


Please cite the Published Version

Sheen, Kimberly Anne  and Luximon, Yan (2021) Effect of in-app components, medium, and screen size of electronic textbooks on reading performance, behavior, and perception. *Displays*, 66. p. 101986. ISSN 0141-9382

DOI: <https://doi.org/10.1016/j.displa.2021.101986>

Publisher: Elsevier BV

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/627124/>

Usage rights:  [Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0](#)

Additional Information: This is an Author Accepted Manuscript of an article published in *Displays*.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Title: Effect of in-app components, medium, and screen size of electronic textbooks on reading performance, behavior, and perception

Names of Authors: Kimberly Anne Sheen and Yan Luximon

Affiliation of Authors: Manchester Metropolitan University and The Hong Kong Polytechnic University

Corresponding Author:

Dr. Yan Luximon

School of Design

The Hong Kong Polytechnic University

Hung Hom, Kowloon

Hong Kong SAR

Email: yan.luximon@polyu.edu.hk

Telephone: +852 2766 5493

Word Count: 6503

Declaration of Interest: Conflicts of interest: none.

Effect of in-app components, medium, and screen size of electronic textbooks on reading performance, behavior, and perception

Although many app-based textbooks are available for students, reading have not been thoroughly outlined. This study aimed to understand how changes from paper to electronic textbooks have affected the academic reading task, investigate student users' perceptions of in-app components and screen sizes, and identify issues affecting in-app components and task requirements. A mixed factorial design experiment was employed. Results showed that there were no significant changes in comprehension and time spent reading between print text and the iPad. Yet, student highlighting, notetaking, and reading behavior and perception significantly changed based on condition. In addition, students struggled to use in-app components and found them frustrating especially when accounting for sentence splitting. The findings presented can assist in understanding the changes in student reading behavior, which can be used to improve interface design of future e-textbooks.

Keywords: electronic textbooks; reading performance; behavior; perception

1 Introduction

Over the past decade, universities across the world have been shifting their practices related to textbooks, adopting electronic textbooks (Chesser, 2011), and publishers such as Pearson are signing commitments to be “digital first” (Overland, 2019). It has been predicted that, once their peripherals are better suited for academic purposes, tablets will be applied for computing throughout academia (McFadden, 2012). This is due to the requirements of academic reading, which require critical thinking, active, and in-depth reading (Schilit et al., 1998). Activities such as notetaking and highlighting are commonly used by students to assist in meeting these requirements and require those peripherals (Liu, 2005; O'Hara and Sellen, 1997). However, these essential features are not yet perfected, especially in e-readers such as the Kindle, and other essential aspects

such as color for text and images is not with some technology (Ferguson et al., 2010). Past research on electronic textbooks vary, but it has been found that students will still use paper due in part to the increase in cognitive overhead demanded by typing compared to writing on paper (Schilit et al., 1998; Liu, 2005) and the lack of high-quality peripherals. When students do discuss e-textbooks, they would rather get their primary e-textbook on a tablet or iPad rather than an e-reader like a Kindle (Weisberg, 2011).

1.1 Reading performance

The findings of past research on reading performance is mostly related to display size and working memory. However, research on display size and its effect on comprehension and recall are often contradictory. Some studies have asserted that the electronic medium itself does not have a negative impact on comprehension and in some instances improves comprehension results (Connell et al., 2012; Bridgeman et al., 2003); however, various aspects of presentation, such as sentence splitting across pages due to smaller display sizes, are likely to negatively influence comprehension because they overload working memory (Dillon, 1992). Working memory is limited; consequently, the complexity involved in a learning task increases the cognitive load on working memory, thereby impeding the acquisition of learning material (Sweller, 1994).

Academic achievement and working memory have a long-established correlation (Yuan et al., 2006). Certain aspects of electronic texts, such as hypertext and typing notes, have been shown to increase cognitive load (DeStafano and LeFevre, 2007; Schilit et al., 1998). While some individual aspects of recall are also found to be hindered by electronic text (Morineau et al., 2005), past research has found that various types of e-readers including those using e-ink technology do not significantly impact learning (Weisberg, 2011).

1.2 Reading behaviors

Although past research has revealed minimal differences in student comprehension of material when changing mediums, changes in student behavior have been observed. Certain aspects of physical textbooks have been shown to be used more often, such as summaries and questions (Woody et al., 2010). Moreover, several studies have established that students spend more time with electronic textbooks (Connell et al., 2012; Daniel & Woody, 2013; Morineau et al., 2005); however, this was found to be opposite in a separate study (Shepperd et al., 2008).

Additionally, various functions available through electronic textbooks, such as notetaking and built-in dictionaries, have actually been reported to be beneficial to students (Demski, 2011). Yet, these in-app components do not provide the same allowance for a variation in meaning of highlights, underlines, and notetaking that paper does, causing reading to be disrupted in electronic texts (O'Hara and Sellen, 1997). Supporting activities are important to student success and an in-depth understanding of the material (Schilit et al., 1998; Kulhavy et al., 1975; Slotte and Lonka, 1999). The design of the in-app components for highlighting and notetaking is important since reading from screens often cause a change in behavior marked by less engagement with the material (Liu, 2005). Past research has found that electronic highlighting and notetaking have several issues such as the blending of electronic notes in with the text, something that is not inherent to physical texts (Schilit et al., 1998), and that electronic notetaking is completed after the reading or with long periods of editing (O'Hara and Sellen, 1997).

1.3 Research gap

Although the past studies described above have briefly addressed this subject, the debate on switching to electronic textbooks remains unresolved. In an article in

Scientific American, Jabr (2013) claimed that reading from a screen is more taxing for the reader, that devices are distracting, and that students are less likely to approach the electronic medium with a learning mindset. This has led other researchers and technology commentators to comment either critically or positively (Jones, 2013; Jones, 2014). Much of the current discussion is either theory or conjecture; thus, it is necessary to approach this topic from a more structured and experimental standpoint. In addition, much of the aforementioned research involves comparing only a physical textbook to an electronic edition or comparing different e-readers (Weisberg, 2011). One example of this is the Kindle app, which has been studied in relation to other reading applications, but not compared to itself with respect to different screen sizes (Jardina and Chaparro, 2013).

