S. (Eds.) Cognitive Aspects Of Electronic Text Processing. Norwood, NJ: Ablex. (1996)

Dillon, A. Designing Usable Electronic Text: Ergonomic Aspects Of Human Information Usage, CRC Press (2004)

Dillon, A. & Gabbard, R. Hypermedia As An Educational Technology: A Review Of The Quantitative Research Literature On Learner Comprehension, Control, And Style. Review Of Educational Research, (1998) 68, 322-349.

Fletcher, J. M. Measuring Reading Comprehension. Scientific Studies Of Reading, (2006), 10**,** 323-330.

Frazier, L. and Rayner, K. Making And Correcting Errors During Sentence Comprehension: Eye Movements In The Analysis Of Structurally Ambiguous Sentences. Cognitive Psychology, (1982), 14**,** 178-210.

Horney, M. A. & Anderson-Inman, L. Supported Text In Electronic Reading Environments. Reading & Writing Quarterly, (1999) 15, 127-168.

Hyona, J., Lorch Jr, R. F. and Rinck, M. Chapter 16 - Eye Movement Measures To Study Global Text Processing. In: Hyona, J., Radach, R., H. Deubela2 -

J. Hyona, R. R. & Deubel, H. (Eds.) The Mind's Eye. Amsterdam: North-Holland. (2003)

Hyrskykari, A., Majaranta, P., Aaltonen, A. and Räihä, K.-J. Design Issues Of Idict: A Gaze-Assisted Translation Aid. Proceedings Of The 2000 Symposium On Eye Tracking Research & Applications, ACM Press, (2000), 9-14.

Iqbal, S. T. and Bailey, B. P. Using Eye Gaze Patterns To Identify User Tasks. The Grace Hopper Celebration Of Women In Computing (2004).

Liu, Z. & Huang, X. Gender Differences In The Online Reading Environment. Journal Of Documentation, (2008) 64, 616-626.

Liversedge, S. P. and Findlay, J. M. Saccadic Eye Movements And Cognition. Trends In Cognitive Sciences, (2000) 4**,** 6-14.

Loizides, F. & Buchanan, G. R. Performing Document Triage On Small Screen Devices. Part 1: Structured Documents. Proceedings Of The Third Symposium On Information Interaction In Context, ACM, (2010), 341-346.

Marshall, C. C. & Bly, S. Turning The Page On Navigation. Digital Libraries, 2005. JCDL'05. Proceedings Of The 5th ACM/IEEE-Cs Joint Conference On, IEEE, (2005), 225-234.

Mckay, D. A. Jump To The Left (And Then A Step To The Right): Reading Practices Within Academic Ebooks. Proceedings Of The 23rd Australian Computer-Human Interaction Conference. Canberra, Australia: ACM. (2011), 202-210

Nugrahaningsih, N., Porta, M. and Ricotti, S. Gaze Behavior Analysis In Multiple-Answer Tests: An Eye Tracking Investigation. Information Technology Based Higher Education And Training (Ithet), 2013 International Conference On, IEEE, (2013) 1-6.

O'Hara, K. & Sellen, A. A Comparison Of Reading Paper And On-Line Documents. Proceedings Of The ACM SigCHI Conference On Human Factors In Computing Systems. Atlanta, Georgia, Usa: ACM. (1997), 335-342.

Rayner, K. Eye Movements In Reading And Information Processing: 20 Years Of Research. Psychological Bulletin**,** (1998), 124(3), 372-422.

Rayner, K., Chace, K. H., Slattery, T. J. and Ashby, J. Eye Movements As Reflections Of Comprehension Processes In Reading. Scientific Studies Of Reading, (2006), 10, 241-255.

Sibert, J. L., Gokturk, M. and Lavine, R. A. The Reading Assistant: Eye Gaze Triggered Auditory Prompting For Reading Remediation. Proceedings Of The 13th Annual ACM Symposium On User Interface Software And Technology. California, USA: ACM, (2000) 101-107.

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y. and Yang, F.-Y. Visual Attention For Solving Multiple-Choice Science Problem: An Eye-Tracking Analysis. Computers & Education, (2012), 58, 375-385.

Tsianos, N., Germanakos, P., Lekkas, Z., Mourlas, C. & Samaras, G. Eye-Tracking Users' Behavior in Relation to Cognitive Style within an E-learning Environment. Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on, (2009), 329-333.

Underwood, G., Hubbard, A. and Wilkinson, H. Eye Fixations Predict Reading Comprehension: The Relationships Between Reading Skill, Reading Speed, And Visual Inspection. Language And Speech, (1990), 33**,** 69-81.