Reading on Small Displays: Reading performance and perceived ease of reading

PIIASTIINA TIKKA

PhD

2013

Reading on Small Displays: Reading performance and perceived ease of reading

PIIASTIINA TIKKA

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

Research undertaken in the Faculty of Health and Life Sciences

April 2013

Abstract

The present thesis explores and discusses reading continuous text on small screens, namely on mobile devices, and aims at identifying a model capturing those factors that most influence the perceived experience of reading. The thesis also provides input for the user interface and content creation industries, offering them some direction as to what to focus on when producing interfaces intended for reading or text-based content that is likely to be read on a small display.

The thesis starts with an overview of the special characteristics of reading on small screens and identifies, through existing literature, issues that may affect fluency and ease of reading on mobile devices. The thesis then presents six experiments and studies on reading performance and perceived experience when reading on small screens. The mixed-methods research presented in the thesis showed that reading performance and subjective perception of reading fluency and ease do not always correspond, and perceived experience can have a strong influence over an end-user's choice of whether to access text based content on a small display device or not. The research shows that it is important to measure interface quality not only in terms of functionality, but also for the user experience offered – and, ideally, to measure experience through more than one variable.

The thesis offers a factor model (mobile reading acceptance model) of those factors that collectively influence subjective experience when reading via small screens. The key factors in the model are visibility of text, overview of contents, navigation within the contents and interaction with the interface/device. Further contributions include methods for cost-efficient user experience testing: a modified critical incident technique and using an optical character recognition to gauge legibility user experience at early design iterations.

Table of contents

List of figures

List of tables

Acknowledgements

Thank you,

Professor Pam Briggs (nominated also for a Giant Gold Star)
Mr Paul Agnew
Dr Lib Orme (also a Giant Gold Star nominee)
Dr Lisa Thomas, Dr Amy Fielden, Dr Angie Johnson & the rest of the PaCT gang
Professor Agnes Kukulska-Hulme
Dr Lynne Coventry
Professor Dr Heinz von Kiosk
Äiti ja Isä

And for enabling early work on the thesis

Nokia
Oulun Yliopiston Säätiö.
Pohjois-Pohjanmaan Kulttuurirahasto

Author’s declaration:

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the School Ethics Committee for each applicable study.

Name: Piiastiina Tikka

Signature:

CHAPTER 1

Where the research questions are presented and challenges regarding reading on small screens are introduced.

Introduction

Vast amounts of electronic content are being accessed on mobile telephones today: the internet, e-books, e-mails, social networking and numerous other specialized applications on smartphones bridge the gap between desktop and mobile computing for everyday consumers at an increasing rate. Larger and larger consumer groups are being brought within reach of the mobile internet by manufacturers who are developing and producing ever cheaper mobile devices for the emerging markets where, realistically, a mobile device is the most feasible access point to internet content, owing to poor physical network infrastructure. To cater for this market, various mobile phone retailers' (and manufacturers') device portfolios in 2012 and 2013 show that affordable smartphones do not come with 4" top-of-the-range screens but are closer to the 3" size and have much lower resolutions than the so-called high-end phones. The high volume of such small-screen, low-resolution devices as an important access point to all manner of electronic content gives the topic of the present thesis specific relevance.

Mobile devices with small screens (sometimes referred to as small displays) constitute an umbrella term used in the present thesis to refer to consumer electronics that have a display in the region of 4" or smaller, and which are typically carried around to work, school, hobbies, etc. These devices would also come with some degree of computer and/or communications technology. Most typical of such devices would be mobile phones, but in using the above-

mentioned nomenclature the present thesis does not intend to rule out the findings from applying to devices such as personal digital assistants (PDAs) or other small electronic readers, dictionaries, etc.

For convenience, the research in the present thesis was done mostly with reference to mobile phones. In addition to referring simply to mobile phones or mobile devices, the terms ‘feature phone’ and ‘smartphone’ are also used in some contexts. A feature phone is a low-end (less expensive) mobile phone that offers a number of mobile services and functionalities that are typically associated with the more expensive, high-end smartphones. A feature phone may come with a colour display, web browser, e-mail, music player, camera and/or touch screen, but typically the feature set is limited. A smartphone is a high-end product that can be expected to have all of the de facto features and applications, as well as more powerful processing power and larger quantities of memory and storage than a feature phone. The test participants in the present thesis research were typically high-end phone owners or consumers who were likely in the course of time to move on to high-end phones. In any event, as high-end phone features tend to filter down to lower-end models with time, responses from the participant groups used remain valid and applicable to present day feature phone development work.

Analysts such as Mary Meeker of Morgan Stanley (gigaom.com, 2010) predicted in 2010 that mobile internet use (including social networking) will overtake fixed internet by 2015. At the same time companies such as Samsung and HTC are aggressively pushing internet-capable devices at emerging and developing markets in countries such as India and China at prices that are affordable even to those well below middle-class income level. Nokia talks of “The next billion” as a business strategy (Nokia.com, 2011), aiming to introduce mobile internet to a billion new users through their devices. It is not uncommon to see commuters in cities like Beijing or Tokyo using their mobile phones to read e-books or other text content on their way to work and back. E-book applications and stores such as Amazon’s Kindle are presently widely available to smartphones.

Mobile phones, by nature, are used on-the-go. Some activities require more sustained attention than others; for example phone calls can be made while walking on a street, but viewing a website tends to be more difficult on the move and is more likely to be done in a stationary position. An on-the-go context also requires a varying level of attention, as mobile users interact with their device while they navigate through crowds or try and make sure they catch the right bus home. The interference caused by context of use has been recognized by a number of countries, where handling a mobile telephone has been banned while operating a motor vehicle. Oulasvirta et al. (2005) found that attention given to mobile device tasks dropped from over 16 seconds under laboratory conditions to about four seconds in mobile, on-the-go, conditions. It has also been established that such activities as talking on a mobile or text messaging while walking slow us down and can even make it difficult to walk in a straight line (Lamberg, and Muratori, 2011).

Yet it is not only the context of use that places challenges on the interaction, whether with the mobile device or with the environment: the same types of content that have typically been accessed on desktop computers and laptops, or in the case of books, on e-readers, are increasingly being accessed on smaller mobile devices ranging from screen sizes below 3” to displays that are about 4.5” in size. Such development can be seen for example in the number of photography, music, video, mobile movie and television services and mobile gaming applications that are presently available for smartphones and many feature phones. How does the restricted display real estate affect information accessibility, by which is meant the availability of information to the end user and its comprehensibility as regards the way it is presented, and how is information accessibility affected particularly as regards text content? (In the present thesis information accessibility does not refer to disabled accessibility as often defined in usability engineering.)

The present thesis aims at identifying features and factors in the performance and perceived experience of reading text on a small screen, namely, a mobile phone. Among all the content that can now be accessed on a mobile phone a

considerable amount is text: news, entertainment, humour, books, blogs and just about any website that is aiming at disseminating information. The purpose of the thesis is to gain further understanding and knowledge of how the limited screen size of mobile devices affects reading performance and perceived experience of fluent and comfortable reading. In particular, the present research is interested in comparing performance with experience: does good performance also mean satisfaction in terms of experienced reading fluency and comfort, or do the two differ? Does reading format (mobile device vs. paper) affect reading performance? A further aim of the research is, then, to make use of the knowledge gained about perceived experience in generating practical quality evaluation support and tools for the content and mobile device industry for design iteration purposes.

Much of the existing research on reading tends to focus on the cognitive processing involved in reading and on legibility and the visual aspects of reading. An additional perspective into reading that requires some attention is the experience an individual has of the said process. From some of the literature discussed in the following chapters we can see that experience and performance in reading do not necessarily go hand in hand, and that the subjective experience of the variables that constitute the activity of reading can have a strong effect on how individuals perceive the said activity: is reading easy and fluent; is it comfortable and an enjoyable process, or is it something quite reverse? The following chapters also refer to research that has already identified something as subjective as motivation as one key factor in reading performance. The questions that need to be asked, then, are: what other subjective factors might influence the reading process, and how? Do performance and experience go hand in hand when reading text on small displays?

Electronic content on small screens potentially poses an additional dimension for the activity of reading both in terms of performance and in the experience of that process. The research questions in the present thesis approach the small screen interfaces, and reading text on them, particularly from the user experience design angle. ‘User experience’ is used throughout the present

thesis to cover affective aspects of human-computer interaction, including but not limited to usability and perceived experiences of effectiveness and utility. While research so far offers guidance for improving overall readability on the more conventional formats of print and computer displays (as discussed in the following chapters), it would also seem necessary to examine the reading process as user experience when the same content is displayed on a small screen like a mobile phone. Pushing for technological improvements such as increasing display size and resolution or creating content that can be scaled down to almost any display size present one approach to improving the reading experience, but they may not be enough: user experience design will have to address the design of the content itself as well as the design of the interface.

Device manufacturers are aware of difficulties in information visualization with small display sizes and the said technological solutions offered at the time of writing the present thesis come in two guises: sharper screens that allow ever more intricate rendering of content (the so-called retina displays lead the way with all three major manufacturers, Apple, Nokia and Samsung) and the tablet devices that are perhaps best described as a compromise between a laptop and a very large mobile phone. These technological improvements together with content and interface design aim at improving the overall user experience. The quality of that user experience is typically measured from a variety of angles ensuring both basic functionality and efficiency of a system and gaining insight into the subjective experience of using the said system. The wide spectrum of user experience quality assurance is necessary because (as seen in both the literature and in the experiments presented in this thesis) the functional axis and the subjective experience of using a system do not necessarily correlate positively. Methods that produce objective measures of interface quality, such as task-based usability testing or reading comprehension tests, provide an insight into the performance and functionality level of a system. To complement this level, collecting subjective responses to the system will provide attributes that may help explain performance: why does a seemingly usable system not feel comfortable to use, etc.?

Chapters 1 - 4 of the present thesis explore the existing literature and knowledge regarding the reading process, legibility and readability, and subjective experience of the reading process. These chapters also discuss various approaches to assess quality and user experience of interfaces. Section 2.2 illustrates some of the differences between larger and smaller devices as well as considering the special physical condition that affects reading on small displays: mobile context of use.

As important as it is to understand the performance aspects of what happens with small screen interaction and information acquisition in terms of measurable factors, it is also important to gain further understanding of the perceived experience that goes with that performance: there is no particular reason to assume that good performance as such equals good experience. User satisfaction is an integral element of interface ergonomics, as can be seen from the definition of user satisfaction defined in the ISO standard “The comfort and acceptability of the work system to its users and other people affected by its use” (ISO-9241-11).

The experiments and the studies in Chapters 5 - 10 of the present thesis first establish actual performance of reading on a small screen and what practicalities are involved with small screens and mobile interfaces. From this performance aspect the thesis then moves on to looking at the perceived experiences that go with small screens and reading, finally aiming at integrating these two realities of practice and perceived experience into conclusions that can be used in content and interface design to improve text-based communications on small screens.

All experiments and studies presented in the present thesis follow and adhere to the ethics guidelines at Northumbria University. The Northumbria University Ethics Committee has reviewed and approved all of the studies. In all studies either a paper copy or an electronic form of the research aims, informed consent and post-test debriefing information were made available and in one-on-one experiments participants were offered the said forms in paper format to

keep. All participants were also informed about the possibility of withdrawing their data from the experiment should they wish to do so.

CHAPTER 2

In which small displays are reviewed in their typical context of use, and the differences in content visualization between large and small screens are illustrated.

Mobile reading in the real world

This chapter aims at providing some insight into the ways mobility and the size of mobile devices can affect accessibility to content, attention and interface design of small display devices. Electronic content is rapidly becoming ubiquitous in areas with a high number of users for mobile and internet services. Through convergent networks (networks that enable rich media content distribution of data, video and telephony) the web with all its contents is available nearly everywhere and at any time, and all communications methods and web services are used on more than a single device. It is possible for people to own desktop computers and/or laptops, smartphones with affordable data rates, and tablet computers. There are public terminals available for various services in public offices; internet banking is now offered via apps on the smartphone and our own content, be it photos or documents for work or study, is increasingly held in data clouds so that the said content is available to its owner on any fixed or portable device. Further, such content can be shared through clouds and services with other users practically in real time.

With all the versatility of how and when we access content it has become next to impossible to know what type of device will be used for accessing particular communications and content. The ways we communicate with our friends and colleagues are as versatile as are the ways of accessing the internet: social networking is all available on the smartphone in just the same way as it is on the laptop or the tablet computer; the chat you had going on your laptop can be continued seamlessly on the mobile phone; all the contacts have been consolidated throughout your devices and services so that from a technological point of view all your connections, networks and content really can be within grasp at all times.

When sharing any content, be it a link to a website, a photo or a status update on a social network, it is no longer necessary to consider what means the recipient has for accessing the content. Mobile device users who were interviewed about their mobile reading attitudes and expectations (Chapter 9) typically assumed at the time that they would delay until they had access to a "proper computer" before accessing some complex content. Many respondents considered mobile phone screens to be too small and the interfaces too unwieldy when using the mobile for large amounts of content. Interview participants who were already in possession of more advanced touch screen phones were not so apprehensive. Such reluctance to adapt to the mobile internet was also identified for example by Kaasinen & al. in their 2009 paper discussing the various factors which make end-users perceive the mobile web as not offering a good enough user experience (Kaasinen & al. 2009). Since the time of the interviews both the processing power and screen sizes of typical mobile phones has changed dramatically: following a web link from a text message or from someone's status update on Facebook can be as simple as tapping the screen, and the content is there.

Use of the mobile web has even prompted websites that are aimed at making finding information and content easier particularly on the mobile phone. Skweezer.com converts a large and complex website into a simpler form if you happen to have a phone that does not easily support full web experience or a

website does not happen to offer a mobile version. There are various sites designed to offer reference type information about restaurants or movies and other events that has been formatted to support small display (mobile device) viewing. There are even websites like webonyourcell.com that list various other sites that the service considers handy to have when you are “on the go” (news, yellow pages, sports, Froogle, etc.). Such trending and useful websites are listed on sites such as About.com’s Web Trends: Top Mobile Web Sites You Have Bookmarked (About.com’s Web Trends, 2013).

Section 2.1, below, examines the challenges of information accessibility on devices with small displays from a broader perspective of the mobile internet. As noted already, practically all electronic content in everyday life is also available on the mobile, including text-heavy content such as (print) news and sport and other journalistic content, blogs, and any websites with information content. Section 2.2 illustrates some of the differences in content design and rendering between two different-sized electronic displays.

2.1 Mobile devices and text in context of use

One area of research on small screens and mobile devices that sheds light on the issues with text on small screens is the mobile internet: many of the challenges with web content on small screens appear to be similar to those with continuous text on small screens. (Kaasinen & al. 2009; Kaikkonen, 2009; Oulasvirta, et al., 2005; Qiu, et al., 2004; Shrestha, 2007) . Essentially, the problem has to do with fitting vast amounts of content on a small screen with the added mobile context of use. In the present thesis, ‘mobile context of use’ refers to accessing content and using a device while on the move or in between other tasks and events in everyday life, as opposed to more static context of use with a desktop computer use. Obviously, even with paper and desktop size screens content will be larger than its viewing area, but the particularly small interface of a mobile phone or some other device like a mobile phone seems to accentuate the issue. Also, websites are likely to carry

their information in text format and thus the problems of reading on a small screen go hand in hand with mobile internet usability.

A typical mobile web issue has involved navigating a website that is both higher and wider than the viewing area on a mobile phone: a website requires scrolling and panning (moving a canvas of content both vertically and horizontally) in order for all of the content to be seen one piece at a time. At present, more and more websites have opted to solve the problem(s) by recognizing the devices that access the website and, where necessary and feasible, offering a mobile-specific version of the site. In these cases (the URL string typically comes up as “http://m.websitename…”) the content on the page is organized into a single-column layout that fits onto the phone screen without the user needing to do any sideways panning.

Usability studies on mobile web use, for example Shrestha (2007), found that a major problem with browsing the internet on the mobile involved locating information, even when it was presented in a single column. The amount of scrolling that was required made the experience poor and frustrating. Another key element in viewing content on a mobile screen that has come up with mobile internet use has been the need to optimize presentation: size of text and images (zoom level), layout adjustments and even semantic conversion (re-structuring content information order and grouping to help small screen use) (Qiu, et al., 2004). The optimization and navigation needs can well be generalized to other forms of content, not only websites. Hyvärinen, et al. (2005) concluded in their study of link placement in mobile banking applications that efficiency of action can take a backseat to experience of meaningfulness: in their study a more efficient design was seen as more complicated in terms of content navigation. Krug (2006) crystallizes this dilemma in the title of his usability text book: *Don’t Make Me Think!* All these concerns also apply to reading on a small screen. Reading text and finding the information there that you need requires navigation and interaction with the device: it requires scrolling, panning and page changes, depending on what

type of device or application is used, and in order to be readable the contents require optimized presentation.

The close relationship between web design and readability is also seen in such papers as Zibell‘s (2000) commentary on Klare’s writing principles, which suggests that most readability principles also apply to web design. Essentially, Klare’s principles of style, type and pleasantness are compared by Zibell to the three core parts of web design: architecture (how content is structured and organized), interface (the platform that presents the architecture to the user) and interaction (the way the interface elements guide the user to take action with the interface). Klare’s definitions of ‘readability’ in terms of legibility, ease of reading and comprehension match, in Zibell’s view, the goals of web readability: architecture ensures that the user sees everything that is intended to be seen; interface ensures ease of use (reading); the user’s interaction with the interface is correct if the interface has been understood correctly. Further points of contact are found in the principles of producing readable text (Klare, 1963, in Zibell, 2000) and usable websites (Nielsen, 2000, in Zibell, 2000): they must both be created for the user/reader; if these match the reader’s/user’s ‘desire’, the experience is positive and the usability/readability score will be higher.