Recently, the designs of several devices, such as the iPad Pro, Microsoft Surface, and the Lenovo Yoga, attempt to apply new technology such as the stylus to realize the power of laptops and the mobility of tablets to fill the need for high-quality peripherals. These changes in new technology have given rise to the two driving research questions of this study: *Do the common screen sizes of mobile technology have an effect on student academic reading performance and behavior?* and *Do the in-app supporting components for highlighting and notetaking change the supporting activities of students?*

To address these questions and based on the previous research, the following three hypotheses were developed for this study.

H1) *There will be no significant changes in time spent reading, comprehension and perception when moving from paper to electronic medium of the same size and format.*

H2) Screen size and medium will change reading performance, perception of the reading task, and reading task behavior.

H3) The performance, behavior and perception of the supporting components will change in the electronic devices, especially when compared to paper.

2 Method

This study investigated the reading performance, behavior, and perception through a single electronic textbook using three mobile devices with different screen sizes, together with a paper control. The application used to interact with the material was the same for all three devices. Two in-app components, highlighting and notetaking, were also investigated to identify any differences when moving from paper control to three different sized electronic devices. The use of an e-reader, and thusly e-ink technology, was excluded due to the fact that at the time of the experiment they did not yet support color, which is essential to textbooks in several fields such as science or engineering to convey meaning (Ferguson et al., 2010).

2.1 Participants

There was a total of three qualifying factors for participation in the experiment. First, the participant had to be a current student at the university. Students were chosen as experiment participants because of their familiarity with academic texts. Second, students were required to have normal or corrected vision. Finally, students were required to have a native language other than English and pass a recall and comprehension reading pretest. This pretest was based on SAT recall and comprehension questions. There were three passing score categories: low, medium, and high. If the students did not pass the pretest, they were disqualified from participating in the experiment. Education level was not considered as a qualifying factor for this

experiment because a student's reading level did not necessarily coincide with their education level.

2.2 Experimental design

A mixed factorial design experiment was applied in this study (Table 1). There were three main independent variables in this experiment: screen size, medium and supporting activity. Three different devices with three different screen sizes were used: iPhone 6s, iPad Mini, and full-sized iPad. A paper control was used with text that was the same size and layout as the iPad. All sizes outlined in Table 1 are measured diagonally. In addition, students were placed in individual groups who did different supporting activities including notetaking, highlighting, or nothing while reading.

In-app component Group (Between Subject)	Medium (Within Subject)			
	Paper Medium	Electronic Medium		
	Paper Control (A4 Size)	Mobile Phone (4.7 inches)	Mini-Tablet (7.9 inches)	Full Sized Tablet (9.7 inches)
	A (n=31)	B (n=31)	C (n=30)	
	Read with Paper Control	Read with Mobile Phone	Read with Mini-Tablet	Read with Full Sized Tablet
	Read and Highlight with Paper Control	Read and Highlight with Mobile Phone	Read and Highlight with Mini-Tablet	Read and Highlight with Full Sized Tablet
	Read and Take Notes with Paper Control	Read and Take Notes with Mobile Phone	Read and Take Notes with Mini-Tablet	Read and Take Notes with Full Sized Tablet

Table 1. Mixed factorial design of the experiment.

The dependent variables are identified as reading performance and subjective impression. Reading performance was defined as the ability of the students to recall and comprehend information they have read, behavior related to the number of times they paged backwards, and the time spent reading. Time spent reading was calculated by the number of words in the individual reading divided by the time the student spent on the reading. A post-test, similar to that in Connell (2012) was used to identify the

participants' ability to recall and comprehend a text, Paging and reading time were recorded with a video camera. Subjective impressions were defined as the impression of amount of text, screen size, ease of page turning, format of text, readability, and highlighting or notetaking, if applicable. and a questionnaire was used to elicit their subjective impressions on completing the task under the various conditions.

Each of the three mobile conditions had a set number of sentence splits, lines of sentences on a page, and number of words per line based on the screen size. These variables were preset by the researchers and remained consistent throughout the experiment. The fixed variables for this experiment were eye height, viewing distance, and posture while reading, which were controlled using a fixed stand. The chapter students were asked to read was randomized. Participants read a different chapter on all three of the mobile device sizes and the paper.

The four settings were used in three different devices including mobile phone, mini-tablet, full sized tablet, and a paper control. The paper control was the same size as the full-sized iPad so as to eliminate the confounding factors for analysis of the differences arising from using physical or electronic mediums. Additionally, participants were placed in one of three groups who completed different tasks: reading alone, highlighting while reading, and taking notes while reading.

For balance, participants were separated into the three different groups based on the three pretest classification scores: high, average, and low. An equal number of high, average, and low scores were distributed across the groups consisting of students undertaking different tasks. A total of 31 students were assigned to Group A and completed the readings with the three different screen sizes and paper conditions. A total 31 students in Group B completed the same process but were requested to use the built-in highlighting function or to highlight directly on the paper. The 30 students in

Group C completed the same process but were requested to use the annotation tool while using the devices and take notes directly on paper when using paper.

2.3 Equipment

Three mobile devices were used during this experiment: an iPhone 6s, iPad Mini, and iPad (Table 2). The three forms of mobile devices used were chosen based on their prevalence of usage within the university. All the devices used the same operating system to ensure the least possible difference within the app and subsequent interactions. The text size, brightness, and layout of the devices were preset to ensure that the conditions were the same across all the devices. The devices were presented to the students on a stand and they were not allowed to hold the devices or alter the state of the devices except to change the page, take notes, or highlight depending on their assigned group.

Display Features	iPhone 6s	iPad Mini	iPad
Screen Resolution	1334x750	1024x768	1024x768
Screen Size (diagonal)	4.7 inches	7.9 inches	9.7 inches
Width	2.64 inches	5.3 inches	6.67 inches
Height	5.44 inches	7.9 inches	9.4 inches
Margins	0.4 inch	1 inch	1 inch

Table 2. Screen sizes, margins, and resolutions of mobile devices.

To ensure ease of access across the devices, the Kindle app was selected as the application to present the textbook. Moreover, during a previous study, students at the university had exhibited a general familiarity with the app. The textbook chosen for the students, *American Civilization: An Introduction* (Mauk and Oakland, 2013), was written in English by professors at a Scandinavian university for use in their classrooms when teaching American history. This subject and book were chosen for several

reasons. Because this experiment was conducted with students who did not natively speak English, using a book written with this in mind was important. In addition, the topic of American history was selected because students were highly unlikely to have prior learning experience on the subject, which could then have influenced their comprehension or recall scores. Finally, history was chosen as the general reading area because students can be expected to have experience reading history texts, and there are no common learning strategies associated with the field that would place a student at the university at an advantage. Four individual chapters were chosen from the textbook and educational reading experts deemed them appropriate for the experiment because they were similar in length and reading level.

A Sony HDR-PJ440 Handycam was also used to video record the students interacting with the mobile devices during the reading sessions. The video camera was placed on a tripod located behind the left shoulder of the participants.

2.4 Procedure

Participants were briefed on the experimental procedure and signed a consent form. After taking a reading comprehension and recall pretest, students were assigned to one of the three groups. Before the students began reading, they were briefly shown how to use the app and any functions they were required to use. They also had the opportunity to try navigating within the book and opening the annotation tool or using the highlighting function.

The student then began the reading assigned to the condition. When the participant felt comfortable with the section of text they had read, they informed the facilitator they had completed the condition. Participants were allowed to spend as much time with the text as they wished. After each condition, a rest period of 3 minutes was completed, students filled out a questionnaire regarding their experience during the

reading task, and they were given a post-reading test. Subsequently, the next condition began. After all the conditions were completed, students were asked to compare their experiences in all the conditions and report their general impressions and any issues they encountered.

Each session ranged from 1 hour to 1.5 hours based on the individual's reading speed. Participants were paid for their time after completing the experiment. If a student did not successfully pass the pretest, they were thanked for their time, paid a nominal amount, and dismissed. Participants were allowed to withdraw from the experiment at any time, however all participants chose to complete the experiment.

2.5 Analysis techniques

The quantitative data collected during the experiment was the time a student spent reading, comprehension of the material, and the perception of the various aspects related to reading the material under the different conditions. This data was then analyzed using descriptive statistics. These descriptive statistics included the mean, median, and SD. Interactions between the groups and conditions (the paper control and the three screens) were first investigated using two-way multivariate analysis of variance (two-way MANOVA). When no interactions were found, a nonparametric test, the Kruskal–Wallis H test, was selected to investigate the results in order to more effectively take into account the differences between the students' habits and approach the ordinal data resulting from the scale questions. Qualitative data that was gathered during open-ended subjective questions on the questionnaire and observations of the readings was coded based on the themes identified. Some qualitative data is presented quantitatively as frequencies.

3 Results

3.1 Participants

92 students participated in this research. All of the students were non-native English speakers with varied ethnic backgrounds including but not limited to Chinese, Indian, Turkish, and Polish. 51 of those students were female while 41 were male. The age of participants ranged from 18 years to 50 years with an average age of 25 years and a standard deviation of 6 years. Participants came from all disciplines across the university. The majority of the participants came from the engineering and design disciplines. Specifically, 27 students reported they were from design, 26 from engineering, 18 from medicine, and the others were from various disciplines across the university. One participant reported studying for a higher diploma, 57 reported studying for an undergraduate degree, 11 for a master's degree, and 23 for a PhD degree.

3.2 Reading performance

3.2.1 Time spent reading

This experiment found that the time spent reading changed not only between the paper control and the various mobile device sizes but also between groups. Yet not all of these were found to be significant when analyzing the time spent reading with the mixed ANOVA or the Kruskal-Wallis H test. Table 3 presents the average time spent reading for each condition in words per minute (wpm) by group and condition.

Group	Condition	n	Median	Mean	SD
A	Paper	31	118.52	121.29	36.19
	iPhone	31	110.22	116.07	38.45
	Mini	31	122.07	129.63	38.17
	iPad	31	108.77	127.01	41.84
B	Paper	31	89.36	96.01	28.50
	iPhone	31	96.13	98.80	30.03

	Mini	31	101.50	106.45	34.70
	iPad	31	108.25	104.48	30.61
C	Paper	30	92.58	92.45	34.22
	iPhone	30	96.66	95.84	32.92
	Mini	30	93.14	103.68	35.60
	iPad	30	97.01	99.91	35.61

Table 3. Time spent reading in words per minute by group and condition.

When investigating the change between paper and full-sized iPad with the mixed ANOVA, there was no significant difference in mean reading speed, $F(1, 82) = 2.594$, $p = 0.111$, partial $\eta^2 = 0.031$. Moreover, the Kruskal-Wallis H test was conducted on the time spent reading for each condition separated by group, thus investigating the effect of display size within each group. Significance across groups was found for the paper condition ($X^2(2) = 11.816$, $p = 0.003$), the iPad Mini Condition ($X^2(2) = 8.343$, $p = 0.015$), and the iPad condition ($X^2(2) = 7.014$, $p = 0.030$). The distribution of the time spent reading in the iPhone 6s was the same across all of the groups ($X^2(2) = 4.418$, $p = 0.110$). The Kruskal-Wallis H test was also conducted on the time spent reading for each condition with combined groups, investigating the effect of display size and component. No significance was found ($X^2(3) = 4.004$, $p = 0.261$).

3.2.2 Comprehension scores

Comprehension and recall were measured using a short multiple-choice test based on the reading. The test was conducted after a 3-minute break after each reading. The median and mean comprehension scores categorized by group and condition are shown in Table 4 below.

Group	Condition	n	Median	Mean	SD
A	Paper	31	6.00	6.03	2.11
	iPhone	31	5.00	5.00	2.10

	Mini	31	6.00	5.48	1.93
	iPad	31	5.00	5.55	2.17
B	Paper	31	5.00	5.42	1.96
	iPhone	31	4.00	4.58	2.05
	Mini	31	5.00	5.52	2.59
	iPad	31	5.00	5.39	2.73
C	Paper	30	5.00	4.90	2.19
	iPhone	30	5.00	5.63	2.39
	Mini	30	5.00	5.37	2.14
	iPad	30	5.00	5.00	1.89

Table 4. Comprehension score for each device by group and condition.