2.2 Same content, different screens

The present section illustrates the common differences between viewing text-heavy content on a larger format (laptop) screen and a smartphone screen. The aim is simply to draw attention to the physical differences in size and to how the size differences affect visual design of content. The devices used for the examples were a 13” laptop and a 3,5” mobile device. The images have not been produced in actual size or in proportion to each other, but they aim primarily at illustrating the differences in content handling between the two different device types.

In the first example the basic google.com site is shown as it is seen on a laptop screen and on a mobile device (Figure 1). In the example, the contents and layout of the Google main search page can be considered reasonably simple in terms of amount of content and the layout, but all the same producing the site on a mobile device screen has nevertheless required some alterations: instead of the fun Google title art (Doodles) the more formal search engine logo has been given along with just a selection of the links that the full laptop browser version provides. Different mobile browsers may render the graphics differently (for example, the links in the top bar such as Images, Places, etc. may perhaps be rendered in more advanced ways) but essentially the example shows how the small display interface concerns are present in something as simple as the Google front page: visibility of content and enabling interaction, namely, selecting links.

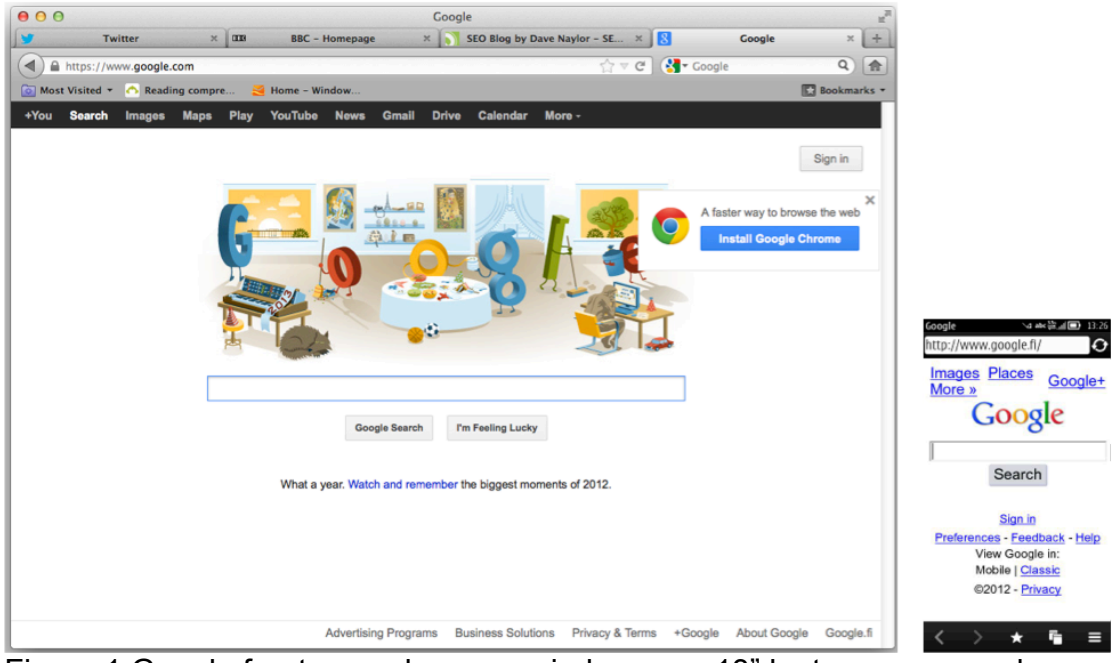


Figure 1 Google front page: browser window on a 13” laptop screen and on a mobile device. Screenshots are not in actual proportion. Date of screen captures: 1.1.2013 (laptop) and 20.12.2012 (mobile).

The second example is from the BBC.uk website, again as seen on a laptop and on mobile phone browser (Figure 2). The front page of the BBC website looks distinctly different on a mobile phone from the laptop browser version: the mobile version design aims at efficiently producing main headlines in

easily selectable elements and with fewer images and less “air” (empty space) around the various elements that make up the page. In the desktop version news items stylistically follow more of a newspaper (print and website) design by anchoring headlines to photographs.

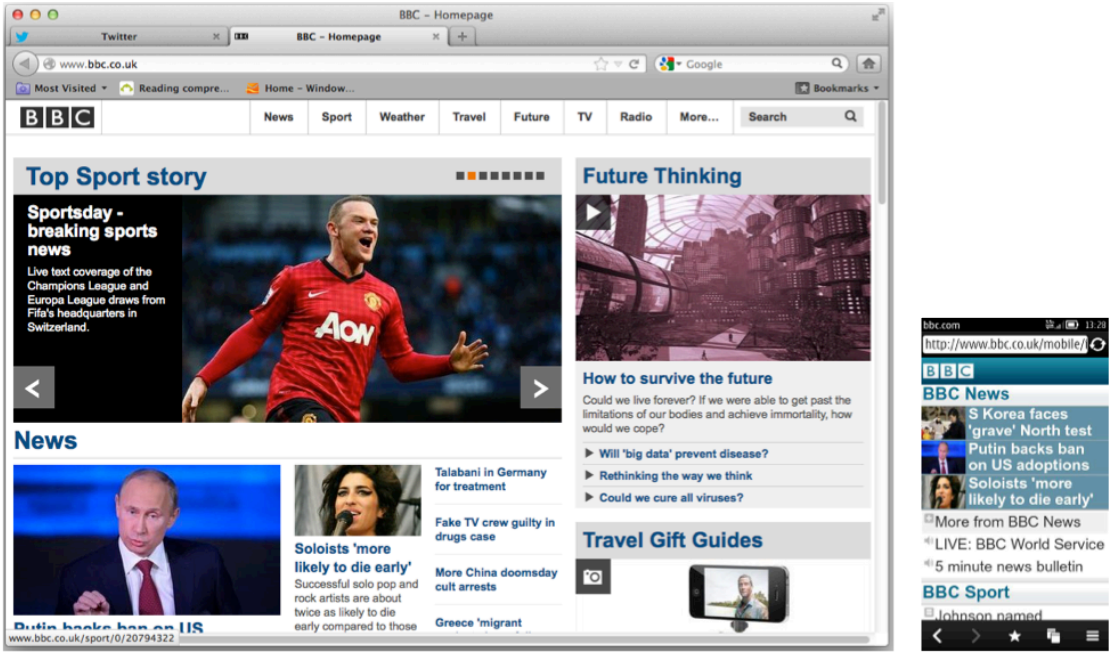


Figure 2 A news item on BBC website on a 13” laptop screen and on a mobile device. Screenshots are not in actual proportion. Date of screen captures: 20.12.2012.

A closer look at a specific news article on the BBC website further illustrates the very different demands the reading medium places on the content. Figure 3 shows the same news item on a mobile screen and on a laptop (same sizes and orientations as in the previous examples). In both cases the news article is, naturally, prioritized, but in the case of the mobile it has been necessary to strip away all extra items such as related news and quick links to top stories. As with the front page example, these extras as well as the section links at the top of the desktop version (News, Sport, Weather, etc.) are elements that enable a flexible and explorative way of browsing the BBC site as a whole. The design encourages more multi-dimensional browsing where the user can find the same articles and items from multiple directions: basically, everything is connected to everything else. The visually separate areas for various

categories of related material support the user's effort in creating a mental map of how information is structured on the site and where he or she needs to click in order to find some specific item.

On the mobile version, however, the supported browsing style is more linear in nature: from a plain list the user selects an item of interest, looks at it and then in order to view more items it will be necessary either to go back a step to the previous list of items or (if the site has been designed with this in mind) scroll to the bottom of the present item to find further options.

Figure 3 BBC news item on a 13" laptop screen and on a mobile device. Screenshot sizes are not in proportion. Date of screen captures: 20.12.2012.

The same reduced navigation support is visible in the following example from Telegraph.co.uk (Figure 4) where again the elements that support multi-dimensional, explorative navigation have been all but removed from the mobile version of the website.

Figure 4 The Telegraph website front page: browser window on a 13” laptop and on a mobile device. Screenshot size not in proportion. Date of screen captures: 20.12.2012.

It is not entirely necessary, however, to hide navigational aids behind menus or to drop them entirely. A blog in the following example (Figure 5) aims at catering for a mobile device audience as well by providing the section/topic links at the top of the desktop version (Home, Who we are, etc.) as layout scalable buttons at the top of the mobile version of the page: the width / size of the screen may well alter from device from device, and the link buttons would simply be wrapped to fit the available horizontal space.

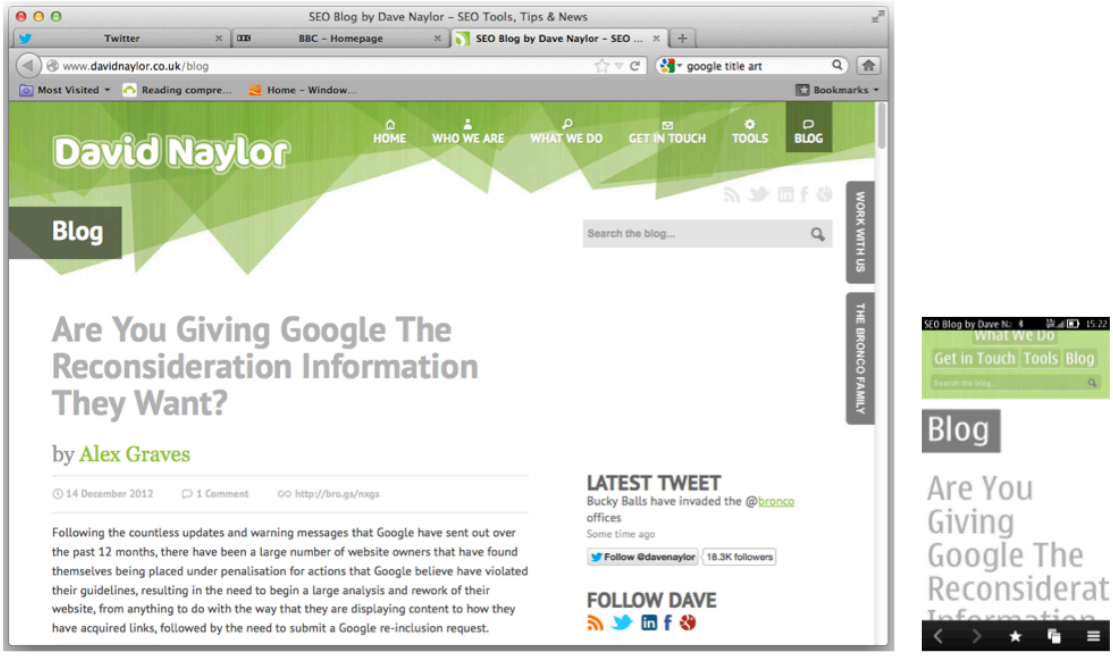


Figure 5 A technology blog: browser window on a 13” laptop and on a mobile device. Screenshot sizes not in proportion. Date of screen captures: 1.1.2013.

Some websites are predominantly aimed at mobile devices, as can be seen in the case of Twitter (Figure 6): The layout of the desktop page follows the mobile-optimised layout where all content is given in a single column. Some of the functions (Home, @connect, etc.) have been spread out to the width of the desktop browser window, but the said bar does not offer any significant additional functionality compared to the mobile screen.

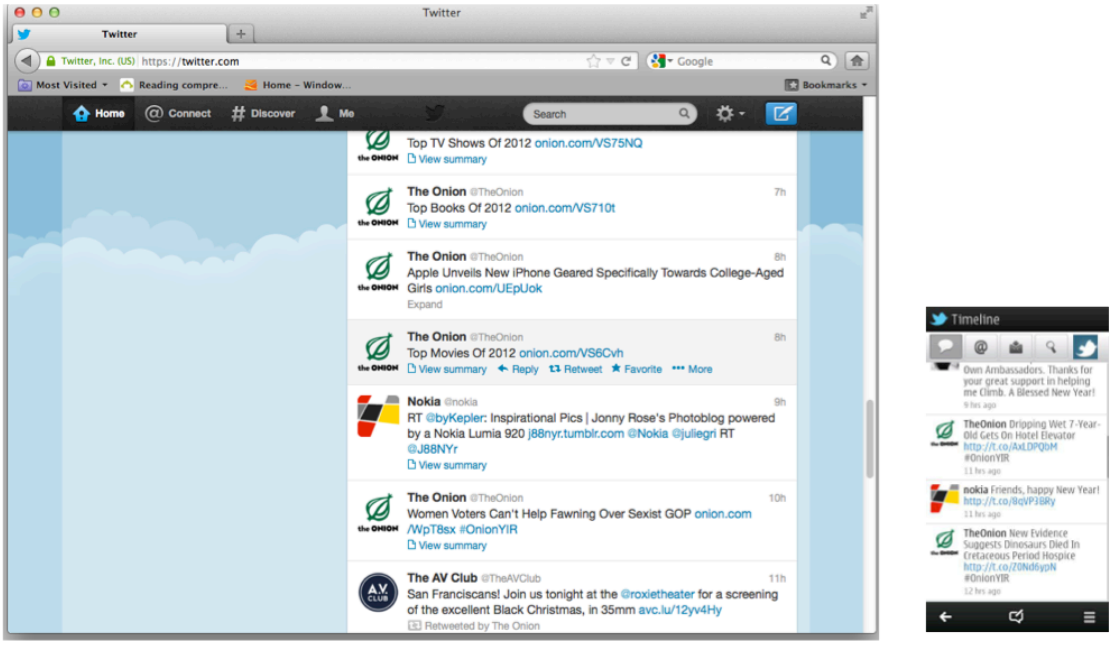


Figure 6 Twitter: browser window on a 13” laptop and on a mobile device. Screenshot sizes not in proportion. Date of screen captures: 20 Dec 2012.

As can be seen from these examples, although mobile phones can render a good amount of text and content on the screen, it has still been deemed necessary to show the content slightly differently from the desktop style. The differences observed in the present section simply go to illustrate that display size has an obvious significance in information and interface design.

In some of the examples (Figures 2, 3 and 4) the content as rendered on the mobile device screen is more pruned and limited compared to how the sites are rendered on the computer screen. This difference evokes concerns over the amount of work required from the end-user to integrate and then comprehend the content that has been fragmented in this way (Sweller, 1994; Sweller and Chandler in Oostendorp, 2003). One of the research aims set for the present thesis was to explore and learn how the reduced screen size affects performance and the user experience of reading on the mobile device screen.

From the website examples it is immediately obvious how much more interaction is required on the mobile device before the end-user has seen all

the same content as is visible almost at a glance on the computer screen. The mobile device is capable of showing text content in a neat and clear single portrait orientation column, much like a column in a magazine. This solution certainly seems functional and tidy and although the end-user cannot see as much of the contents on the mobile as one can on a paper column (or on the computer screen), is there a difference in the performance of reading a column on a mobile screen and other formats? Dillon, et al. (1990) identified that reading comprehension was compromised when a sentence straddled over page division on PC screens.

One difference highlighted earlier on when looking at website examples on computer and mobile screens was in the navigation model. Navigation through a whole site made up of various pages with set links to other pages within that site is more linear and hierarchical on the mobile screen than on the computer. The large format version of a page or site allows the end-user to form an overview not only of the specific content (text) on the screen but also of how that content is organized overall on the website. Research into mobile web usability (for example Shrestha, 2007) suggests that lack of overview can be a concern for end-users in the sense that they would be uncertain about whether they would be able to find specific information on a given site.

This chapter aimed at placing small displays into their typical context of use and at offering an overview of interface and content design issues that typically come up with providing content for both large and small displays. These issues, discussed and studied in the present thesis, comprise: the overall lack of visibility into content, loss of immediate (textual) context, fragmentation of content and increased need for interaction with the device. The following chapter looks into visual and cognitive aspects of the reading process.

CHAPTER 3

Which discusses existing knowledge regarding visual and cognitive aspects of the reading process.

Mechanics of reading

The present chapter discusses some of the existing knowledge and literature regarding how we see and recognize text and words. An overview of these processes is beneficial to understanding the potential special demands small display devices may present to the reading process overall. Understanding the reading process is also necessary in order to answer the research questions presented earlier.

There is a great body of knowledge on how the reading process works physiologically and psychologically, and we can see that when the process is disrupted in any way it is reasonable to expect poorer reading performance results with measures such as comprehension or reading speed. It would, on this basis, also seem reasonable to expect the experience of reading to be affected when the reading process is disrupted.

In the present chapter, section 3.1 discusses some aspects of the physiology of reading: how the eye moves and focuses on text and words when reading. Section 3.2 is interested in the cognitive processes involved in reading, particularly as regards the context of small displays: limited screen space may lead to fragmented content and thus an additional need for information integration, and the mobile context of use may put an additional strain on attentional resources. Section 3.3 discusses the reading process in the

broader terms of personal interest and motivation and an individual reader's ability and experience.

3.1 Word recognition and legibility

At the level of word recognition and legibility, the way the eye functions in reading is an important element in fluency. Each eye fixation on a line of text has an area that is the central fixation point and is the area of sharp focus. Around the sharp focus there is a stretch of less sharp but still visible and partly recognizable area. Beyond this, there is also the periphery. The parafoveal information produces a preview effect in reading, that is, the parafoveal information though not accurate and complete is used for making a quick hypothesis of what is to come, thus limiting the range of what the following characters or words are likely to be. (Balota, and Rayner, 1991. In Besner, and Humpreys, (eds). 1991; Rayner, et al., 2011). For readers whose first language uses a left-to-right writing system the stretch of the parafovea tends to be slightly biased towards the right.

Hyönä (1993), studying the effect of context on eye movements in reading, found evidence of the integrative nature of fixation regression on comprehension. When a word in a text was not considered predictable and was not primed earlier in the text in any way, the fixation regressed to the given word more often and the regressions would occur particularly when the reader reached a sentence boundary. Text passage difficulty also increased both the number of fixations as well as their length in time (ibid.). In small display context, the concern that arises from this finding is that should the display size and content design not allow enough context to be available at one time, regressions (and previews) are not fully possible and comprehension and fluency are thereby compromised.