The Kruskal-Wallis H test was conducted on the comprehension scores for each condition separated by group. No significant differences in comprehension scores were found for the paper condition ($X^2(2) = 4.881$, $p = 0.087$), iPhone 6s condition ($X^2(2) = 3.495$, $p = 0.174$), iPad Mini condition ($X^2(2) = 0.204$, $p = 0.903$), or the iPad condition ($X^2(2) = 0.657$, $p = 0.720$) across all groups. The Kruskal-Wallis H test was also conducted on the comprehension score for each condition with combined groups. No significance was found ($X^2(3) = 2.602$, $p = 0.457$).

3.3 Student behavior

3.3.1 Reading behaviors

There were many changes in the behavior of students when using in-app components to support their reading. Overall, when moving to the Kindle app, students took fewer notes and used the highlighting tool less frequently (Table 8).

Condition	Back Pages (Groups A, B & C)				Unique Highlights (Group B)				Number of Words (Group C)			
	n	Median	Mean	SD	n	Median	Mean	SD	n	Median	Mean	SD
Paper	92	0	1.11	1.63	31	27	32.39	25.82	30	57	72.27	73.3
iPhone 6s	92	2	4.53	5.99	31	10	10.26	8.18	30	4	8.37	12.93
Mini	92	1	2.1	2.88	31	9	11.16	11.43	30	5	13.1	22.81

iPad	92	1	2.15	3.07	31	10	12.52	13.44	30	6.5	14.37	23.25
------	----	---	------	------	----	----	-------	-------	----	-----	-------	-------

Table 8. Mean, median, and standard deviation of back pages, unique highlights, and number of words in notes.

The Kruskal-Wallis H Test was conducted within Groups B and C to better understand the significance and changes across the paper control and three devices. Each highlight was counted not by number of words highlighted, but by each unique highlight made by students. In the paper format, circles, crosses, and stars were counted as a unique highlight. Notes were counted by words. In the paper format, non-word notes were counted as a unique word. The distribution of unique highlights ($X^2(3) = 26.409$, $p = 0.000$) and words ($X^2(3) = 51.838$, $p = 0.000$) were not the same across the four conditions for both groups. In addition, both group and condition were also analyzed by the number of times students paged backwards in the condition using the Kruskal-Wallis H Test. The distribution of back pages across groups was found to be the same for each of the four condition of paper ($X^2(2) = 5.485$, $p = 0.064$), iPhone 6s ($X^2(2) = 0.928$, $p = 0.629$), iPad Mini ($X^2(2) = 0.393$, $p = 0.821$), and iPad ($X^2(2) = 3.882$, $p = 0.144$); whereas the distribution across conditions was not the same ($X^2(3) = 24.498$, $p = 0.000$).

3.3.1.1 Student highlighting and notetaking

During the experiment, students frequently reported that the platform did not support their habits. Analysis of the paper control revealed that 73.3% students in the notetaking group used more visual notetaking styles that are not supported by the app's simple textbook input (Fig. 1). Additionally, 16.1% of students in the highlighting group used other marks such as circling or starring to help identify the importance of the material in addition to simple highlighting (Fig. 2).

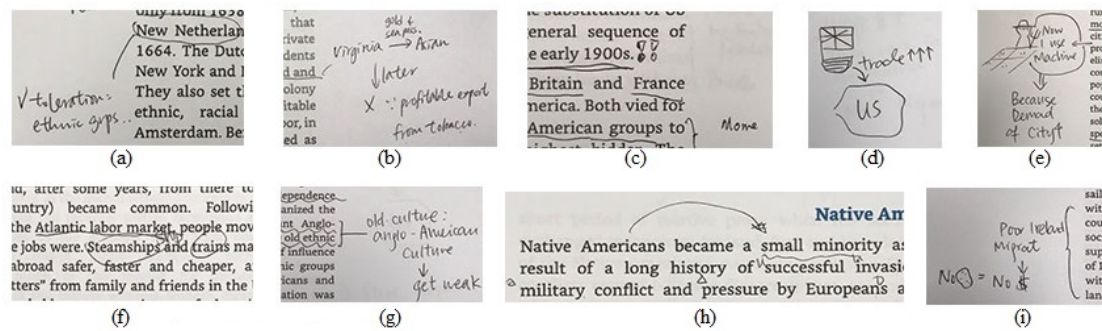


Figure 1. Examples of notetaking behavior not supported by the Kindle app: (a) words with connections, (b) symbols to aid understanding, (c) symbols to indicate importance, (d) images that symbolize thoughts without words, (e) pictures to give full meaning, (f) connections with words & arrows for meaning, (g) subcategories, (h) various symbols for meaning, and (i) symbols with individual meaning as a phrase

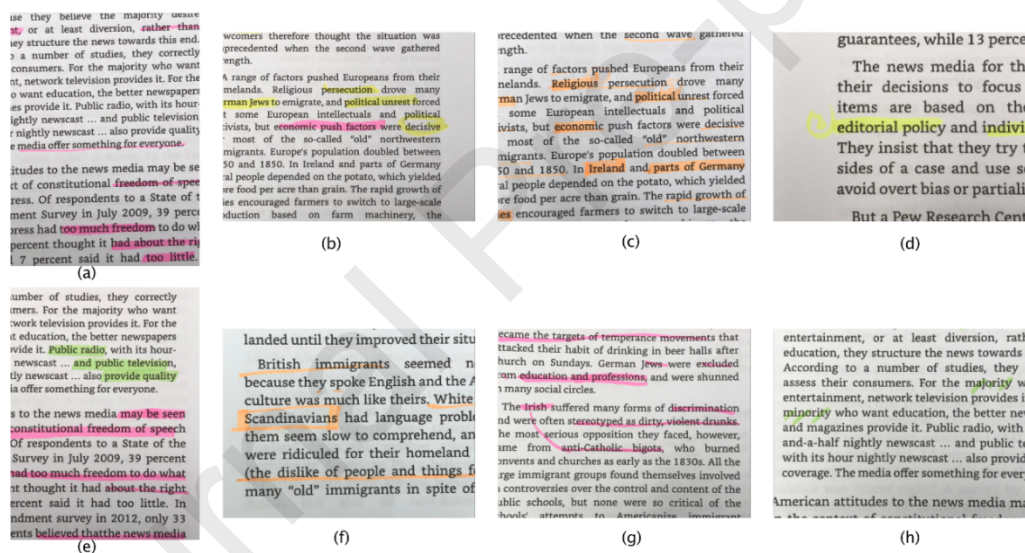


Figure 2. Examples of highlighting behavior not supported by the Kindle app: (a) underlining and highlighting, (b) circling, (c) numbering, (d) using symbols, (e) using multiple colors and underlining, (f) boxing, (g) using connection lines, and (h) crossing words.