Admittedly, some physical interaction from the reader is required even with conventional books and other printed material, for example when parsing sentences and paragraphs across page boundaries or when needing another

look at a word or a phrase. However, the shorter chunks of text that current mobile displays allow at a viewable size and format require far more interaction to scroll or pan the content around more frequently. According to Ellis (1983) some 10-20% of reading time is used on such regressions. One question that arises, then, is that if reading the content requires more frequent page turning, will the paging start interfering with the fixation regressions and previewing that are a necessary part of reading? Admittedly, the experiments in the present thesis do not use eye tracking to establish whether such interference takes place, but this question is an interesting one for further research into small screen reading.

Layout and other visual style elements (line spacing, number of columns, typeface size and format, etc.) also affect the experience of readability (and legibility) of text on electronic displays (Yi et al., 2011; Bernard et al., 2003). Research into most functional typefaces on electronic screens indicates that sans serif typefaces and serif typefaces specifically designed for electronic displays generally do work better for end-users (Bernard et al., 2001). Some typefaces, such as Verdana and Georgia, have been specifically designed for electronic displays (Boyarski et al., 1998). Furthermore, line length and column layout are issues that have been studied in terms of finding the most functional layout designs for text on electronic screens. Dyson (2004) found that, up to a point, a longer line (more characters per line) enables faster reading than a shorter line. However, in terms of layout, the results on finding an optimal column number have been inconclusive (Dyson, 2004; Baker, 2005). Chapter 5 discusses the difference between performance and the experience of that performance and how participants responded to the possibility of adjusting the viewing conditions (and thus presumably legibility).

3.2 Fragmentation of content and split attention: cognitive load concerns

Fragmentation of information in the context of reading continuous text on small display devices means the way text is divided into small chunks of a few words or sentences that are visible on the small display at one time. The reader cannot glance quickly either to preceding or following parts of text in order to aid comprehension and refresh the memory of what has been read just before. The working memory concerns over reading comprehension itself (Just and Carpenter, 1992; Graeser, et al., 1994) also reflect upon content fragmentation and split attention (Sweller and Chandler 1994, in Oostendorp, 2003): the small interface could well produce comprehension differences as compared with comprehension using a larger format such as paper. De Bruijn, de Mul and Oostendorp (1992) studied verbal and pictorial information presented on a computer screen and established that as the information rarely could be displayed on a “single page”, as it were, the learner not only had to integrate the information from the text and picture on each page but also maintain the said information as he or she progressed through the following pages in order to be able to integrate information on the new pages with that on previous pages (Oostendorp, 2003.).

The problem of fragmentation has been acknowledged in the field of information visualization, as illustrated for example in Spence (2007). Spence describes various strategies for dealing with the visualization of data when the display area is considerably smaller than the content (allowing for an acceptable level of legibility in presentation). One of the concerns highlighted as the source of the problem is fragmentation of information when only a small part of it is visible at any one time. There is lack of visibility into content that has been read just prior to currently visible text alongside a similar lack of visibility into text that immediately follows the currently visible text. Shrestha (2007) found that at perceived experience level, one mobile web browsing concern for mobile device users was being unable to gain a sufficient overview of a website and thus find some specific piece of information there.

To put the problem of fragmentation into the context of small, mobile device displays, there may be considerable limits regarding how much text is displayed at one time on the screen. If the propositions are too loosely packed it will be difficult to physically see the entire sentence and related propositions (required for coherence) and interacting with the interface will be necessary in order to bring either previously read text or the upcoming text into the visible area. It may not even be possible to see a full paragraph of text at one time, depending on the genre and medium. The necessity of interaction demands splitting active attention between the reading process and the interaction process, thereby adding to the cognitive load. Packing the proposition more densely, however, may lead to yet another problem, where the propositions are too dense for fluent comprehension to be possible. The trade-offs in readability on small screens thus seem to point to balancing proposition density on the one hand against legibility and syntactic complexity on the other. Working memory is an important part of language comprehension: it stores the partial and complete products of what a person is reading or listening to as the language is integrated into meaningful ideas from successive words (Just and Carpenter, 1992).

Fragmented content on an interface that is not automatically clear would, according to cognitive load theory produce extraneous cognitive load (Sweller, et al., 1998) during a reading process. Reading range (a measure of working memory capacity) also varies from individual to individual (Daneman and Carpenter, 1980), which makes it difficult to gauge acceptable degrees of content fragmentation. Individual differences not only in reading ability but also in familiarity with the device used for reading produce various levels of both extraneous and intrinsic cognitive load from reader to reader. Interrupting the reading flow not only by splitting content onto several smaller “pages”, as it were, on a mobile device screen but also by demanding a frequent device interaction in the form of turning those electronic pages could both be anticipated to add to working memory resource requirements.

The cognitive load may decrease with the reader's level of experience of the reading medium. With a conventional book, the only task additional to the basic reading process expected of the reader is basically to turn the page at the right time and find the new starting line on that page. After all, as classical modular theories on memory illustrate, working memory is a limited capacity and long-term memory is a potentially unlimited capacity. Things that have been learned before are stored as schemata in long-term memory from where they can be applied to new situations. (Sweller, et al., 1998). For most people, reading a book (possibly excluding the so-called digital-native generation and defining "book" in a way that allows for cultural and language-dependent interpretation) is an activity learned and reinforced in some form from a relatively early age.

It is likely that for most of us there is an automated schema in place for reading a page and then turning it, thereby requiring very little in the way of working memory and attention. Likewise, the more common electronic interfaces become, the more familiar consumers will be with them, creating automated schemas from at least some of the more de facto interactions such as pinch-zooming, page panning long-press and double-tap on touch screen mobile phones.

On the other hand, in a situation where **a)** a reader is faced with fragmented information (a very small amount of content from a longer text is visible at one time) and **b)** the medium is not a familiar book but an electronic device, the requirement for interaction (to make an effort to find the next piece of information) and the method of interaction (not turning a page but performing some other function to bring up more information) will both call for cognitive processes more intensive than with a conventional book. Working memory would already be under higher demands owing to the fragmented nature of the information, and at the same time the reader must keep interacting with the mechanism that allows him or her to proceed with the reading, be it scrolling with hardware buttons, dragging the page using a touch screen or whatever else a typical small display device may offer. The process of

bringing more information onto the display, then, typically calls for added attention to stop the interaction when the expected piece of text is finally on the screen: typical interaction problems include scrolling too much or too little, thus making it difficult to find the next new line of text to read. With interfaces that do not require scrolling or panning, the attention the interaction demands may lessen.

The simultaneous activities of reading and attention-heavy methods of interaction are very likely to tax the reader's cognitive resources, and the question is, how will all this affect the reading fluency and perceived fluency? In other words, having to both comprehend fragmented information and interact with an interface that is not yet a fully formed automated schema in the long-term memory causes a high cognitive load owing to a high interaction level between all of the factors involved (Oostendorp, 2003). Bearing in mind, however, that the interaction with a conventional book is something that has developed over time and is something that the reader has learned, a different interface should in turn be learnable at least to some degree and thus the cognitive load caused by the device interaction should lessen as the reader gains experience of the device and its interface.

3.3 Reading strategies, reading fluency and comprehension

The various factors affecting reading fluency and comprehension indicate a number of areas where reading process fluency might be compromised when reading takes place on a device with a small display. The present section draws together some theories on reading processes, perception and cognitive processing, discussing them in the context of small display devices with the aim of providing some conceptual frame for the thesis.

Samuels and Eisenberg (1981) discuss the factors of text format, legibility and readability in the context of creating models of the reading process, focusing on why the various factors have an influence on the reading process. The external factors include the physical characteristics of written materials,

meaning legibility (visibility and physical recognition of text including text size, contrast, available light, print size, orthography, etc.) and format of text (layout-related elements such as column width, size of margins, size of page, etc.). Along with the physical characteristics there are the language style factors that typically fall under the heading of readability (functionality, complexity and communicativeness of written language, often measured through word frequency, sentence construction and density of proposition). (Samuels and Eisenberg, 1981). The reading process factors identified by Samuels and Eisenberg are observable also in Chapter 9 (an interview study on attitudes and expectations on small screen reading) and Chapter 10, where similar factors were extracted from interview data on perceived aspects of small display reading.

Physical text factors affect reading strategies. Afflerbach et al. (2008) define a reading strategy as a conscious decision to direct the reading process as a means to an end. By recognising a style and layout format readers will adopt a specific strategy that supports their favoured progress through the given text (Samuels and Eisenberg 1982). For example, a familiar narrow column width on a newspaper page can make the reader adopt a quick glancing strategy to skim through a whole page in order to identify articles and headings that are interesting. The reading process in such a case takes place within a familiar contextual field. The selected reading strategy can also depend on the purpose of reading (motivation) (Afflerbach et al., 2008): for example, reading for the purpose of deep learning or a need to simply read through a piece of text as fast as possible are strategic decisions by a reader that depend on the purpose of reading.

Text format and device factors provide readers with cues for adopting a reading strategy. How well or how frequently such strategies are used is dependent on the reader's individual reading ability. Afflerbach et al. (2008) suggest that reading skill, as opposed to a reading strategy, is an automated process that takes place without conscious decisions on the reader's part. Samuels and Eisenberg (1981) appear to advocate a view that does not rule

out the conscious/automated distinction between strategy and skill but leaves room for assuming that strategies, too, can be learned to a point of automation and that employing a strategy does not need to be a fully conscious decision but can be an internal expectation of what a given text will be like based on the recognition of format or genre. In this sense, what Samuels and Eisenberg call 'reading strategy' is something not entirely dependent on intention or motivation, but also something that follows stylistic recognition.

Readability elements can follow conventions that are familiar to the reader: for example, scientific articles on a given subject matter are likely to follow a set format with regard to vocabulary and linguistic construction, while a crime novel will follow a number of genre conventions that the reader may be familiar with. Any breach of the conventions typically causes a breakdown in the reading strategy, which in turn negatively affects comprehension and overall reading fluency. Such stylistic disruption is a well-used technique in creative writing such as poetry, and mixing genres within a text is a way of applying surprise and additional perspective to the interpretation of a text: for example, *The Autobiography of Alice B. Toklas* from 1933 by Gertrude Stein produced one such surprise when the style of a biography was used to convey a story of an entirely fictional character. The reader's experience of reading different text types is part of the existing knowledge base and thus part of the comprehension process. The look and feel of text presentation cannot, always, be reproduced in exactly the same way on a mobile display as they are in their "original" reference format. Can readers transfer strategies and stylistic expectations from a format that they know onto a mobile device even if the visual cues for presentation conventions no longer apply?

Text comprehension requires the presence of contextual information in order for the reader to be able to employ a reading strategy suitable for a given syntactic style. For example, a long and syntactically complex sentence may in certain cases be more effective at communicating the intended meaning

than splitting the sentence into two simpler sentences that at first glance might appear easier to comprehend. Straying from the style convention abruptly, even to a simple structure, may result in poorer comprehension. (Samuels and Eisenberg, 1981). However, other literature on syntactic complexity points to generally acknowledging that complexity makes language processing (comprehension) poorer because of the necessary extra cost to working memory (Just and Carpenter, 1992; Gibson, 1998, Graeser, et al., 1994).

Gibson (1998) highlights two specific components in this resource cost in his Syntactic Prediction Locality Theory (SPLT): first, there is the memory component (used to determine how much of the computational resource is needed to store partial sentences while waiting to be able to integrate further input) and secondly, there is the integration cost (how much of the computational resource is needed to integrate new elements into the partial sentences). Gibson's theory emphasises locality; the longer the distance between the items that require integration, the higher the resource cost. The importance of distance has also been emphasized by Hudson (1995): longer distance integration requires more resources. Dillon, et al. (1990) discovered a negative effect on reading comprehension when reading sentences that had been split across screens: subjects had to return to a previous screen to re-read the earlier part of a sentence more often than was necessary when sentences did not split across screens. Just and Carpenter (1992) also make room for acknowledging individual differences in working memory capacity for language, suggesting that the individual differences account for differences in adult readers at various elements of language comprehension.

Related to syntactic complexity is the idea of proposition density, where the rate at which the propositions (or "idea units") are presented in the text has a bearing on reading comprehension and fluency. When a high number of important and relevant ideas are presented very tightly together, the cognitive system is taxed more than in a more loosely packed text. (Samuels and Eisenberg, 1981). On a mobile device screen, the compromises in content

design may be great: a naturally fluent pace of proposition density might spread content over several pages (or screens) and require a lot of interaction (page turning), whereas high proposition density might decrease the overall sense of fluency and comprehension.

The small displays of mobile devices seem to be in a disadvantaged position compared to conventional text media in terms of text format and readability. The limitation of the screen size seems to compromise either visibility or available context. Less context and poorer visibility of style elements may prevent the reader from using reading strategies efficiently. While Samuels and Eisenberg suggest that there are times when syntactic complexity can be a more efficient way of conveying meaning, put in the context of many small displays that do not show an equivalent number of words per screen compared to a book page or a desktop computer screen, the implication would be that added syntactic complexity paired with increased interaction (page turning, scrolling, etc.) would be a combination that heavily taxes the reader's cognitive resources.

To assist readability, a text can employ various methods in overall text construction in order to prompt and help a user employ the pre-existing knowledge base. A fairly basic requirement for the reader to be able to use his or her pre-existing knowledge is coherence from one proposition to another. Further methods for providing order and coherence in a text include using organizing markers such as titles, headings and, for example, outlining an upcoming section in an introduction (advance organiser). (Samuels and Eisenberg, 1981). At the very basic level these elements also include, for example, grouping units of communication into meaningful paragraphs. Coherence, where one proposition has a meaningful connection with the next, is basically a form of priming, where recognition of one item alerts the reader to more efficiently recognize any upcoming items that are related to the first one.

The above-mentioned methods for providing organizing markers still apply in mobile reading. The limitation of the screen real estate will restrict how many of these markers are visible to the reader at any one time, resulting in a lack of overview and the reader's inability to anticipate other content yet to be seen. Being unable to form an overview would then limit employing reading strategies from early on in the reading.

There are great differences in how much text and other content is visible at a comfortably legible level at any one time between various text interfaces. The figures in section 2.2 illustrate the difference between screen sizes for essentially the same content: Figure 4 shows a newspaper website as seen on a laptop screen and the same site as captured off a mobile device screen. With such differences even with what is essentially the same content, it does not seem unreasonable to question whether readers, when looking at text content on a mobile device, will be able to employ the familiar reading strategies they have acquired and learned earlier. Further to that, it may be necessary to consider the possibility that a pre-existing reading strategy may even disrupt comprehension when reading text on a small display: after all, the changes to the overall text formatting and lack of available context may disrupt the application of existing knowledge and the cues and prompts of advance organizers the reading strategy may rely on. If this is indeed the case, there should be a clear reduction in comprehension and speed (fluency) between a conventional format text and a mobile device version of the same text.

CHAPTER 4

Where literature and research methodology for reading performance and user experience are overviewed.

Performance and experience

In the previous chapters the focus has been on the mobile context of use and the special demands small displays place on interface design. In the present chapter the focus is on existing knowledge of how small displays affect the performance of reading, and on how the performance can be measured. Alongside performance, the present chapter discusses measures of experience and how user experience relates to performance. Earlier studies have indicated that reading performance and perceived user experience do not always coincide. For example, Laarni (2002) found that while he was able to make recommendations on optimal display types for various presentation methods based on participant performance data, the perceived ratings for usability and preference differed from those for performance. In other studies reading and comprehension performance was measured when reading on electronic displays, varying the visual elements such as text size, typeface and format (Bernard et al., 2003) or column width and line spacing (Yi et al., 2011), and performance data again differed from perceived feedback: even when the performance measurements did not show any significant difference between conditions, test participants were able to differentiate at the level of experience and express a preference.

The present thesis is interested in the differences between reading performance and the subjective experience of reading on small display

devices. Interviews with mobile phone users (Chapter 9) showed that the expectation of reading on the mobile phone was that it would be more difficult or irksome than reading the same text in a conventional format (paper). Cognition is complemented by emotion and emotion that is embedded in the perception of a system, product or a task has a great effect on the overall experience of that product, system or task: format and presentation seem to influence user experience beyond simply being usable or practical (Norman, 2004). The areas of measurement in the present thesis are, then, two-fold: first, the need to establish if reading on a small screen does produce poorer results than reading in a conventional format and, secondly, the extent to which the experience of reading compares to the reading performance.

The present chapter looks into various methods of evaluating and measuring reading performance and readability as well as into methods for evaluating and testing user experience (UX), discussing these methods in the context of the present thesis and its research questions.

4.1 Readability and reading performance

The study of readability is a specific field devoted to evaluation of the functionality, complexity and communicativeness of written language. At its broadest, readability study takes into account all those aspects that influence reading and comprehension. There are various definitions of readability, summarized in DuBay (2004): Klare, Dale and Chall, McLaughlin, and Hargis have all defined readability with a slightly different emphasis on such readability factors as clarity of writing and how easy it is to comprehend the text, not excluding the characteristics of the readers of the said text (DuBay, 2004).

There are a number of tools devised in readability studies to measure and evaluate readability of texts. Gilliland (1972) lists formulae, tables and charts, sentence completion and Cloze procedure, objective question and answer techniques and subjective assessment. Readability assessment of the three

main elements of readability (as stated in the definitions) involve

1) Ease of reading
2) Interest
3) Ease of understanding.
(Gilliland, 1972).