The most common approach to reading was identifying key words, terms, or concepts, and then reading while focusing on those; 45 of the participants reported this method. The second most common approach (20 participants) was to simply read though the material. Other approaches were not very common and varied greatly.

Fifty-four students reported that their approach to reading changed across the conditions (59% of participants); this was most common for the iPhone (35 participants; 38%). Seventeen students reported that they changed their approach between paper and the three devices, and two students reported a change only for the iPad Mini and iPhone. The commonly reported reasons for the change were eye fatigue, loss of focus when using electronic devices, increased sentence splitting for the iPhone, screen dimming due screen time, and the fact that supporting activities were more difficult to use in the electronic medium.

3.3.2 Observations and reported issues

Several issues were identified during the experiment through facilitator observation and through students' reports after each task. Some of these were related solely to the specific mobile device and others were found across all of the devices.

Device specific issues were found on the iPhone 6s. This was due to the small screen size. Students frequently reported issues reading the material due to the limited information on the screen. Students also reported difficulties in notetaking and highlighting. The notetaking issues were related to the small size of the keyboard, and the highlighting issues were frequently related to the increased sentence splitting caused by the small screen size. Students had to highlight text on two different pages and reported that this was difficult and often time consuming.

Issues with the in-app components used in both Groups B and C were reported across devices. Highlighting was reported to be difficult to complete without using more than one highlighting movement to cover the complete sentence. In addition to the student reports, the facilitator observed students experiencing issues when attempting to make an existing highlighted section longer or shorter; they occasionally completely

removed the highlighted section and started again. Students also reported that they often went back or forward a page while attempting to highlight a passage.

It was observed that students also frequently struggled with the annotation tool. Students reported that the keyboard was not ideal for inputting their notes. Many students reported frustration with the fact that they could not move the textbox popup and thus could not see the text that they were referencing. Instead, they had to spend more time repeatedly opening and closing the textbox until they had edited their notes to their satisfaction. In addition, the facilitator observed many students becoming confused when attempting to access the annotation component. When they selected a word or phrase, students would initially look at the larger dictionary, thesaurus, and Wikipedia boxes that pop up below the toolbar. Some students even reflexively attempted to select those options. Students also showed frustration with the way the notetaking icon was represented. Several students deleted the note to try and select a phrase again, only to have the same icon appear. A few of these students then used the highlighting component to identify the corresponding phrase.

3.4 Student perceptions and preference

3.4.1 Student perceptions

Perceptions based on various aspects of the conditions were elicited from participants immediately after the reading under that condition was completed. When students were asked to think about these aspects, they were requested not to take into account any other condition that they may have already completed. The perceptions of the amount of text presented and the screen size were elicited using a 5-level scale in which 3 was ideal. Table 5 presents the median, mean, and SD of these two perceptions. The perception of turning pages, format of the text, general readability of the text,

highlighting and notetaking were elicited using a 5-level scale that ranged from 1 (*very bad/difficult*) to 5 (*very easy/good*). Table 6 presents the median, mean, and SD for the perception of turning pages, format of text, and general readability of the text. Table 7 presents the perceived ease of highlighting and notetaking under the four conditions.

Group	Condition	n	Perception of Text Amount			Screen Size		
			Median	Mean	SD	Median	Mean	SD
A	Paper	31	3.00	2.74	0.68	2.00	1.81	0.54
	iPhone	31	4.00	4.06	0.77			
	Mini	31	3.00	2.94	0.51			
	iPad	31	3.00	2.65	0.55			
B	Paper	31	3.00	2.90	0.47	2.00	1.87	0.62
	iPhone	31	4.00	3.68	0.95			
	Mini	31	3.00	2.90	0.30			
	iPad	31	3.00	2.68	0.65			
C	Paper	30	3.00	3.07	0.69	2.00	1.73	0.58
	iPhone	30	4.00	4.00	0.91			
	Mini	30	3.00	2.87	0.57			
	iPad	30	3.00	2.87	0.43			

Table 5. Perception of the amount of text presented and screen size by group and condition.

Group	Condition	n	Turn Pages			Text Format			Readability		
			Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
A	Paper	31	4.00	3.65	0.84	4.00	3.45	0.81	4.00	3.84	0.58
	iPhone	31	4.00	3.61	0.72	2.00	2.26	1.06	2.00	2.55	0.93
	Mini	31	4.00	4.00	0.58	4.00	3.61	0.62	4.00	3.58	0.81
	iPad	31	4.00	4.03	0.71	4.00	3.52	0.85	4.00	3.52	0.93
B	Paper	31	4.00	4.03	0.75	4.00	3.58	0.85	4.00	3.90	0.79
	iPhone	31	4.00	3.65	0.76	2.00	2.32	0.91	2.00	2.42	0.92
	Mini	31	4.00	3.48	0.85	4.00	3.77	0.62	4.00	3.81	0.54
	iPad	31	4.00	3.61	0.76	4.00	3.39	0.88	4.00	3.61	0.76
C	Paper	30	4.00	3.93	0.94	4.00	3.60	0.86	4.00	3.57	0.94
	iPhone	30	3.00	3.33	0.96	2.00	2.30	1.09	2.00	2.30	0.84

Mini	30	4.00	3.63	0.89	4.00	3.67	0.80	4.00	3.60	0.89
iPad	30	4.00	3.77	0.73	4.00	3.77	0.63	4.00	3.80	0.85

Table 6. Perception of the ease of turning pages, text format, and readability for each device by group and condition.