The first item on the list, ease of reading, concerns basic reading skills (word recognition, error rates, eye fixations, etc.) and physical aspects of visibility and legibility. The second is – understandably – studied from a human interest point of view. Studies of the third group have focused on words and sentences (their length, frequency, etc.). The results derived from studying these three areas of definition separately are not, according to Gilliland, comparable (ibid.).

Readability formulae can be useful after a user and task analysis phase has been completed and there is a clear understanding of what is being measured and with whom. Such reader-oriented aspects as interest and motivation cannot be left out of the equation, and user and task analysis may provide some data that help to negotiate the interest and motivation variance among readers. For example, instructional text is used with the aim of completing some task. The readers therefore share a common motivation. A history book for schools, however, will have a readership where some readers already have an interest in the subject matter, while other readers have no personal motivation to read and study regardless of possibly having to pass exams on the subject later. First impressions, then, are that a readability formula may be of some use in the first example, provided that the readership is otherwise homogenous enough, but that there may be difficulties with the second example group. Redish (2000) argues in her criticism of Klare's *The Measurement of Readability* (1963) that instead of trusting a readability formula, better and more accurate readability assessment results would be achieved through usability testing documentation with a sample of the target audience.

The idea that exact mathematical (and statistical) methods could be imposed

on a sample of written text is, probably, quite appealing if someone is looking for an engineering solution for their language evaluation needs. However, as the adverse comments by Bruce & Rubin and Anderson & Davison (for example) in Davison, & Green, (eds.) 1988 exemplify, these formulas at best give very little insight into the true level of language complexity and into how readable or unreadable a piece of writing is. According to these articles, readability formulas ignore some of the very central human factors in how people read, comprehend and perceive written text. For example, replacing the influence of a person's prior knowledge of a particular topic he or she is reading about with a vocabulary difficulty rating (Davison & Green, 1988) does not cover the full range of that reader's life experience, genre literacy or any personal competence in reading. There are no means of grading a reader's level of interest, prior knowledge, motivation (whether natural/personal or conscious direction of focus), which means that in effect formulas will overlook all individual aspects of comprehension and reading ability. The result, therefore, is that we are left with formulas that are practically incapable of producing usable data for evaluation of human communication.

Readability formulas tend to focus on items such as word difficulty levels and syntax complexity as charted from a small sample of text. Owing to the limitations on what can be deduced from a text sample for statistical analysis, results of this kind of analysis should not be used for revising texts (Davison & Green, 1988). That being the case, an immediate question arises: if a method that attempts to reveal unnecessary complexity and hidden difficulty in texts cannot be used in trying to improve those texts, precisely what purpose do the methods serve? As Bruce & Rubin (in Davison & Green, 1988) found, simplification of a given text reduced its cohesion and overall communicativeness. At times intricate issues may require intricate communication, which, in turn, may make a given text look complex when evaluated using such crude tools as readability formulas. Bruce and Rubin also point out (ibid.) that it is possible to devise texts which seem to be very readable when tested with the measures used, but which are in fact impossible to comprehend owing to a lack of genuine desire to communicate.

Then again, formulas as means of estimating difficulty and therefore accessibility level of a given text for large masses are presently used regularly to ensure minimum readability hygiene. For example, the State of Florida requires a given Fleisch-Kincaid minimum grade level for insurance documents. (Florida Laws: FL Statutes, 2013).

Cloze procedure has been used in readability assessment as well, but it has come in for some criticism, owing to the extent of variance in test setting. It has also been suggested that it is possible for test subjects to fill in the gaps from their experience and knowledge of familiar speech and language patterns. Despite shortcomings, however, with proper test setting Cloze procedures should have some use for readability studies. A major benefit of the method is that it efficiently takes the individual reader into account: the test reflects the full reading ability and linguistic capability of the reader and in this manner satisfies all aspects of the earlier definitions of readability. Also, the test itself matches a reader with a text intended for that reader. (Gilliland, 1972). Cloze procedure has since gained general recognition as a valid test method for reading comprehension (Williams et al., 2011). Cloze tests focus on the individual's ability to understand a text and rely more on the relationship of words than merely counting the frequency or length of words (DuBay, 2004).

Question and answer techniques have, in practice, not only been used to measure the difficulty of a passage but to provide a base line against which other measures can be compared. These question and answer techniques are primarily intended for measuring the comprehension of a passage, and do not produce data for assessing overall readability (such as fluency). There are also various technical shortcomings that limit the applicability of the question and answer technique. For example, although it is possible to assess whether a given text is easy or difficult to understand, it is impossible to determine which are the factors that lead to such a conclusion (whether the subject matter is too difficult for the reader or the phrasing too complex). Also the

testing situation (for example time limits) may affect the result. (Gilliland, 1972) Question and answer techniques may work in places where other measures prove unreliable, as for example with groups with learning difficulties or other intellectual disabilities: methods such as time-on-task say nothing about the quality of the reading event, and readability formulas may not be able to take into account the special needs of such a specific target group. Question-based methods include subjective assessments of the reading experience (ease/efficiency) that measure metacomprehension (although this might still be difficult for readers with intellectual disabilities) and objective comprehension questions. (Huenerfauth, et al., 2009). A question based approach may also be beneficial if the objective is to learn about the experience of reading and the experience of information acquisition, or if the participants cannot be tested in controlled test situations such as laboratories (Waycott and Kukulska-Hulme, 2003).

Because it is so difficult to find accurate quantitative methods for measuring overall readability, the qualitative paradigm may have something to offer researchers and practitioners. Gilliland refers to a study by Moyle (1971, in Gilliland, 1972) on consistency of grading carried out by a group of people rather than by individuals. The study showed that grading of books was more consistent when this was done by a committee (in this case a group of experienced teachers) than when undertaken by similarly experienced teachers individually (ibid.). Expert evaluation techniques used in usability engineering rely on setting clear targets and acceptability guidelines regarding the experts' assessment. In these evaluations the evaluators are expected to review a given user interface within certain parameters (Nielsen, 1993).

There are a great number of further studies into assessing readability. Michael Pressley (in Farstrup and Samuels, 2002) refers to a number of thinking-aloud methods that have been used in studying the active comprehension strategies of good readers. Thinking-aloud is also a useful technique in usability engineering. In such a test, the test subjects are asked to voice their thoughts on what they are doing, thinking and/or feeling. The consistency of results

derived from the tests referred to by Pressley (over 40 published studies), indicating certain mainstream strategies common to good readers, suggests that thinking-aloud may also be an effective test method in text evaluation as defined in the context of the present study. The test method itself is highly subjective and the interpretation of results requires a considerable input from the test moderator, but it would seem highly feasible to use the technique as a complementary test method alongside any other tests made on text and information accessibility.

Baker, et al. (in Davison and Green, 1988) mention, along with prediction formulas, “judgment of expert writers” as an alternative to actual testing. They find this judgment to be inadequate as a stand-alone, but when the evaluators in their example test were given user test data, such as transcripts of test subjects’ thinking aloud, the evaluators were able to improve the readability of the tested documents significantly.

Baker, et al. go on to suggest a number of strategies for testing users (readers) in real scenarios, for example by a thinking-aloud test. One test setting designed for a technical manual audience is to have the test subjects perform a given task following the directions in the manual and think out loud while performing it. Also, the test must be calibrated so that it reflects the practical information goals of the readers. In other words, if the text is used as a reference point only, it is not reasonable to test how well the readers have learned it. (Baker, et al., in Davison and Green, 1988).

As can be seen from this overview of methods for assessing comprehension and readability, the tools for the job are varied and none seem to be more comprehensive in nature than others. In the present thesis there is a strong impetus to discover a performance level that could be compared to experience. The methods selected for the experiments therefore ranged from modified Cloze method for reading comprehension for reasons of controlled test administration and clear scoring to open-ended questions on readability. The experience aspect was approached through Likert type grading, which is

a typical usability engineering industry method (discussed further in Section 4.2.1).

4.2 User experience and usability

When a human operated system functions in such a way that its operator is able to perform all the intended tasks and achieve the intended user-goals safely and without loss of data or content a minimum level of usability has been achieved. If these tasks have been completed in a way that also leaves the system user feeling relatively confident and happy about the way the system worked, a certain level of overall user experience (UX) has been reached as well.

Harry Brignull, a UX consultant and writer of 90percentofeverything blog says much the same:

> "Old-school Usability espouses the idea that user activities are onerous tasks that they want to get out of the way as soon as possible. While this is true in some cases, usability is now widely understood to be more of a hygiene factor – something that can cause dissatisfaction if missing, but its presence cannot take you beyond lack of dissatisfaction."
> (90percentofeverything, 2009)

'Usability' in the present chapter is used as a generic term to denote all such activities and methods that are used in improving or studying user experience. Usability in general falls under various categories in literature and research: human-computer interaction (HCI), user-centred design (UCD), man-machine interface (MMI), human-machine interface (HMI), operator-machine interface (OMI), user interface design (UID), human factors (HF) and ergonomics (Nielsen, 1993). User experience (UX) can, bearing in mind the quotation from Brignull, be seen as a complete service path of which usability is a part. The user's perception of how well he controls the environment in question, the quality of the content (relevance, trustworthiness, etc.) and the overall interface appeal (functionality, visual look) are elements which are ways of describing the broader scope of UX (Snitker, 2004).

The word ‘experience’ in itself contains an element of feelings and emotions (*Collins English Dictionary*, 1999). The design of systems, objects and interactions affect their users not only in a functional sense (“do A, gain B”) but they also often coax an emotional response: a system worked well, using it was fun and the service it provided was pleasant or efficient. Donald A. Norman suggests in his *Emotional Design* (2004) that when it comes to products the emotional side of a product’s design is more important than its practical side, and that essentially usability is one component of a design of a given product (others mentioned are aesthetics and practicality). A look at technology blogs and audience commentary on newspaper websites with any new product or operating system announcements from the biggest mobile device platform manufacturers shows a strong emotional response to these systems or their brands: fierce debates by the so-called fan boys over the pros and cons of whether Android’s latest update is a better system than the most recent iOS by Apple or by Microsoft illustrate just how emotional technology users can be over the products they use in their everyday lives.

The interview study in the present thesis (Chapter 9) indicated negative attitudes towards reading continuous text on a mobile device screen. The initial responses by participants who had never really read anything more continuous than text messages on their mobiles deemed reading to be too difficult, strenuous and awkward for them to want to do it. Research into emotions and decision making shows that mood affects not just decision making but information processing (Schwarz, 2000): a negative mood tends to push individuals towards detail-oriented bottom-up processing strategies, whereas a positive mood nudges individuals to opt for more heuristic, top-down processing strategies.

4.2.1 Usability testing: ensuring minimum UX hygiene

Usability in itself is an abstract, and requires closer definition through more concrete concepts in order to gain a more tangible shape. These more

concrete concepts are: learnability, efficiency, memorability, errors and satisfaction. In addition, usability is always relative to a set task and set users and the optimal usability defined for one such set may not be directly applicable to another. Furthermore, it may not be possible to end up with the optimal result in all the listed categories (for example, a highly learnable system will probably not reach optimal efficiency levels in the beginning), and usability trade-offs are necessary. This means balancing the importance of the usability categories according to usability goals and criteria set in a development project. (Nielsen,1993).

Usability engineering and information design use similar methods for defining their ultimate goals. At the core of these methods is a close analysis of users and their tasks. Nielsen (1993) lists three areas in which users (even within a basically homogenous group) will differ from each other on an individual level: their experience with the system in question, with computers in general and with the task domain in question. However, typically user experience is discussed on a simplified axis of novice – expert (or somewhere in between). The difference between system expertise and domain expertise can be considerable in cases where an established domain expert is adopting new software. Expertise in the domain field allows for specialised terminology in the interface, but lack of experience in the system requires usability parameters appropriate for a novice.

Other aspects in which users can differ from each other are (for example): age and gender (which are relatively easy to observe), spatial memory, reasoning ability and preferred learning style (all of which are more difficult to observe). In addition, there will be differences in the general capacity and ability for carrying out given tasks: some people are simply more productive and efficient than others. Because of these differences it is important in usability engineering to know as much about the actual users as possible. Through an understanding of the user demographics it is possible to set goals for the complexity levels of a given system under development. In addition to users

themselves, it will help to know the context of use (office environment, noise levels, etc.). (Nielsen,1993).

One major aspect of user interfaces and systems that is scrutinized in usability engineering is consistency. The usability activities for coordinating the interface do not only cover what is on the computer screen, but the entire package: application screens, documentation, online help and training material (for example tutorials).

The toolbox a usability researcher has at his or her disposal these days is immense: variations and adaptations of a plethora of methods from ethnographic origin to experimental psychology and heuristic evaluations are available so that a researcher can always pick the right tool for each research need. Cognitive walkthroughs, expert evaluations, focus groups, co-design, interviews and task-based laboratory testing are all typical usability research methods used around the world. The Internet offers various sites such as Usability First (usabilityfirst.com) that list and describe these methods that usability researchers typically use.

In usability research the test protocols not only focus on how well end-users cope with tasks on a system, but these protocols often also include ways of gathering data on the experience of using a system. The methods used typically include interviewing and experience response scales like Likert on specific perceived aspects of the device or system experience like the perceived ease of use. In studies that contribute to the present thesis it was natural to turn to such methods when aiming to measure and understand the user experience involved in reading text on small screens. In the present thesis experience responses have been collected by interviewing (semi-structured interview, open-ended questionnaire), and using the Likert scale in both a web questionnaire and as part of a task-based test session.

Task-based tests aim at discovering mistakes and problems within an interface at a functional level: for example, are all the necessary items visible and does the user understand what they mean and how to manipulate them? To explore the experience of using a system/interface, there are a number of set questionnaires that have been developed for industry purposes such as USE (Lund, 2000), SUMI , SUS (System Usability Scale) (Brooke, 1996) and QUIS (http://www.lap.umd.edu/quis/) which all aim at satisfying the ISO 9241-11 usability requirements (correctly titled “Ergonomics of Human System Interaction”). The questionnaires have a different number of questions in a number of categories or aspects that essentially aim at probing responses for how useful, usable and satisfactory the various elements in a tested system are.

The said questionnaires have been designed rather specifically to gauge the perceived usability of a system and although they do query users’ satisfaction and their experience of whether the system was easy to use, the bank of questions can be a bit overwhelming when a researcher is more interested in users’ experience of a specific aspect of an interaction event. In the context of the present study the questions set for the test participants focused on the “experience of reading” more than on how the participants perceived the functionality of the interface. ‘Experience’ was presented to the participants through a range of questions from the importance of some features or styles in creating a positive experience to participant experience of fluency and how comfortable the reading experience was. The questions, in other words, were created to directly address the research questions instead of using a set questionnaire focusing on interface functionality.

The present chapter along with the preceding three chapters have aimed at providing background into the particular characteristics of reading on small screens, based on existing literature. The following chapters will present six experiments and studies into mobile reading.

CHAPTER 5

In which reading comprehension and the subjective experience of reading are measured and compared.

Reading comprehension test on paper and on two mobile device conditions

The present chapter describes the experiment design, data collection and analysis and the results of an experiment in which participants were asked to complete a reading comprehension test on paper and on a mobile phone. In the test text samples were rotated between three conditions: (i) text on paper, (ii) text on a mobile, and (iii) text on a mobile with the capacity for participants to adjust viewing settings on the device. The test was a modified Cloze procedure and the results were compared for significant differences in the number of correct answers. Time on task was also loosely monitored and compared between the conditions for additional insight into task performance.

In the experiment text was presented on paper, on a mobile device screen (3.5” in size) where text was set closely to match the font size and layout of the paper version (the paper font size was close to newspaper print font size and the typefaces were those of some of the original source texts and as such close to those used in many newspapers and magazines, and the text on mobile was set to replicate those parameters), and on the same mobile device where participants were allowed to adjust the viewing parameters (text size, spacing, typeface, margins, background colour and orientation of the device when reading, i.e. portrait or landscape) to their own preferences. A reading comprehension test was devised and given to participants in the different conditions. After the comprehension test was completed, participants also

rated their reading experience in terms of fluency and ease so that the perceived experience could be compared to the comprehension test performance. The experiment was eventually compared to the results from an interview study (Chapter 9) on attitudes and expectations on reading text on a mobile phone.

The three main research questions the experiment was geared to explore were:

1. Does reading comprehension level differ significantly between text formats (paper and mobile device screen)?
2. Does making individual adjustments to the mobile device text appearance and layout improve reading performance?
3. Do perceived experiences on ease and fluency of reading correlate with the comprehension test performance results?

An early study into comparing reading rates between paper and an electronic screen by Gould et al. (1987) discovered conditions under which participants were able to read on a CRT (Cathode Ray Tube) display as efficiently as from a good paper page: quality of the display and the text characters used were in a key position. High enough resolution on the display and typefaces similar to those on the printed paper allowed participants' reading to be as efficient on screen as on paper. Although this research dates back to 1987, and since then the development with displays has taken considerable leaps forward, it is worth noting that display quality and screen size were among the top concerns with the participants in a study on students' responses to reading course materials on PDA devices by Waycott and Kukulska-Hulme (2003).

Theories on reading comprehension that emphasize the need for deduction and inferences as well as the role of working memory in reading comprehension (Just and Carpenter, 1992; Graeser, et al., 1994) considered in the context of split attention and fragmentation of content (Sweller 1994, Sweller and Chandler, 1994 in Oostendorp, 2003) suggest that a size-limited interface (small mobile screen) compared to a larger interface sheet of paper

would present comprehension differences between these formats. There are individual differences in reading range (a measure of working memory capacity) (Daneman and Carpenter, 1980). Increased cognitive load from the reader having to integrate fragmented content (Sweller 1994, Sweller and Chandler, 1994 in Oostendorp, 2003) was anticipated owing to the text being split over various small pages (screens) rather than on one larger area that would not impose physical navigation needs. The expected outcome of the experiment was that the mobile device reading would not only make a difference because of legibility, but also because of the increased need for interaction: turning pages more frequently. The need for interaction would interrupt reading flow and place additional pressure on working memory, thus requiring more work to make sense of the sentences that had words missing here and there. Dillon, et al. (1990) also expected that splitting sentences across screens would interrupt the comprehension process.