Condition	Ease of Highlighting				Ease of Notetaking			
	n	Median	Mean	SD	n	Median	Mean	SD
Paper	31	5.00	4.35	0.88	30	4.00	4.00	1.02
iPhone	31	3.00	3.03	0.88	30	2.50	2.67	1.12
Mini	31	4.00	3.35	0.92	30	3.00	3.13	1.11
iPad	31	4.00	3.29	1.01	30	4.00	3.47	0.90

Table 7. Perception of the ease of highlighting and notetaking for each condition.

The Kruskal-Wallis H test was conducted on the perceptions of the various aspects of the conditions for each condition separated by group. No significant difference across groups was found for the perception of the amount of text shown in any of the four conditions: paper ($X^2(2) = 3.134$, $p = 0.209$), iPhone 6s ($X^2(2) = 3.087$, $p = 0.214$), iPad Mini ($X^2(2) = 0.367$, $p = 0.832$), and iPad ($X^2(2) = 3.076$, $p = 0.215$). Significance across groups was found for the perception of the ease of turning pages in the iPad Mini condition ($X^2(2) = 6.322$, $p = 0.042$). No significant difference was found in the other three conditions: paper ($X^2(2) = 2.883$, $p = 0.237$), iPhone 6s ($X^2(2) = 3.015$, $p = 0.221$), and iPad ($X^2(2) = 4.839$, $p = 0.089$). Also, no significant difference across groups was found for the perception of the readability of the text in any of the four conditions: paper ($X^2(2) = 2.126$, $p = 0.345$), iPhone 6s ($X^2(2) = 0.714$, $p = 0.700$), iPad Mini ($X^2(2) = 1.616$, $p = 0.446$), and iPad ($X^2(2) = 1.622$, $p = 0.444$). In addition, no significant difference across groups was found for the perception of the format of the text shown in any of the four conditions: paper ($X^2(2) = 0.382$, $p = 0.826$), iPhone 6s ($X^2(2) = 0.181$, $p = 0.913$), iPad Mini ($X^2(2) = 1.460$, $p = 0.482$), and iPad ($X^2(2) = 3.066$, $p = 0.216$). Finally, no significant difference across groups was found for the

perception of the screen size in any of the three digital conditions: iPhone 6s ($X^2(2) = 0.810$, $p = 0.667$), iPad Mini ($X^2(2) = 0.021$, $p = 0.989$), and iPad ($X^2(2) = 0.777$, $p = 0.678$).

Finally, the Kruskal-Wallis H test was also conducted with combined groups across all conditions. Significance was found in all perceptions across conditions.

Overall, the distribution of the perception of the amount of text ($X^2(3) = 120.060$, $p = 0.000$), ease of turning pages ($X^2(3) = 8.079$, $p = 0.044$), format of the text ($X^2(3) = 100.310$, $p = 0.000$), readability ($X^2(3) = 107.608$, $p = 0.000$), screen size ($X^2(2) = 149.672$, $p = 0.000$), ease of highlighting ($X^2(3) = 32.460$, $p = 0.000$), and ease of notetaking ($X^2(3) = 21.696$, $p = 0.000$) was different across all four conditions.

3.4.2 Student preferences

After the experiment was completed, students were asked to compare the mediums used for the experiment and report which they felt was worst for academic reading. A total of 84 (91%) participants reported that the iPhone was the worst device for completing academic readings. These reports were relatively evenly spread across the three groups (28 in A, 27 in B, and 29 in C). Two participants reported that the iPad Mini (2%) was the worst. Both of these participants were from Group C. No participants from Group A stated that the iPad Mini was the worst. Similarly, two participants reported that the iPad was the worst (2%). One participant was in Group B and the other was in Group C. Finally, four participants (4%) reported that paper was the worst medium to read on. Three of the participants who reported that paper was the worst were in Group A and one was in Group B. No participants in the notetaking group, Group C, reported that the iPad was the worst.

Subsequently, students were asked to report which medium they felt was best. Many students stated that paper and one of the devices were both the best, unlike when they were asked which was worst, when their answers were more certain. 57 students provided two answers when asked which was best, and one of those students offered three answers. Fifteen of the participants reporting two best mediums were in Group A, 23 were in Group B, and 19 were in Group C. The participant who answered that three were best was in Group B. All of the participants who selected paper asserted that it was the first among their answers.

54 participants (59%) reported that the iPad was the best. Of these participants, 33 (61%) also replied that paper was best. Furthermore, 37 (40%) participants reported that the iPad Mini was the best. Similarly, 22 (59%) of these participants also reported that paper was the best. In total, 57 participants (62%) asserted that paper remained the best medium for academic reading. Finally, three (3%) participants believed that the iPhone was the best device to read from. These three participants also reported that paper was best. Many students explained that their choice of paper was due to nostalgia, although some stated that devices simply did not support their reading habits. This was especially prevalent in Group C. Students also cited eye fatigue as another reason for hesitancy to state that a specific device was superior to paper.

4 Discussion

4.1 Student performance

4.1.1 Time spent reading

In general, the time spent reading across all groups and conditions was not significantly different. These findings are important when examining if there is a change between

paper text and electronic text in the same size and format; this was especially apparent in Group A. While the results were not significant, examining the median time spent reading had some differences, paper resulted in a faster reading time than most of the electronic mediums in Group A, which is in line with previous research (Daniel and Woody, 2013). However, according to the means, reading was completed faster on the iPad Mini and iPad, which is in line with a study that found that the time spent reading decreased when using the electronic version of texts (Shepperd et al., 2008). This shorter time spent reading in the paper medium was not sustained in Groups B and C when the components were introduced. This discrepancy can be attributed to the increase in highlights and notes made by students in those groups when using the paper medium, which differed from the other conditions. Past research has considered this an overall advantage as the inclusion of notetaking increases the time spent with the text (Kulhavy et al., 1975) which then in turn allows students to feel they are deepening their understanding of the text (O'Hara and Sellen, 1997). Not only do students feel their comprehension of the material is improved but it also encourages active reading and in-depth reading which is required for academic reading unlike a novel (Schilit et al., 1998; Liu, 2005; Wolf and Barzillai, 2009; Hartley, 1990).