To summarise, the experiment was a repeated measures design where sets of texts were rotated between three conditions: a paper printout in newspaper-style column layout, a mobile phone with the text layout (text size and column width) similar to the paper layout, and an adjusted mobile phone where participants were allowed to adjust layout and other viewing conditions on the device. Based on earlier literature on reading on electronic displays and the cognitive processing concerns mentioned above the main hypothesis of the experiment was that there would be a significant difference in the reading comprehension results at least between paper and on mobile device screen when layout of a given text is the same as the paper version layout as closely as possible, and that paper would produce a higher score on correct answers.

5.1 Method

Participants completed a reading comprehension test under three conditions: paper, mobile device and mobile device with bespoke viewing adjustments. The test was constructed of three sets of texts and reading comprehension was assessed with a modified Cloze procedure test method. The first two conditions (paper printout and mobile device) were rotated. Finally,

participants were allowed to adjust the layout and appearance of the text on the mobile device screen before completing the last text set. The three text sets used in the experiment were rotated throughout the conditions, as illustrated in Table 1. After completing the test, users were asked to rate their experiences (ease and fluency of reading) on a Likert type 1-7 scale.

5.1.1 Design

The experiment was a mixed-methods approach with quantitative data collected from performance scores (modified Cloze procedure reading comprehension where correct answers provided the score) and participant self-assessment (perceived scores) and qualitative data gathered from open-ended questions. These data produced the dependent variables for the experiment. The independent variable in the experiment was the three sets of the test texts used in the experiments: paper printout, mobile phone and mobile phone that users were allowed to adjust to their liking (layout and other viewing conditions).

The comprehension test administered was a repeated measures design with a modified Cloze procedure test with three text sets made up of three sub-sections each. Latin squares were used for balancing the possible effect of starting order on reading comprehension, as unfamiliarity with the procedure and being in a test situation in themselves are likely to influence participant performance somewhat. The text sets were also rotated from condition to condition to diffuse any potential bias where a specific text might work better or worse in a specific format. Cloze procedure was selected as the comprehension test method owing to Cloze procedure's general recognition as a valid test method for reading comprehension (Williams et al., 2011).

The following rotation matrix shows how the conditions (paper, mobile, adjusted mobile), test sets (Set 1, Set 2, Set 3) and participants were distributed in the test setup. 'Rotation group' in the left column simply indicates that the three basic rotations were further split into two subgroups to further

diffuse the impact of the starting order of the conditions. The full test texts are included in Appendix 12.1 and the perceived ratings and participant background information questionnaire are available in Appendix 12.3.

Table 1 Rotation matrix for reading comprehension test.

Rotation group	Participants	Test 1	Test 2	Test 3
A1	1 - 8	Set1/P	Set2/M	Set3/AM
A2	9 - 16	Set2/M	Set1/P	Set3/AM
B1	17 - 24	Set3/P	Set1/M	Set2/AM
B2	25 - 30	Set1/M	Set3/P	Set2/AM
C1	31 - 38	Set2/P	Set3/M	Set1/AM
C2	39 - 45	Set3/M	Set2/P	Set1/AM

5.1.2 Participants

The 45 participants were a sample based on opportunity, mostly made up of staff and undergraduates at Northumbria University Department of Psychology. Power was determined for three F-tests (ANOVAs) for repeated measures when there is one group of subjects. Without an estimate of how big the differences in scores would be, conventional effect size estimates were used: the desired power level was 80%, α – level was at 0.05, and a small correlation between conditions was estimated (0.25). The estimate correlation is based on the assumption that regardless of reading format or test method, there would typically be an observable difference between those who perform better and those who perform worse in such tests. Power calculations were carried out on power analysis software G*Power (Erdfelder, Faul, and Buchner, 1996), which indicated a minimum sample size of 45 for a three-way ANOVA test.

Table 2 Participant age and gender; N=45.

Answer options	Response per cent	Response count
18 - 25	75.6%	34
26 – 35	15.6%	7
36 – 45	4.4%	2
46 – 55	4.4%	2
Female	82.2%	37
Male	17.8%	8

5.1.3 Materials

In each text set used in the experiment the set was made up of three passages, which were rated as easy, medium and difficult using the Flesch-Kincaid reading ease test. The Flesch-Kincaid reading ease score is a 100 point scale where the higher the score the easier the passage is estimated to be. The rule of thumb is that most documents or passages should settle in the 60-70 range to ensure that they can be understood by the general public. The formula for the score is 206.835 – (1.015 x ASL) – (84.6 x ASW), where ASL means average sentence length and ASW means average number of syllables per word. For the purposes of the experiment, the automated Flesch-Kincaid scores calculations provided by Microsoft Word word processing software were used. (Flesch R,1948).

The texts in question were selected from a variety of sources, mainly magazines, newspapers and periodicals. Some of the texts were originally from magazines but selected for the present experiment from texts that had been used as reading comprehension tests in Finnish matriculation examinations for English as first or second foreign language (these test texts are carefully set and constructed to provide a controlled level of difficulty), or they were published as content extracts in a weekly magazine that collects and summarizes the previous week's news from UK media (*The Week*). The texts were selected on the basis of their reading ease and length (word count). The sets were compiled from three passages each in order to reduce the risk of subject matter influencing the scores too strongly. The texts used in the experiment can be found in Appendix 12.1.

Table 3 Test passage rotation in reading comprehension test.

	Set 1	Set 2	Set 3
Passage 1: Easy	Words: 252 Reading ease: 68.6	Words: 182 Reading ease: 76.7	Words: 183 Reading ease: 67.7
Passage 2: Medium	Words: 230 Reading ease: 55.7	Words: 296 Reading ease: 49.2	Words: 200 Reading ease: 57.8
Passage 3: Hard	Words: 136 Reading ease: 28.0	Words: 122 Reading ease: 33.2	Words: 176 Reading ease: 36.2
Total	Words: 618 Reading ease: 55.4	Words : 600 Reading ease: 55.4	Words: 559 Reading ease: 56.8

The test itself was an open-ended modified Cloze procedure (Enkvist and Kohonen, 1976) test where approximately every 10th word had been removed. While typically Cloze procedure tests tend to remove every 5 – 7 words (Enkvist and Kohonen, 1976), the present experiment aimed at maintaining a reasonable reading flow between the gaps.

The approximation was due to avoiding blanks on words that would require specific knowledge of the subject matter (such as people's names) or other words that could not be deduced from the text and context otherwise. Answers were marked contextually, meaning that more than one answer could be correct depending on the context of the text. The range of all answers given by participants and the selection of accepted answers is shown in the 'Answer sheet' in Appendix 12.2. For determining a fair marking policy for the range of provided answers, a lecturer in English Philology (a native English speaker) with considerable experience specifically in setting contextually marked modified Cloze tests for English philology majors at university level was consulted.

The texts were provided on an A4 sheet in 10pt typeface and in two columns, much as a magazine or a newspaper article would typically look like. On the mobile device (before adjustments) screen the layout followed the words per line count from the paper version, except for presenting the text in a single column. The mobile device used was a first generation iPhone on which the texts were presented in .pub file format on a free e-book reader software Stanza (http://www.mystanza.com/). The basic page-turning interaction on Stanza was a horizontal swipe from right to left (forward) and left to right (back to previous page). Öquist and Lundin (2007) examined various presentation methods for mobile phones, establishing the fastest presentation method that did not compromise reading comprehension. They compared three presentation methods: Paging (page by page), Leading (one line of text steadily scrolling across the screen) and RSVP (Rapid Serial Visual Presentation) and found that Paging, overall, offered the best readability on a mobile device. Based on Öquist and Lundin's findings, there was no reason to assume a negative impact on the results from the page-by-page interface on the selected e-reader software.

Practice

Windscreen wipers

The first patent for windscreen (1)_____ was issued to a woman from Birmingham, Alabama in 1903. Reportedly, Mary Anderson noticed (2)_____ a visit to New York that when it rained the drivers of trolley cars either had to leave the windows (3)_____ and expose passengers to the weather or get out (4)_____ few minutes to wipe their windscreens by hand. Anderson came up with a device that could be operated by hand from inside the car. Sadly, she made little (6)_____ from her invention since no manufacturer could see the value of it.

Source: BBC History Magazine, 2007

SET 1

How I poisoned my family

THE novelist Nicholas Evans once declared: "Guilt is my subject." That was (1)_____ the warm August evening in 2008 when he almost killed half his family. Since then, he says: "I've taken (2)_____ research to a rather extreme degree." Evans and his wife, Charlotte, were staying with her brother and sister-in-law (3)_____ Scotland, when Evans picked some wild mushrooms and fried them up for supper. Within 24 hours they were all in (4)_____ fighting for their lives, as one by one they went into kidney failure. They survived, but remain on dialysis. Evans's 29-year-old daughter, Lauren,

For the Best Pi
Lie Down

SCIENTISTS (13)_____
training 61 people in moto
verbal tasks: tapping a key
sequence, discriminating (
shapes on a computer scre
a list of words. Then the sc
divided (15)_____ su
groups. The first took a na
3, the second (16)_____
milligram caffeine pill, and
placebo (like medicine but
The subjects (17)_____
had been taught earlier an
researchers who did not kn
they (18)_____ in. Th
caffeine had worse motor s
napped or had a placebo. I
(19)_____, the nappe
better than either the caffe
placebo group. On the verb
were best by a wide margin
consumers did no (21)_____
given a placebo. Despite th
performance, caffeine take
reported less sleepiness th
"People think they're smar
(22)_____," said Sar
assistant professor of psyc
University of California, Sa
lead author of the study. "
strong argument for taking
having a cup of coffee."
Source: The New York Times, Dec

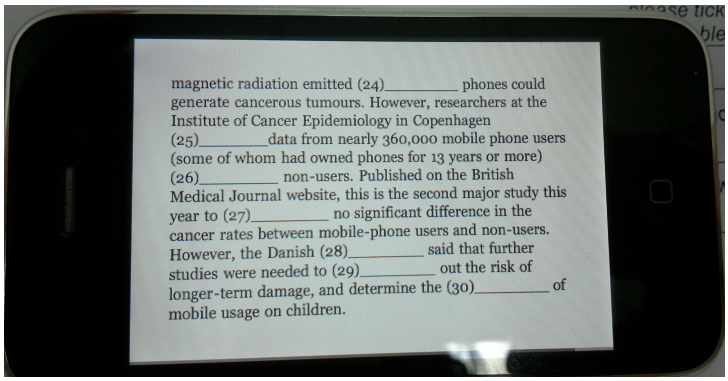
Mobiles Are Pr

Practice:

Windscreen wipers

The first patent for windscreen (1)_____ was issued to a woman from Birmingham, Alabama in 1903. Reportedly, Mary Anderson noticed (2)_____ a visit to New York that when it rained the drivers of trolley cars either had to leave the windows (3)_____ and expose passengers to the weather or get out (4)_____ few minutes to wipe their windscreens by hand. Anderson came up with a device that could be operated by hand from inside the car. Sadly,

Figure 7 Test text sample on paper (left) and the same as presented on the phone screen at the start of the test (right). This illustration is a compilation of a photographic capture of a segment of the paper format and a screen capture taken off the device used in the tests, and shows how the text size and layout were presented to the test participants. A precise replication of the paper format text is provided in Appendix 12.1.



magnetic radiation emitted (24)_____ phones could generate cancerous tumours. However, researchers at the Institute of Cancer Epidemiology in Copenhagen (25)_____data from nearly 360,000 mobile phone users (some of whom had owned phones for 13 years or more) (26)_____ non-users. Published on the British Medical Journal website, this is the second major study this year to (27)_____ no significant difference in the cancer rates between mobile-phone users and non-users. However, the Danish (28)_____ said that further studies were needed to (29)_____ out the risk of longer-term damage, and determine the (30)_____ of mobile usage on children.

Figure 8 1st generation iPhone with a sample of test text; landscape orientation.

The adjustments available to the participants were appearance related and layout related. Yi, et al. (2011) compared column numbers and line spacing in e-readers in terms of readability, comprehensibility and satisfaction and found that such appearance elements had an immediate effect on the three factors they were looking into. Also Bernard et al. (2003) compared typeface sizes, types and formats on computer-displayed screens and deemed that what has been generally found functional on paper may not be similarly functional on a computer screen. Accordingly, including an adjusted mobile in the test set was important and meaningful in order to gauge the effect of participants' personal layout and appearance preferences on their reading performance and experience.

Appearance adjustment options on Stanza application:

- theme (pre-defined list of colour schemes and background textures),
- font (selection of serif and sans serif fonts),
- font size,
- background colour,
- text colour,
- link colour,
- background image and
- (background) image opacity

Layout adjustments:

- alignment (justified, left, centre, right)
- hyphenation
- margins (width)
- line spacing
- paragraph spacing
- paragraph indent

In addition to these appearance and layout settings, participants could turn the device into landscape orientation, thereby displaying the text in landscape orientation as well (still single column, as in portrait orientation but with longer and fewer lines).

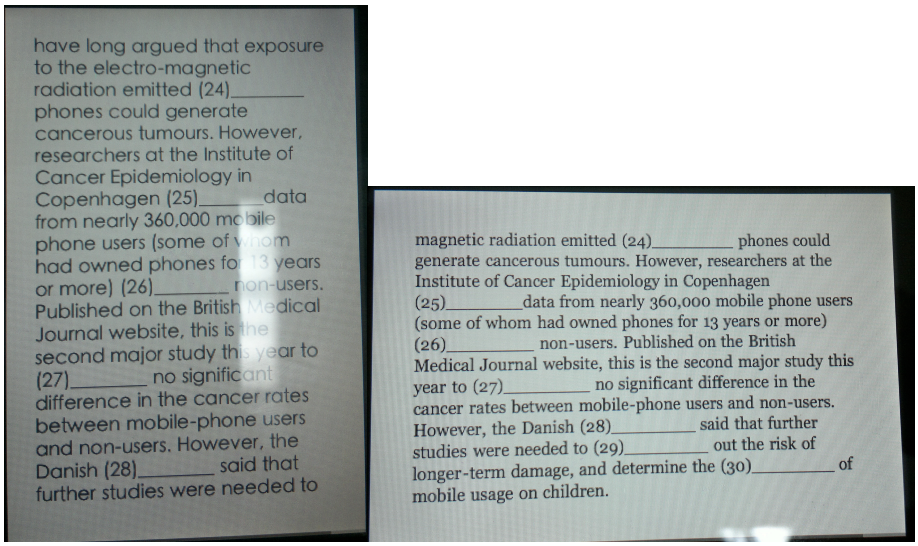


Figure 9 Sample of test text in portrait (left) and landscape (right) orientation.

Along with the comprehension test, participants were asked initially to provide some background information about their computing experience, current mobile phone and mobile phone use. The questions are presented in Appendix 12.3. After the comprehension test participants were asked to fill in a further questionnaire where they rated their experience of completing the test and of generally reading under the three conditions in terms of

a) Ease of completing the test on mobile/paper/adjusted mobile, and
b) Fluency of reading text on paper/mobile/adjusted mobile.

In addition to these questions, participants were asked to indicate which format was best for them (favourite).

The questions in the experiment focused on perceived ease and fluency as generic terms for acceptable user experience that could be used as measures of subjective experience to compare with actual performance measures.

5.1.4 Procedure

The participants were tested one by one and the researcher moderated the tests. Time on task was taken discreetly so as not to add performance pressure. In line with this discretion, timing was not recorded any more precisely than at minute level to provide basic indication on whether

completing each section had taken noticeably longer at any point, compared to other sections. Each text set (see Appendix 12.1 for full text sets) had 30 gaps of one word spread out at regular intervals (modified somewhat, however, to avoid gaps on proper nouns and other items that would require specific knowledge of the topic from the participant). Visually, the gaps were of equal length so that participants could not use the gap size as a clue to determine what type or length word was missing.

Upon arriving at the test room, the participants were told what they were expected to do, and that the experiment had ethical clearance from the Northumbria University School of Life Sciences Board of Ethics. Participants were also assured that the study was not interested in their individual reading comprehension or competence as such, but that the focus was on comparing the different conditions, namely paper vs. mobile device screen.

At the start of the test participants were asked to fill in a short background information survey on a computer provided. The survey simply asked for details such as participant gender, age group and what mobile phone they were using at present. After this the Cloze test procedure was explained in more detail with the help of an example passage (named ‘practice’ on the test sheets). Participants were instructed to complete each task (set of texts) independently, alerting the moderator when they were finished with each. The answers were filled onto a specific answer sheet (paper) to reduce interaction with the actual Cloze text to a minimum. In other words, participants did not have to try and fill the gaps in the text by writing or typing directly onto the Cloze text, as this would have been somewhat complicated and slow on the mobile device. The answer sheet also provided participants with visibility into how many gaps there were in each set.

The final condition in each session was on a mobile device where participants were allowed to change some of the viewing parameters of the device screen: what parameters could be changed and how this was done was shown to the participants and the moderator offered assistance where necessary in making

these adjustments as requested by the participants. Such assistance was necessary for some users who were not confident as to the exact interactions with the particular application software used.

At the end of the tasks participants were asked to fill in a short questionnaire on a computer provided on their perceived assessments on ease of completing the tasks, fluency of reading on the different formats and to indicate which format suited them the best personally.

5.2 Results

Analysis of variance was performed on the performance data (comprehension scores, time on task and perceived ratings on ease of completing the test under each condition). In addition, for the question of screen adjustments preferences, the frequency of adjustments vs. no adjustments was noted. Participants were also asked to comment freely after the test on the ease and fluency of the different conditions as well as on why they selected one of them as their preferred choice over the others. These comments were analysed with the grounded theory method in order to categorize the comments thematically (Järvinen and Järvinen, 2000). The process of grounded theory method was the same as that described in Section 9.1.1.1.