4.1.2 Comprehension

While students often feel that supporting activities such as notetaking improve their comprehension, no significant differences in comprehension were found across devices or with the use of components. This finding is not surprising as the way in which supporting activities were used and how comprehension was consequently tested in this experiment is not a reflection of the real-world situation. This is something that has been found in similar studies in the past (Riley and Dyer, 1979). In this study,

supporting activities were not allowed to be reviewed before the comprehension test in this experiment. They were only implemented to understand the behavior change across conditions and to identify if there were major changes in comprehension while using the components when moving from the paper to the full-sized iPad.

Similar to the findings of this study, past research in physical texts have found that the addition of highlighting and underlining do not have an immediate effect on the retention of material, yet the same study found that active highlighting and underlining are valuable to students as they do, at least weakly, increase the overall retention of the material (Fowler and Barker, 1974). Other studies have shown that supporting activities are more often used by students during revision or paper writing to avoid needing to reread material, thereby providing them with subsequent insights at a different time which allows for the deeper understanding of the material (Slotte and Lonka, 1999; Kuhavy et al., 1979; Dyer et al., 1979).

While there was an absence of significant changes in immediate comprehension, it is still noteworthy that the mean comprehension score for iPhones for Group C was the highest, which is similar to past findings (Connell et al., 2012). This result was not duplicated in the groups highlighting or reading alone. This was most likely due to students retaining the information they were going to take notes on in their short-term memory. Students reported that they had to try to quickly memorize the text due to the increased sentence splitting and the fact that the notetaking input functionality took up the entire screen, unlike for the iPad Mini or iPad.

4.2 Reading behavior

There was a significant difference in the number of unique highlights and notes taken across conditions. The paper condition afforded the most unique highlights and notes. Past research has found that paper allows for many benefits that electronic textbooks do

not have such as integrated reading and notetaking, which allows rich and varied marks to assist in meaning making, and ease of finding the notes or highlights on a page (O'Hara and Sellen, 1997; Schilit et al., 1998).

These benefits and the variance on paper text was also found in this study which showed that many students' notetaking and highlighting habits differed depending on the input supported by the electronic devices. The more visual style of notetaking or highlighting even changed across students or within one student's paper condition. Words were sometimes highlighted fully, underlined, numbered, crossed out, or circled (See Figure 2). In the notetaking groups, students created arrows, timelines, pictures with words, standalone pictures, and other features to help them create associations across the text and within their knowledge base (Figure 1). The decrease of notes and highlights in the electronic textbooks is due to the struggle that students reported because the in-app components do not support their habits.

Although past surveys have shown that hundreds of thousands of students wish to take notes or highlight in their electronic textbooks (Warren, 2010), these components are not yet optimized for students using the Kindle app. In fact, these essential features are not yet perfected in any e-reader (Ferguson et al., 2010). Although all the students in Group B were able to use the highlighting function in this study, albeit with difficulty, past research found that less than 80% of the examined users of the Kindle app on an iPad 3 were able to fully utilize the highlighting function within 1 minute (Jardina and Chaparro, 2013). Although members of Group C were able to use the notetaking functionality, some students refused to take notes in certain electronic devices because of the difficulties associated with the component.

4.3 Student preference

The preferences that students reported mirrored many of the aspects observed during the experiment. The iPhone 6s consistently scored poorly and was thought to be the worst of the reading mediums. It also caused the most changes in reading behavior. Changes in reading behavior were often quite noticeable across the conditions and groups. For example, back paging was found to be significantly different in all the devices and paper textbook, and the iPhone 6s yielded almost double the average number of back pages. Students reported this phenomenon after reading on the iPhone 6s and believing that it was due to the sentence splitting and lack of text available to them on the screen.

The perceptions of the various aspects and the overall reported best device showed differences. At the end of the sessions, students most often reported that the iPad was the best device for completing academic readings and that paper was the best medium overall. However, the perceptions of the individual aspects revealed a different response. The iPad Mini and paper were most frequently rated the best. However, many students stated that there was not a substantial difference between their perception of the iPad and the iPad Mini when questioned. Overall this study had similar findings to past research which has repeatedly found that students will still prefer to use paper for in-depth reading due to various underlying reasons (Schilit et al., 1998; Liu, 2005).

5 Conclusion

The findings from this study mostly support all three of the original hypotheses. In support of H1, the findings revealed that there was no significant change in comprehension and perception of aspects of the reading when moving from paper to the iPad. In addition, in support of this same hypothesis, the time-spent reading was very similar for students reading in a print medium in the same size and format as the

electronic textbook. H2 was mostly supported by the findings of this experiment. Perception of the reading task and the behavior of notetaking, highlighting, and back paging differed across the four conditions. While the time spent reading did differ across the conditions, this change was not significant; thus, it does not support H2. The marked decrease in time spent reading was when students used the iPhone 6s in all groups and students reported the least satisfaction with reading on that device in general due to their greater difficulties in reading the material and using the functions due to the sentence splitting increase. In line with H3, students created significantly more unique highlights and notes in terms of number of words and different styles in paper format than any of the electronic devices. In addition, students struggled to use the functions in general and found them frustrating.

Although using electronic textbooks may not have a significant effect on student reading performance, their behavior does change substantially. The Kindle app in its current iteration is unsuitable for academic reading because it does not support the way that students highlight and take notes, behaviors that are considered important for students who must subsequently revise for examinations or write papers. However, this investigation and subsequent findings show that students perform similarly to the paper control and are open to the use of devices with a screen that has a diagonal size of 7.9 inches or above. If supporting activity inputs are better designed to support student needs when highlighting and notetaking, students would likely find electronic textbooks of this size suitable for their academic needs.