5.2.1 Participant background

Self-assessment (perceived) computing skills:

Participants were asked to self-assess their computing fluency before starting the comprehension test: “On a scale from one to seven, where one means "not fluent at all" and seven means "extremely fluent", how fluent would you say you are with computers and technology in general?” On the whole, participants tended to perceive their computing fluency to be at the higher (more fluent) end of the scale, no one estimating themselves at the very low end at all (rating 1-2).

Table 4 Descriptive statistics: perceived computing fluency.

	N	Minimum	Maximum	Mean	Std. Deviation
Perceived computer skills ratings	45	3.00	7.00	5.4000	.91453

Current mobile phone

Participants were also asked to provide information on which mobile phone they were using at present. The purpose of the question was to see if there was a clear majority of so-called smartphones or touch-screen phones, which tend to have larger screen size than conventional hardware keypad phones.

Table 5 Participants' current mobile phone.

	Frequency	Percent	Valid Percent	Cumulative Percent
Blackberry	2	4.4	4.4	4.4
Blackberry Bold	5	11.1	11.1	15.6
Blackberry Curve	5	11.1	11.1	26.7
Blackberry Torch	1	2.2	2.2	28.9
HTC	1	2.2	2.2	31.1
HTC Desire	2	4.4	4.4	35.6
HTC Desire S	1	2.2	2.2	37.8
HTC Desire Z	1	2.2	2.2	40.0
HTC Mozart	1	2.2	2.2	42.2
HTC Sensation	1	2.2	2.2	44.4
iPhone	2	4.4	4.4	48.9
iPhone 3	1	2.2	2.2	51.1
iPhone 3gs	2	4.4	4.4	55.6
iPhone 4	10	22.2	22.2	77.8
iPhone 4s	1	2.2	2.2	80.0
monoblock	1	2.2	2.2	82.2
Nokia (not a smart phone)	1	2.2	2.2	84.4
Nokia (touchscreen)	1	2.2	2.2	86.7
Samsung	2	4.4	4.4	91.1
Samsung (touch)	2	4.4	4.4	95.6
Sony Ericsson (non-touch)	1	2.2	2.2	97.8
Sony Ericsson k800i	1	2.2	2.2	100.0
Total	45	100.0	100.0	

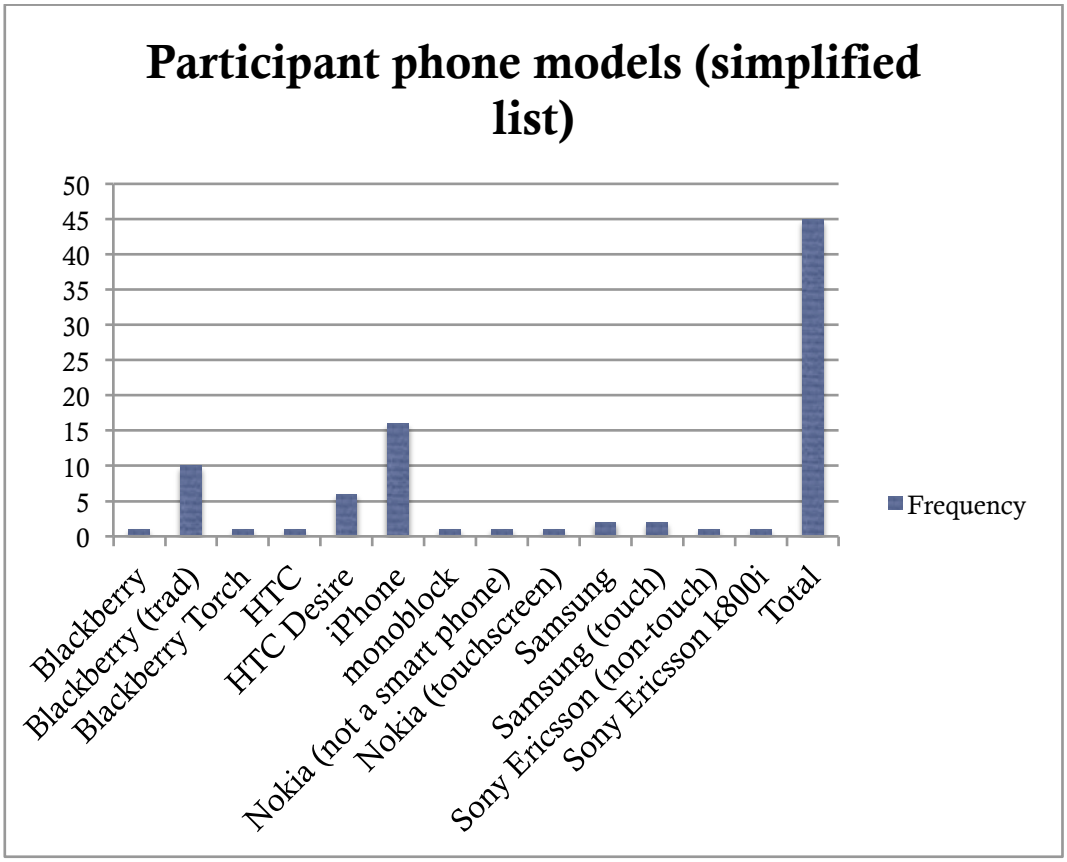


Figure 10 Simplified chart of test participants' phones at the time of the experiment.

Table 5 lists the phones used by test participants at the time of the experiment. The noteworthy element in the list is that at least 14 of the models are full touch phones with screen sizes approximately between 3” and 4.5” and a single hybrid with a full size touch screen complemented by a sliding hardware keyboard. In the bar chart the phones have been grouped by brand and by information given by participants about what model they have exactly. Participants were allowed to fill in the information freely in the questionnaire and not everyone necessarily remembered the exact model of their phone.

Models that can immediately be identified as touch phones or other phones with similar screen size account for some 60% of the devices. The next largest

group is participants with traditional Blackberries with a hardware keyboard. These phones are geared by design towards frequent text-based messaging.

5.2.2 Analysis of variance: scores, time on task, perceived ratings

Analysis of variance was performed on the dependent data from the participants: test score (correct answers), time on task (minutes), perceived ease of completing tasks in each condition (1 to 7 score, 7 being the positive “extremely easy” response,) and perceived fluency of reading text under the given formats (1 to 7 score). The conditions are referred to as paper, mobile (no adjustments by participants) and adjusted mobile (adjustments to presentation of text allowed, though not compulsory). 95% confidence level was used throughout. Maximum score per text set (and, thus, condition) was 30.

Table 6 Descriptive statistics for correct answers, time on task, perceived ease of completing tasks and perceived fluency of reading.

	Paper	Mobile	Adjusted mobile	N
Score (correct answers)	23.49 (2.897)	23.31 (3.302)	23.27 (2.484)	N = 45
Time on task (minutes)	8.80 (2.258)	9.09 (2.532)	8.14 (2.075)	N = 44
Perceived ease of completing tasks (1 – 7 score)	5.74 (.978)	4.84 (1.344)	5.44 (1.297)	N= 43
Perceived fluency of reading text in the given formats	5.77 (1.179)	4.70 (1.286)	5.36 (1.399)	N = 44

A one-way repeated measures ANOVA on the test scores (correct answers) showed that there was no significant effect of text format on the number of correct answers. Despite the ‘adjusted mobile’ condition being the third and last one for each participant, there seemed to be no effect on the number of correct answers.

A one-way repeated measures ANOVA on the time on task (minutes) determined that there was a significant effect of text format on time on task, Wilks’ Lambda = .814, $F(2, 42) = 4.803$, $p<.05$. Post hoc analysis using Bonferroni correction revealed that time on task differed significantly between mobile and adjusted mobile conditions ($p<.05$). Participants took more time to complete the tasks on a mobile device ($M = 9.09$) than on a mobile device where they had had the option to adjust the viewing parameters ($M = 8.14$). There was no significant difference between paper ($M=8.80$) and mobile or adjusted mobile, but paper condition seemed to fall somewhere in between the two other conditions in terms of time on task.

A one-way repeated measures ANOVA on the perceived rating on ease of completing the task under the given conditions showed a significant effect of

text format on perceived ease, Wilks' Lamda (2,41) = .717, $p<.05$. Post hoc analysis using Bonferroni correction revealed that perceived ease of completing task differed significantly between paper and mobile conditions ($p<.05$). Participants rated completing the test from paper format text (M = 5.74) as easier than completing the test from mobile device format (M = 4.70).

A one-way repeated measures ANOVA on the perceived rating on fluency of reading the given text formats showed a significant effect of text format on perceived fluency of reading, Wilks' Lambda = .798, $F(2,42) = 8.676$, $p<.05$. Post hoc analysis using Bonferroni correction showed that perceived fluency of reading differed significantly between paper and mobile ($p < .05$) as well as mobile and adjusted mobile ($p<.05$). Participants rated fluency of reading to be better on both paper (M = 5.77) and on an adjusted mobile device (M = 5.36) than on a mobile device (M = 4.70).

To summarize, there was no significant difference in task performance (correct answers) between the conditions, but in time on task participants took significantly longer to complete the tasks on the mobile than on the mobile where adjustments to viewing conditions were available. There was no significant difference between paper and mobile phone or paper and adjusted mobile in terms of time on task. Completing the test on paper was perceived to be significantly easier than on a mobile phone (perceived ease) and participants rated fluency of reading as significantly better on both paper and adjusted mobile device than on the mobile device without adjustments to the viewing conditions.

A clear majority of participants made some level of adjustments to the viewing conditions of the text on the mobile device. Without differentiating further the magnitude or focus of those adjustments, 39 participants out of 45 made some adjustments. This accounts for 86.7% of the participants. Owing to the clear minority of participants who did not adjust the screen in any way, it was not feasible to compare the results between those who made adjustments and those who did not.

Table 7 Making adjustments to the mobile device screen. N=45.

	Frequency	Percent	Valid Percent	Cumulative Percent
No adjustments	6	13.3	13.3	13.3
Adjustments made	39	86.7	86.7	100.0
Total	45	100.0	100.0	

Table 8 Descriptive statistics for perceived ease and fluency between formats.

	N	Minimum	Maximum	Mean	Std. Deviation
I found the paper format of the test easy to complete.	43	3.00	7.00	5.74	.98
I found reading on paper format to be fluent.	44	2.00	7.00	5.77	1.18
I found the mobile phone format (without modifications) of the test to be easy to complete.	45	2.00	7.00	4.89	1.34
I found reading on the mobile phone format (without modifications) to be fluent.	45	2.00	7.00	4.71	1.27
I found the mobile phone format of the test that I had modified myself to be easy to complete.	45	1.00	7.00	5.49	1.29
I found reading on the mobile phone format that I had myself modified to be fluent.	45	1.00	7.00	5.36	1.38

In addition to asking participants to rate the three reading formats in question, they were also asked to nominate their preferred format. Paper was, as perhaps could be expected, a clear favourite while personalized adjustments to text visualization (adjusted mobile) was the next favourite. Section 5.2.3 addresses the perceived reasons for the choice of favourite in more detail.

Table 9 Preferred reading format.

	Frequency	Percent	Valid Percent	Cumulative Percent
Paper	28	62.2	62.2	62.2
Basic mobile	6	13.3	13.3	75.6
Adjusted mobile	9	20.0	20.0	95.6
Other	2	4.4	4.4	100.0
Total	45	100.0	100.0	

5.2.3 Analysis of free participant comments on perceived ratings

Having rated the ease of completing the tasks on paper and fluency of reading on paper, participants elaborated on their scores in their own words. Grounded theory approach was adopted for analysing the open-ended comments users volunteered in the questionnaire at the end of the comprehension test. The questions simply asked why participants had rated as they did with the intention that the questions would not limit participants' answers too tightly and the participants would have the opportunity to voice any reasons or concerns that they had thought of. These open-ended comments from participants were based on their subjective experiences immediately after the comprehension test.

An interview study on the expectations and attitudes towards reading on small displays used grounded theory to analyse the responses and to formulate a model of relationships and causality of the discovered categories (Chapter 9). The same technique was used in the present experiment to analyse the responses in the present experiment in order to gain a "before and after" view of these subjective experiences and attitudes. It should be emphasized, at this

point, that the collected qualitative data is by no means pure “before-after” data in the sense that the participants in the interview study in Chapter 9 and the participants in the reading comprehension test (present chapter) were not the same and the generalized observations and findings cannot be interpreted as progression or development of individual participants’ views on the matter. The grounded theory analysis process is described in more detail in Section 9.1.1.1.

The open-ended responses on *why* the participants rated the various text formats in terms of ease and fluency of reading the way they did were reduced to conceptual statements (open coding) (Järvinen and Järvinen, 2000). These statements were, in turn, arranged under categories (ibid.). The process of analysis then differed from a full grounded theory approach in the final part: instead of performing the selective coding analysis (ibid.) solely on the data from the open-ended responses from the post-comprehension test questions and forming an independent model of relationships and causalities between the categories, the categories from the axial coding phase were projected against the categories discovered in the said interview study. The purpose of projecting the categories in this manner was to examine how many of the initially discovered categories from the interview study (basically, the “before” state) were still present in the post-comprehension test experience and to then observe if the relationships and causality of the relationships could still be seen in the new data.

5.2.3.1 Categories of statements

The questions were set with reference to the three text formats used in the comprehension test. For the first two conditions (paper and mobile), participants elaborated on a simple “why?” question once they had rated their experience. A mobile device with the possibility to adjust appearance and layout of text encouraged responses in which participants compared the adjusted mobile condition of reading to the previous two. Because of this change in the answer style, the answers to the first two formats (paper and mobile) were analysed and categorized separately, and the answers were

then compared to the mobile reading acceptance model (Section 9.2) established from the said interview study. The following sections outline the categorization for each text format. The adjusted mobile responses were analysed in the same manner (reduced to conceptual statements, then categorized).

The following two sections explore the categorized participant responses in the context of the questions as presented in the end-of-test questionnaire. These sections are followed by a review of the categorized responses to adjusted mobile device reading experience. The latter is presented individually, as the comments very much tended to compare the last condition of the test (adjusted mobile) with the first two. At this point participants now quite fluently employed a more analytical approach to their experience, aiming to explain and verbalize more accurately what was different about their performance and experience with the third condition. Hence, the categorization of the statements went up a somewhat different path from the first two conditions. The categories that emerged from these statements were more explanatory of the shortfalls or the benefits of the other two test conditions.

In the following sections the example statements from participants are presented exactly as they were typed into the (electronic) questionnaire at the test. Since the comments were made during the test session without corrections themselves, using an alien Scandinavian keyboard layout on an unfamiliar laptop (researcher's laptop), we should not pay too much heed to the spelling and typos.

5.2.3.1.1 Text on paper

Two major groups of comments emerged from all the comments made on completing the test from paper format text and on perceived fluency of reading on paper: familiarity or convention of reading on paper, and paper facilitating overview/preview type activity as well as ease of navigation within the text.

Familiarity or convention here simply refers to participants stressing a preference for the paper format because they are more used to it or more familiar with it. In the second large group (approximately equal in size to the familiarity group) participants expressed the importance and positive influence good overview and preview capacity had on reading and test completion. Bearing in mind the nature of the modified Cloze procedure test, scanning and backtracking (moving back within the content to check earlier parts, sentences or phrases) within the test texts will have been necessary for many in order to fill the gaps in the texts in a meaningful way. It is not surprising, then, that many of the comments in this area mention scanning, skimming and backtracking specifically.

Familiarity and convention:

The most typical reason given for how participants scored the paper format experience was simply that paper format was familiar and conventional. As explained earlier in this chapter, the starting order of paper/mobile format was rotated in the test so that half the participants started with paper and the other half with mobile (no adjustments). The rotating starting order aimed at balancing potential bias that might develop on the basis of what the participant tried in the test first.

> *I'm familiar with the format. It's easy to read because of the large paper size.*
>
> *I am used to completing tasks on paper and am familiar with comprehension tasks on paper from english exams*
>
> *I think I found the paper format of the test easier to complete because that is what I'm used to reading from. For my university work I always print out journals rather than reading them off a screen as it seems easier for me to read.*
>
> *Because I'm more used to reading this sort of text on paper.*
>
> *It is more natural. It allows for easier scanning of the page, which makes identifying the missing word easier*

Importance of overview/preview and navigation:

> *Because I had the whole text on the page and I could go back and read with out flicking through the screens.*
>
> *Some of the missing words were hard to fill in but being able to look at all of the extracts as a whole made it easier to asses how I was doing and gave me a clearer idea of what words were fitting in smoothly and which weren't.*
>
> *I could read all the text at once at quickly look back over previous sentences easily.*
>
> *It was easier to read as I had a preview of the text to follow. Also allows scanning and skim reading.*
>
> *Because I had the whole text on the page and I could go back and read with out flicking through the screens.*
>
> *I could read all the text at once at quickly look back over previous sentences easily.*

Additional, smaller, group headings that emerged from the comments were **interaction** (namely, no need for much in paper compared to mobile) and **layout**. Participants commented that overall, paper was easier on the eye and elements such as clear margins around the text made the viewing experience better. The need to interact with the mobile device for turning the page, as it were, quickly became an irritant for some of the participants.

On the whole, there were three comments that preferred mobile over paper, with all three participants finding excessive overview distracting. In the mobile device comments there was also a small group of participants who found the limited interface size (and thus limited amount of text visible at once) helpful for completing the test.

5.2.3.1.2 Text on mobile device (no adjustments)

In participant comments on reading on the mobile device before they were allowed to make adjustments, there was a wider array of topics with some

overlap. Also, the comments did not fall so strictly into only negatives or positives per category.

As was perhaps predictable, the topic of overview was a dominant one. The majority of comments under this category lamented the lack of overview and the difficulty of preview and easy backtracking within text. Many comments also pointed out that it was difficult to navigate back to an exact position within the text because the passages straddled over so many pages on the mobile screen. Related and partly overlapping with the overview issue was that of content fragmentations: having to turn pages (horizontal swipe across the screen in the Stanza application user interface) very frequently appeared to break the reading and task performance flow.

Overview and navigation:

> *Without being able to see the text as a whole it was much harder to refer back to specific points and consider the missing words in context.*
>
> *It wasnt as fluent as the paper format and I found myself flicking back and forth more often to make sense of the sentence before I could fill in the gaps.*
>
> *It was still fairly easy to read on the phone but because of the fewer amount of words I could only see at a time, it made it less fluent, and slowed down reading speed.*
>
> *It was easy enough to read, but when I wanted to go back and read when I did not understand the sentence completely it was harder to find where I could clear up the ambiguity as I had to flick around the screen.*
>
> *[mobile was] less fluent as sometimes I had to turn the page and then go back to make sure I had gathered the correct meaning of the sentence which allowed me to select the appropriate word, obviously on paper this was not the case.*

Together with comments that essentially pointed to overview and/or preview and backtracking possibilities, the interaction aspect was raised as well: having to turn pages.

Navigation:

> *only less fluent as sometimes I had to turn the page and then go back to make sure I had gathered the correct meaning of the sentence which allowed me to select the appropriate word, obviously on paper this was not the case.*

> *Less fluent than paper because turning the page is required and screen goes dark after a while too.*

> *The small screen meant my concentration was broken when changing pages. It was more distracting, going back and forth.*

> *I have the same phone so am familiar with how to use it. It was less fluent than paper because I had to swipe across to read the next page*

More than with the participant comments on reading on paper, there were a number of comments pointing out that reading on mobiles was an everyday activity these days and there were no problems with either lack of overview or fragmentation of content. These comments were categorised under **familiarity and convention**:

> *i am very used to using my own mobile all the time, so i am used to the smaller screen and adapting how i read on it. The text itself and the stories it told also seemed to make more sense.*

> *I found both were easier as the passages were easier to flick back to where I needed to look to fill in the missing words.*

> *im used to using mobile to read*

Many participants, however, also commented on being not so fluent or familiar with reading on the mobile:

> *I am less used to reading in this format. I had to move the screens forward at times to understand some of the sentences.*

> *Again I am used to reading things on a mobile phone but not completing work or a comphrehension on it*

Further categories emerging from the body of comments were legibility and layout. Whilst, again, some participants were happy with the way text is displayed on mobile phone screens owing to their long experience with the device and in some cases happening to own the same device as was used in the test, it was also noted that text for many users felt too cramped and tight on the screen, and the font size seemed too small. With the latter it should be noted that in the test setup the mobile device font size was set to be as close as possible to the print size on the paper version.

Legibility and layout:

> *The writing was a bit cramped and small and reading off of a screen is a bit more strainful to my eyes than reading from paper.*
>
> *The text was quite small and the gaps between the lines were small. The reading wasn't very fluent as you had to stop to change screens often and also sometimes it was necessary to go back a page to reread what had come before.*
>
> *On the phone the text came in smaller chunks so it wasn't as fluent because I had to keep changing the page to read the next bit.*
>
> *i found it more difficult than paper as on a mobile the words seem more clumped together making it harder to read.*

One category that did not come up from the comments on text on paper was that of the content, that is, the topic of the text sets given to participants to complete on the mobile. Some five comments out of the 45 mentioned that basically there was nothing wrong with the reading, but that somehow the topics were perhaps harder on the mobile version.

5.2.3.1.3 Impact of adjustments on mobile screen reading user experience

Five distinct categories of comments stood out when participant comments were analysed on ease and fluency of reading and completing the test on the mobile after they had been allowed to adjust the viewing parameters:

- Identifying more text per screen as a source of improvement in user experience.

- Identifying that more text per line with landscape orientation of the device made for a better user experience (often coinciding with slightly larger text).
- Obtained sense of improvement in general in the reading experience and performance owing to the changes participants had made.
- Conclusion that adjustments improved the experience compared to the mobile device that had not been adjusted, but stating that the experience was still not quite good enough or not as good as reading on paper.
- Perception that alterations made no observable difference in the (perceived) user experience.

Identifying more text per screen as a source of improvement in user experience.

The comments under this category came in two guises: either participants felt that getting more text per screen improved their performance and experience in a fairly straightforward fashion, or they acknowledged that although there was perhaps improvement, there was still too much need to keep turning pages.

> *The text/font changed helped, but the size limitations of the screen meant I still couldn't read the text as a whole. Kept moving back and forth.*
>
> *The text itself was easier to read however you still had to stop to flip back and forth between pages as much as with the unmodified format*

Fluency of reading and completing the test seemed to be strongly linked to being able to see more text on the screen than in the earlier condition with the mobile device. More text allowed participants to see more text and keep sentences and passages more complete so that the backtracking and preview effect that was so well liked with the paper part of the experiment was regained to some degree. However, the trade-off was that there was a need to balance between finding a font size that was clearly legible even when the

device was held in a comfortable position or at a suitable reading distance and still having the expected amount of text on the screen visible at any one time.

> *I think it improved my ability to read it more easily. But it didnt really help with the fluency issues as i still only had a limited amount of text available at any one time. It was hard to find a balance between font size being readable enough to comfotably hold it at a reasonable distance and the amount of text available on one page.*
>
> *The text itself was easier to read however you still had to stop to flip back and forth between pages as much as with the unmodified format*
>
> *It was more fluent than the previous time but I still had to flick back and forth as some the sentences were incomplete on the page.*
>
> *Having to switch from page to page in order to gain enough context to be fill the gaps in the text was, clearly, still an irritant and a major reason why the mobile device even after adjustments did not work as well for the participants as paper version.*
>
> *More text per page so less interruptions to move pages.*
>
> *Because the margins were increased I could view more of the text at a time, which increased fluency, with fewer distractions turning pages. But there's more page turing than the paper format.*

Identifying that more text per line with landscape orientation of the device made for a better user experience (often coinciding with slightly larger text).

A number of participants turned the device into landscape orientation. On the test device this resulted in the text being presented in the selected orientation and also in a slightly larger font. In the comments, having longer lines of text was considered an improvement in general fluency and for better maintaining at least sentence level context. Some participants felt they did not have to turn the pages quite as often as in landscape.

> *I found it was quicker to read when the text was slightly larger as I felt I made less errors and it was easier to read as more text appeared horizontally.*
>
> *The words were large and clear, with little infomraton on the screen so I didn't try to miss bits of the text out. I am also used to having the*

screen landscape and so this probably made it easier for me as it was more usual.

Obtained sense of improvement in general in the reading experience and performance owing to the changes participants had made.

Participants who changed the appearance and layout settings typically reported at the end of the questionnaire that the changes that they made improved the fluency of reading and/or legibility of the text:

i felt when i had modified teh phone it became more easier due the fonat begining slightly more of an easier font to ready and increasing the spacing between the lines made it more clearly to distunguish between the rows of text.

The format i chose had bigger font and was more spaced out which made it easier to read, although i found this comphrehension harder than the other 2.

I turned the phone sideways to make it easier to read

After the size of the font had been modified I found it easier to read and therefore allowed me to read more fluently and almost like i was reading a passage on paper.

I made the font size bigger which made it easier for me to read this is because the words looked more spaced out and alot less closer together - was easier to distinguish the words.

With the writing being a bit bigger and the sentences spaced more widley it was easier to process the words smoothly and find an appropriate word to fill the gaps.

In those comments stating that the few changes that were made improved the reading experience, the focus of improvements seemed to be strongly on legibility of text: larger text. These participants clearly favoured being able to see text without additional effort over the amount of text that was generally on the page at one time. Typically, in this category of comments, the appreciated features included being able to see words clearly when they were bigger and more clearly spaced out on the page. Also increasing the line spacing seemed to improve the viewing conditions for these participants.

From some of the comments in this category it was also possible to conclude that simply having the freedom to adjust the conditions could in itself be a factor in the reading experience: further research into the correlation between actual improvement in performance and the ability to adjust the viewing parameters would probably be helpful for user interface designers not only in the mobile device industry but more broadly as well in terms of finding the right balance between interface simplicity and functional settings architecture.

Conclusion that adjustments improved the experience compared to the mobile device that had not been adjusted, but adding that the experience was still not quite good enough or not as good as reading on paper.

In this category of comments improvements were acknowledged, but immediately counterbalanced with a note on how the change was not quite enough. Typically, a participant would notice that legibility of individual words or a piece of text would improve, but the overall fluency of reading did not: it was still not possible to see a functional amount of content on the screen in order to obtain a good grasp of the whole context of individual passages. In addition, the comparison to paper format was very much at the top of participants' minds and even after adjustments the mobile device screen simply did not feel as easy and fluent as the paper version of the texts.

> *This was a little easier because there was more text on the screen but still wasn't as fluent to read as the paper text.*
>
> *I benefitted from the modified screens, particularly the larger font. The text, however, seemed more complex than earlier examples and so my reading felt less fluent.*
>
> *After changing the font type and size as well as margins I was able to fit more text on screen without sacrificing legibility as much as with the unmodified version (still not as clear as paper). Cut lines and quick fade out were still a problem, however.*

Perception that alterations made no observable difference in the (perceived) user experience.

The final category of comments was the observation that the adjustments did not affect the reading experience. Again, some consideration was given to perhaps improving legibility, but the experience of ease and fluency was not improved enough to make a difference.

> *I felt that the layout of the text on the phone did not make much difference when completing the test essay.*
>
> *The font size was made bigger but I don't think this affected how easy the task became, it was very similar to the previous one. The white background was kept as it is easier to read things from a white backround as books, papers, articles etc are all on white backgrounds and we are used to this.*
>
> *I dont think changing the font changed the ease of reading ability.*
>
> *I dont think the modifications I made helped much. With more line spacing it was easier to read but then that meant there were less words on each page so it didnt help with the fluency of the text.*

One interesting additional, albeit small, group of comments that should be mentioned here concerns those participants who did favour mobile device screen over paper. The comments pointed out that for some participants the fragmentation of content was actually a positive element that improved their reading experience: breaking text down into small one-screen bits was experienced as less distracting and improved concentration on taking in the content that was actually there (the eye seeking backtracking or preview did not break concentration).

> *i found it easy to read as the writting was clear and broken down into chunks due to the small screen so no other distractors, like other chunks f text, were on the screen at the same time*

5.2.3.1.4 Preferred format (paper, mobile or adjusted mobile)

As was seen in the analysis of variance, a clear majority (62.2%) of participants selected paper format as the most fluent and easiest format for perceived ease and fluency in reading. Text on mobile with adjustments was favoured by 20% and text on mobile without any adjustments was selected as the favourite by 13.3%. A choice of “other” was also offered in the

questionnaire to see if participants had further experiences of options, such as e-readers, that they might favour.

The reasons for picking paper format were quite homogenous: participants were most used to reading on paper and it was the familiar format.

> *I'm most used to doing this kind of work on paper so find it easier, I don't read books online either.*
>
> *i am not brilliant wth technology. i am much happier with an old fashioned approach of pen and paper!*
>
> *It is easier for me to read as I am used to how it is presented and I am familiar with reading passages in this way and so feel more comfortable when reading off paper.*
>
> *I am more familar with it.*
>
> *I think this is because it is the format I am more familiar with.*

Participants who picked paper also pointed out that the wider view (overview) of the text was a factor: having to interrupt reading to flick through the pages or not seeing full sentences or phrases at once tipped the scales to paper instead of the more limited viewing experience on the mobile.

> *I think it's because I am used to it and because all of the text fits onto the page so not broken into sections by having to turn the page frequently.*
>
> *Everything fit onto one page so it was more fluent to read, rather than swiping across to read the next page like on the phone*
>
> *It was the easiest to read and the most fluent are there is ore text on the page it is much easier to scan through it than to flip back and forth between the pages.*

The reasons for picking adjusted mobile format tended to focus on appreciating user control over the interface:

> *I had more control over the device, and the layout of the text making it easier to read and take in information.*
>
> *I think that being able to personalise the screen - font size and background colour specifically - helps speed up my reading. In the*

real world, I prefer reading on screens as I am able to customise these features.

i am more used to reading off electronic equipent than off paper as most of my communication is via text and all work i complete is on a computer

you are able to change the font to how you are more used to seeing text in everyday circumanstances and the way in which you rmore familar in reading. i feel being able to personalise the phone and increasing the spacing made the experience better i was able to more clearly distinguish between the words and the different lines.

The relatively small number of participants who opted to keep with the basic mobile without any personalized adjustments simply acknowledged that what was presented to them on the mobile was clear and functional enough. Some of these participants did not bother with adjustments at all in the test situation.

i just felt that the stories were much more fluent, and made more sense to me. i found the software used to present the text like a book, and so when i had to change the page, it was like flicking over to a new page in the book. this could be due to the fact i am pretty much on my iphone 24 hours of a day anyway, and thus i am very used to reading long text from my phone.

its what im used to

There are fewer words at once, which seemed to hold my attention much better than on a printed page.

It was simple and easy to use, and im used to using this format.

A further study would be required to establish how many end-users would actually adjust any settings on their electronic devices if this had not been a test situation where they were specifically offered a chance to do so. The researcher's personal experience from nearly ten years of usability testing mobile telephones globally is that typically in mobile phone concept studies and usability tests, when asked, end-users say that they like the possibility of personalizing their devices and adjusting certain settings, but if one were to take a peek at their phones there and then, the only adjustments would typically be a specific background image and set of shortcuts (if applicable to their phone). Many people simply do not change any default options or

settings in various systems or interfaces even when doing so would lead to obvious benefits (Samuelson and Zeckhauser 1988, in Thaler and Sunstein 2008).

5.3 Discussion

In the original hypothesis based on interview and survey data on mobile device users' experiences and expectations, a significant difference in performance scores was expected. However, there was no significant difference in the overall number of correct answers between the conditions. The significant differences between the conditions were observed in time on task and in the perceived ease and fluency: completing the tasks on a mobile device took longer than on paper and significantly longer than on an adjusted mobile. The perceived preferences and observations on ease and fluency still supported paper as the easiest, most fluent and most preferred format. These open-ended comments on ease, fluency and preferred format echo the attitudes and expectations recorded in the interview study that produced the mobile reading acceptance model (Chapter 9).

These results, then, prompt several questions. If the expectation based on user interviews so strongly points to a significant difference in performance, why is the performance so uniform throughout the test conditions? Are end-users so accustomed to mobile interfaces for their communications and access to various content that at a practical level the interface and format really make little difference? Or was the administered test too easy, and are the results so similar throughout the conditions because of ceiling effect?

In terms of correct answers, then, performance did not show significant differences whereas time on task did show some significant differences. Further experiments would be required to assess whether the narrow performance margins with correct answers scores is due to ceiling effect or if the participants' ability to complete the test is due to a level of familiarity with mobile devices combined with the lack of time restriction in the test set-up. A

study into reading behaviour and reading devices by Grzeschik et al. (2011) looked into various electronic reading devices (conventional LCD screens, a smartphone and a dedicated e-reader) and compared the reading rates to the rates when reading books or printed pages. The study found that within the relatively small sample of participants in the test, the reading rates did not suffer from the reading device – which was not the expectation and the researchers' view of common opinion on the matter. Grzeschik et al. (2011) concluded that it was more likely that differences in reading rates between reading devices were due to individual reading behaviour as well as to what type of text was in question (scholarly or otherwise). The interview study on attitudes and expectations pointed to the importance of internal motivation: the reason for reading. Also, the present experiment imposed external motivation onto the participants: they were expected to read carefully enough to be able to complete the test. In the light of these conclusions, it is not, perhaps, surprising that participants did perform close to the same high level in terms of correct answers.

In the comments participants gave on their preferences when rating ease, fluency and favourite format one category emerged clearly above others: **convention and familiarity**. Not only was paper format given as the most fluent/easiest/best of the three but when participants picked one of the mobile conditions, the typical reason was their familiarity with mobile devices. Typically, participants said they were used to paper for most reading, but deep familiarity with mobile devices was not really surprising either: in the self-assessment questions which participants answered prior to the comprehension test, the participants described themselves as having above-average computing skills and a clear majority were already using phones that can be classified as 'smartphones' (phones that have internet connectivity for web surfing, e-mailing, and other more advanced applications). The mobile phone ownership profile of the interview group in the mobile reading acceptance study was different: these participants had predominantly low- to mid-price basic phones with either no or limited internet connectivity. It would, therefore, make for an interesting further study to explore what (if any)

changes in perceived attitudes and expectations to mobile device reading would be observed if a similar interview study was carried out with a participant sample who were mostly already using smartphones.

Another aspect of perceived fluency and ease where the scales tipped towards paper format was the format facilitating **easy overview and navigation** within the content. In the comments in the context of paper, participants highlighted the ease of overview and within-content navigation, whereas with mobile conditions there were observations that the format did not support these functions so well. A mobile web browsing usability study comparing mobile browsing and desktop browsing found that participants' main concerns were to do with poor ability locating specific content on a page that only showed a small fragment of the entire content at one time (Shrestha, 2007). Interaction was another element perceived in the same manner: paper is easy to handle and moving within content that was mostly presented on a single sheet was easier than having to flick from page to page on an interface that showed only a small fragment of the content at a time. Yet the test scores in terms of correct answers did not show any significant difference, which suggests that perhaps the time on task difference could be explained by the additional need to navigate and interact physically with the test material on the mobile, which was the first of the two mobile conditions participants encountered. Adjusted mobile was the last condition and may have benefited from participants becoming familiarized with the text procedure and with the test device.