To understand more holistically how this may affect comprehension or recall, a similar experiment could be repeated in which students return at a later time and review their notes before taking the comprehension and recall post-test. Additionally, a similar experiment could be conducted to identify differences in paper reading and that of e-

readers using an e-ink technology. E-paper materials were not considered in this experiment as they were not found to be prevalent with students and it has been predicted that the uptake in tablets are likely to make e-readers obsolete (Gross, 2013). However, the benefits of e-paper, such as the ability to be read in sunlight, and the invention of color e-paper in 2020 may make electronic displays more suitable for academic reading due to their similarities with paper. Overall, findings from this experiment can assist in understanding how student behavior changes and what design changes should be made to current electronic textbook services to enable full academic support.

Funding: This work was supported by the Hong Kong Fellowship Scheme (1-904Z).

References

- Bridgeman, B., Lennon, M.L., Jackenthal, A., 2003. Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Test Performance. *Appl. Meas. Educ.* 16(3), 191-205.
- Chesser, W.D., 2011. The e-textbook revolution. *Libr. Technol. Rep.* 47(8), 28.
- Connell, C., Bayliss, L., Farmer, W., 2012. Effects of eBook Readers and Tablet Computers on Reading Comprehension. *Int. J. Instr. Media.* 39(2), 131-140.
- Daniel, D.B., Woody, W.D., 2013. E-textbooks at what cost? Performance and use of electronic v. print texts. *Comput. Educ.* 62, 18-23.
- Demski, J., 2011. Ell to go: Two schools transform their Ell programs by giving students around-the-clock access to some of the latest mobile devices. *T.H.E. J.* 38(5), 28.
- DeStefano, D., LeFevre, J.A., 2007. Cognitive load in hypertext reading: A review. *Comput. Hum. Behav.* 23(3), 1616-1641.
- Dillon, A., 1992. Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics.* 35(10), 1297-1326.

- Dyer, J.W., Riley, J., Yekovich, F.R., 1979. An Analysis of Three Study Skills: Notetaking, Summarizing, and Rereading¹. *J. Educ. Res.* 73(1), 3-7.
- Ferguson, J., Giordano, J., Jerome, E., Lipkin, A., Pinnolis, J., Widmer, L., Spring 2010. LTS E-textbook analysis.
http://bir.brandeis.edu/bitstream/handle/10192/23919/EtextbookReport_LTS-RIS-Spring2010.pdf (accessed 15 June 2017).
- Fowler, R.L., Barker, A.S., 1974. Effectiveness of highlighting for retention of text material. *J. Appl. Psychol.* 59(3), 358-364.
- Gross, D., 2013. As tablets boom, e-readers feel the blast. CNN.
<http://edition.cnn.com/2013/02/28/tech/gaming-gadgets/tablets-replacing-e-readers/index.html> (accessed 20 February 2020).
- Hartley, J., 1990. Textbook design: Current status and future directions. *Int. J. Ed. R.* 14(6), 533-541.
- Jabr, F., 2013. Why the brain prefers paper. *Sci. Am.* 309(5), 48-53.
- Jardina, J.R., Chaparro, B.S., 2013, September. Usability, engagement, and satisfaction of two e-textbook applications. In *Hum. Fac. Erg. Soc. P.* 57(1), SAGE Publications, Los Angeles, pp. 482-486.
- Jones, J., 27 Feb. 2014. Reading Comprehension: Paper or Screen?. dmlcentral. Digital Media and Learning Initiative. Blog. (accessed 27 July 2014).
- Jones, J., 5 Nov. 2013. How Does Electronic Reading Affect Comprehension? dmlcentral. Digital Media and Learning Initiative. Blog. (accessed 27 July 2014).
- Kulhavy, R.W., Dyer, J.W., Silver, L., 1975. The effects of notetaking and test expectancy on the learning of text material. *J. Educ. Res.* 68(10), 363-365.
- Liu, Z., 2005. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *J. Doc.* 61(6), 700-712.
- Mauk, D.C., Oakland, J., 2014. American civilization: an introduction. London: Routledge.
- McFadden, C., 2012. Are Textbooks Dead? Making Sense of the Digital Transition. *Publish. Res. Q.* 28(2), 93-99.
- Morineau, T., Blanche, C., Tobin, L., Guéguen, N., 2005. The emergence of the contextual role of the e-book in cognitive processes through an ecological and functional analysis. *Int. J. Hum-Compt. St.* 62(3), 329-348.
- O'Hara, K., Sellen, A., 1997, March. A comparison of reading paper and on-line documents. In *Proc. SIGCHI Conf. Hum. Factor. Comput. Syst. ACM*, pp. 335-342.

- Overland, S. 16 Jul. 2019. Pearson Turns the Page on College Textbooks as Digital Courseware Demands Grows. Pearson. Blog. (accessed 25 July 2019).
- Riley, J.D., Dyer, J., 1979. The effects of notetaking while reading or listening. *Lit. Res. Instr.* 19(1), 51-56.
- Schilit, B.N., Golovchinsky, G., Price, M.N., 1998, January. Beyond paper: supporting active reading with free form digital ink annotations. In *Proc. SIGCHI Conf. Hum. Factor. Comput. Syst. ACM*, pp. 249-256.
- Shepperd, J.A., Grace, J.L., Koch, E.J., 2008. Evaluating the electronic textbook: is it time to dispense with the paper text?. *Teach. Psychol.* 35(1), 2-5.
- Slotte, V., Lonka, K., 1999. Review and process effects of spontaneous note-taking on text comprehension. *Contemp. Educ. Psychol.* 24(1), 1-20.
- Sweller, J., 1994. Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* 4(4), 295-312.
- Warren, J.W., 2010. The Progression of Digital Publishing: Innovation and the Evolution of E-books. *Int. J. Book.* 7(4).
- Weisberg, M., 2011. Student attitudes and behaviors towards digital textbooks. *Publish. Res. Q.* 27(2), 188-196.
- Wolf, M., Barzillai, M., 2009. The importance of deep reading. *Educ. Leadership.* 66(6), 32-37.
- Woody, W.D., Daniel, D.B., Baker, C.A., 2010. E-books or textbooks: Students prefer textbooks. *Compt. Educ.* 55(3), 945-948.
- Yuan, K., Steedle, J., Shavelson, R., Alonzo, A., Oppezzo, M., 2006. Working memory, fluid intelligence, and science learning. *Educ. Res. Rev.-Neth.* 1(2), 83-98.