Considering that 10-20% of reading time is spent on regressions (eye moving backwards saccades during reading) (Ellis, 1984), it is not hard to hypothesize that on a small mobile display that displays a fraction of the words per page compared to a conventional printed sheet the regressions are interrupted more often by page breaks. Hyönä (1993) also found evidence of the integrative nature of regression on comprehension: in non-primed setting with unpredictable words fixations regressed particularly when the reader reached a sentence boundary. Furthermore, moving from easier passages to more

difficult passages, not only do the eye fixations on a line tend to be longer, but the number of regressions also increases (ibid.). The test material in the present experiment was designed to progress from relatively easy passages to more advanced ones and the last part of each test set would be likely to demand an increased number of regressions.

Eye fixations do not only jump backwards, but the eye previews up-coming content as well during reading, picking up information just to the right of the fixation point (parafoveal preview benefit) (Rayner et al., 2011). The parafoveal pre-processing plays a significant role in decisions on skipping familiar and predictable words in the reading process (ibid.). In the same way as frequent page change could affect regression, the preview benefit could also be compromised. In short, frequent page changes interrupt the processes that enable fluent continuous reading.

Dillon, et al. (1990) compared reading comprehension on smaller and larger screen areas, looking into the effect on comprehension when a sentence was split over a page change. The present experiment seems to have arrived at many of the same conclusions as Dillon et al. in finding that, without time restriction, there was no major effect on comprehension between the various test conditions, but participants' perceived preference was still clear: a larger screen area was deemed better. Earlier in this chapter the impact of split or fragmented content and the need for (distracting) interaction with smaller interfaces was referred to as grounds for hypothesising that there would be a significant difference in reading comprehension performance between paper and mobile device screen. This was not the case, as concluded, but it is still important to acknowledge that regardless of performance, participants were still able to differentiate between the experiences.

One of the most interesting observations in the time on task results is that there is a significant difference between mobile and adjusted mobile conditions, and paper format falls in between these two. Admittedly, the mean for time on task with paper is closer to adjusted mobile than mobile, but the

difference in either direction is not statistically significant. Looking at the comments from participants who selected adjusted mobile as their preferred reading format, one category rises above the others: preferring the freedom to make the adjustments. The impact of these adjustments, judging from the comments on ease and fluency ratings, came through maximising either the amount of text per screen and accepting a trade-off of font size related legibility or having longer rows of text (landscape mode), which still allowed clearly legible font size.

Bernard et al. (2003) found that although performance-wise there were no significant differences in reading speed and accuracy between the tested typefaces (size, type and format), participants were still able to point to a relatively clear perceived favourite. The trade-off that was seen in the reading comprehension test when participants adjusted the screen suggest that participants were well aware of improved local legibility of words if they were to use a larger typeface, but also that they would lose some continuity and sentence context if the typeface was too large.

With no significant difference in performance in terms of correct answers, the present experiment echoes the results from the study by Bernard et al. insofar as the measurable differences in performance as an outcome of changes in text appearance. Bernard et al. (ibid.) refer to Tinker (1963) and others maintaining that varying typeface and size of text within common parameters does not tend to have any significant effect on its readability. The possibility of ceiling effect in both the present experiment and in the study by Bernard et al. (ibid.) can be overcome by means of, perhaps, a time restriction on the tasks to see if a finer grain of difference in performance can be detected. The fact that perceived preference is expressed by participants as clearly as it was, suggests that there is, indeed, a finer grain in the task performance that should be studied. Furthermore, additional investigation into the effect of specific adjustments vs. simply knowing that adjustments are possible would be needed in order to establish if the perceived forced format significantly affects the perceived user experience of reading on a small display.

Mobile (either adjusted or plain) preference was also explained in terms of familiarity and convention, as discussed earlier. Pairing the familiarity aspect with the freedom to make adjustments might well be a combination that allows mobile reading to be perceived to be as fluent and functional as reading on paper. Would further experimentation, then, using a sample of participants specifically screened to include end-users who are already used to reading on their mobiles (advanced mobile readers who regularly use websites, e-mail, e-books, etc. on their phones) indicate familiarity and convention as the key element for positive perceived mobile reading attitude and experience? Of course, it must be remembered that in the test set-up in the present experiment the adjusted mobile condition was always the last condition of the three to be tested, and (as pointed out previously) the possibility cannot be excluded that the improved time on task in the adjusted mobile condition was due to the practice that participants gained from the first two conditions. Further experimentation with alternative condition rotations would be needed to examine the effect of condition order upon test scores.

The open-ended comments on the conditions and their ease and fluency are in line with the interview study on the topic and therefore perhaps the most interesting result from the present experiment is the degree of influence of perceived fluency and ease over actual performance on overall user experience. In other words, while performance on the primary tasks remained steady throughout the conditions, users still perceived noticeable differences in user experience. In terms of designing user interfaces for reading and designing textual content for mobile devices user acceptance of the interface is likely to be determined by user attitudes prior to using a given interface, attitudes and perceptions regarding reading on mobile devices, as well as the perceived fluency of interacting with the device and the user interface. Most of the comments on ease and fluency highlighted the irritation and disruptiveness caused by having to flick through pages too frequently and lacking the overview ability when only a small amount of text was visible at any one given time on a mobile device – even after adjustments had been made. The

participants also appreciated having control over the interface to obtain the viewing conditions they personally preferred, and regardless of whether the adjustments produced actual improvements to the test scores or not, these participants typically reported improved user experience.

Legibility (in the form of text size) was seen as one of the issues with reading on the small screen in the present experiment. The following chapter explores the experience of legibility in a word-search experiment, comparing subjective experience with task performance.

CHAPTER 6

Which presents a legibility experiment into word recognition on a mobile device screen, and how various light conditions affected user experience.

Word search on small display: legibility in mobile context

In the reading comprehension test presented in the previous chapter, participants pointed to having to compromise between the amount of text on a mobile screen and legible text size. The present chapter describes and discusses the design, administration and results of an experiment in legibility on a mobile phone screen when the viewing conditions were less than ideal. Mobile devices are used under a variety of external contexts and, thus, various light conditions. This means that the screens are viewed in complete darkness and bright sunlight and anything in between, both indoors and out and on the move. The interface on the devices will also be made up of various graphical elements and patterns. The patterns, selected colour themes and the overall design of interface components may not always be designed with set content in mind, and there is therefore potential for conflict in terms of colour contrasts and such. In the experiment the viewing conditions (reflections on the screen, different designs of background "wallpaper") were simulated. Legibility is an important part of the overall reading experience, as seen in the responses in the experiments presented in chapter 5, and it is therefore necessary to explore how various circumstances affect legibility as such, and also how possible problems with legibility affect the reading experience.

For example, Bernard et al. (2003) compared the effects of text size and format on computer displays and found that although performance rates did

not differ significantly when using typical, everyday typefaces and text sizes, test participants were still able to identify what their preferences were at reading experience level. Such a difference between performance and perception was also visible in the reading comprehension experiment described in Chapter 5. Further to that, legibility and readability of content on a mobile screen were elements respondents in a reading attitudes and expectations interview (Chapter 9) and in a survey on the same theme (Chapter 10) brought up as something that had a strong influence on the likelihood of their wanting to access text format information on a mobile device with a small screen.

Studies into contrast and luminance in reading have found factors that can help designers predict legibility of text content. Sharff and Ahumada (2000) studied the combined effects of textured backgrounds and text contrast, finding that while both had some effect on readability (measured in a search task), the effect was most notable when the background texture frequency was high and text contrast relatively low. Mills and Weldon (1987) illustrated early on how readability on computer screens is dependent upon a number of factors, such as contrasts, font styles and overall resolution of the computer screen. On the whole, a considerable body of research has been undertaken to establish optimal viewing and legibility conditions, such as Legge and Rubin (1986) on the role of contrast, Legge, et al. (1987) on wavelength effects in vision, Pastoor (1990) on legibility and subjective preference in terms of colour contrasts in text, and Dyson (2004) on the effects of physical layout when reading from a screen.

Cattell (as early as 1886) observed that word recognition (at least for shorter words) was faster than recognition of individual letters. He explains this to be due to words being used more frequently than individual letters and thus being a more familiar unit. Cattell touched on what has later been recognized, through further experimentation by Reicher (1969) and Wheeler (1970), as Word Superiority Phenomenon (WSP). The WSP supposes that real,

meaningful, words have some element that supports their correct, immediate recognition, unlike nonsense words or isolated characters.

The present experiment, part of an experiment that aimed at exploring the possibilities of using an optical character recognition (OCR) system in legibility assessment of various graphical user interfaces (GUIs) (Colley et al., 2012), analyses the participant and OCR performance data and the perceived legibility ratings for the purpose of examining whether there is a meaningful correlation between performance and perception. The author of the present thesis was a member of the research group for the study into OCR use, focusing on participant experiment planning, execution and the variance analysis.

The present experiment analysis, therefore, aims at finding out the correlations between performance and perception throughout the test conditions of background type and light effect (reflection). The hypothesis for the experiment was that task performance and perceived legibility ratings would correlate positively. In other words, changes in the visibility and legibility of the words on the mobile device screen would also be seen in the subjective assessments of legibility. Should such a positive correlation be true, subjective assessment (perceived legibility rating) on its own could be used in interface design testing as an indicator of the functional legibility level.

6.1 Method

Participants were asked to count the occurrence of one particular word from a collection of words on specifically designed screenshots of mobile device screens (three sets of 16 images, which show the independent variables of background image and surface reflection). The recognition score from the OCR system was also included in the correlation analysis. After each image, participants were asked to state how many times the specified target word occurred in the image and then to rate the legibility of the image (screenshot) on a Likert type scale from one to seven, where one represented extremely

poor legibility and seven represented extremely good visibility. The correct count of target word from both the OCR system and test participants and the perceived legibility scores from the participants provided the dependent variables in the experiment. The experiment focused on background and reflection as the examined variables because light is a considerable disruption in visibility with all electronic displays from digital cameras to laptop computers and because on mobile interfaces it is not always possible to control the background graphics the text content appears on.

Experiments with shapes of words suggest that the Word Superiority Effect is also present when nonsense words have the same Bouma shape (the outline shape of a word or a cluster of letters) (see Figure 11) as a meaningful word (Haber and Schindler, 1981; Monk and Hulme, 1983). Leading on from such observations and theories suggesting that word recognition does, indeed, typically happen at least partially on the basis of word shape, the present experiment was designed to present participants with lists of words where the words all had the same or close to the same word (Bouma) shape. The purpose behind the similarity in shape was to ensure that participants had to look at (and read) all the words in the sets instead of jumping from word to word on the basis of shape alone.

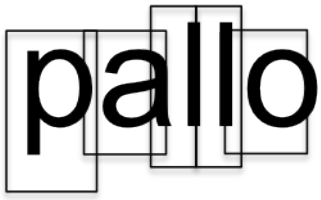


Figure 11 Bouma shape of a word: the outline shape of a cluster of characters.

As the focus of the present analysis of the experiment lay in the correlation of performance scores and perceived ratings, a quantitative approach was selected. The data analysis does not extend to calculating relative effectiveness (Paas and van Merrienboer, 1993, 1994), as the images used in the test were not designs aiming at finding the most effective colours, contrast, or other graphical elements, but rather to examine the effect that various levels

of graphical distortion and obscuring have on the legibility of words. It should be emphasized that the test setting was artificial in that the tests were done in a laboratory and the light effects used on the mobile phone screens were simulated by means of graphics. The experiment aimed at removing such variables as moving or shading the phone when light distracted seeing the screen so that a fixed comparison could be made with the OCR system. For fuller ecological validity, a similar test would have to be repeated in realistic conditions, but controlling the independent variable would then be extremely difficult.

The participants were randomly assigned to one of three groups of ten participants in this between-subjects design and the test images (three sets of 16 screenshots that presented the background and light conditions) were rotated between the groups (Table 10). In addition, the order of presenting the images in the test was rotated so that half of each group started the image viewing from one point in the series, and the other half from another point. The images are described in more detail in section 6.1.2 (Materials) of the present chapter.

Analysis of variance was conducted on the performance scores (correct count of target word) and on perceived legibility ratings, and the two sets of scores were checked for correlation. In the analysis in Colley et al. (ibid.) correlation was also checked between participant scores and the OCR system.

Table 10 Image set rotation in word search experiment.

Group/image set	Image set 1	Image set 2	Image set 3
Group 1A	a – h, i - p	a – h, i - p	a – h, i - p
Group 1B	i – p, a - h	i – p, a - h	i – p, a - h
Group 2A	a – h, i - p	a – h, i - p	a – h, i - p
Group 2B	i – p, a - h	i – p, a - h	i – p, a - h
Group 3A	a – h, i - p	a – h, i - p	a – h, i - p
Group 3B	i – p, a - h	i – p, a - h	i – p, a - h

6.1.1 Participants

Thirty participants were tested in one-on-one sessions. The participants were all professionals with years of experience of various mobile phones. The age range was 27-42 with 14 female and 16 male participants. All participants declared normal or corrected-to-normal eyesight. The participants were all employees of a telecommunications company in Finland from both technical and non-technical positions and were native Finnish speakers.

6.1.2 Materials

Three sets of 16 images were created for the experiment (Colley et al., 2012). In all, sets were made of a series of background and foreground graphics, which simulated viewing conditions with or without background “wallpaper”, black or white text, and two types of simulated light conditions. The sets all produced the graphics variants, and differed from each other in the number and position of the target words that participants were asked to count. Figure 12 illustrates a 16-image set, arranged according to background and light conditions. The purpose of having three sets was to ensure randomization.

Figure 12 Test images in word search experiment. From top left D L E M; Second row: B J G O; Third row: H P A I, Bottom row: F N C K. The figure is reproduced here with permission from Colley, Huhtala, Tikka and Häkkilä, 2012.

Table 11 Image type summary.

	No light effect	Light effect
Plain background	D, L	E, M, B, J, G, O
Wallpaper background	H,P	A, I, F, N, C, K

The analysis was interested in the effects of background and light conditions. Table 11 summarises the images used in the experiment under the examined categories of reflection and background type. The plain background images have a 50% grey as the background and the text is white and black. The wallpaper background is an example of a gradient wallpaper that can be found on mobile phones. The light conditions simulated reflection, cloud and bright ambient light (Figure 12). Figure 13 shows a test image in an actual device in

the test situation. The Latin squares rotation matrix for administering the test images to the participant groups is illustrated in Table 10.

Figure 13 Test text image on a device.

Test participants were given the task of counting the number of times a set word appeared in the image. All the words were in lower case and mostly resembled each other in their Bouma shape. A word list was used instead of full sentences so that participants could not use their general knowledge of language to guess the occurrences of the given target word. The latter concern is familiar from Cloze procedure tests on comprehension (Gilliland, 1972). The word participants were asked to count was ‘pallo’ while the other words on the screen, for example ‘pulla’, ‘pöllö’, ‘valli’, etc. are all real words in the Finnish language, and all test participants were native Finnish speakers.

A time limit was set on the basis of pilot tests where pilot users counted the target word from an image that did not have any visual distortions or contrast gradations. The users were instructed to try and read at their natural pace, though the test premise fully acknowledged that being in a test situation with a stopwatch would affect the reading pace somewhat. While being aware of an overall time limit discouraged the participants from slowing their reading pace too much, knowing that they were not being timed for how quickly they performed the task made for no additional anxiety or nervousness and a close-

to-natural reading pace was feasible. Participants' individual approaches to tasks showed various task strategies: where one participant would simply read the text on the screen line by line and be happy with the result, another would read through and count and then use the remaining time to check the result before declaring it to the researcher. To time the reading task, a slightly different test protocol and instructions would be required.

6.1.3 Procedure

A series of words were displayed as pictures (30 words per picture, 16 pictures) with varying backgrounds and visual effects simulating various light conditions (see Figure 12).

In the test session, participants were initially introduced to the task and then given a warm-up task in order to familiarize them with the test procedure and task type and thus minimize the effect of surprise and novelty on the performance. The tests were carried out in the participants' native language and the words on the test images were in their native language as well. Test images were handed out to the participants by the moderator one at a time, presented on a mobile phone screen.

Upon completing their count, participants stated their answer, followed by an immediate assessment of the perceived legibility. Participants were shown a card with the assessment scale to help with the assessment. The scale, 1-7 was a Likert-type scale where 1 represented extremely poor legibility and 7 represented a level of legibility where participants felt there were no problems in seeing and recognizing the words.

6.2 Results

The images used in the experiment fall into four groups:

1) Plain background without light.
2) Plain background with light.
3) Wallpaper background without light.

4) Wallpaper background with light.

Variance analysis was carried out with both performance data and subjective ratings data. The analysis uses a .05 significance level throughout.

6.2.1 Analysis of Variance

6.2.1.1 Word recognition performance

The experiment was interested in the main effects of light and background type on the performance on word recognition as well as the interaction effect of light and background type. The experiment was a two-by-two analysis of variance, where there were two background types and two light conditions (Table 12).

Table 12 2x2 ANOVA design of the study.

	No light	Light
Plain background		
Wallpaper background		

There was no significant main effect of no light on task performance, Wilks' Lambda = .991, F(1,29)= .251, p= .620; light on test images (M=71.11) did not lead to a significantly different performance from when no light was visible (M=71.67).

There was no significant main effect of background type on task performance, Wilks' Lambda= .999, F(1,29)=.021, p= .886; wallpaper type background (M=72.22) did not lead to a significantly different performance level from plain type background (M=70.56).

There was no significant interaction effect of light use and background type, Wilks' Lambda= .994, F(1,29)= .165, p= .687; use of light did not lead to a significantly different result on plain type background from that on wallpaper type background.

Table 13 Means and standard deviations for participant task performance.

	No Light	**Light**	**Overall**
Plain	71.67 (31.303)	69.44 (33.945)	70.56
Wallpaper	71.67 (23.194)	72.78 (25.702)	72.22
Overall	71.67	71.11	71.39

6.2.1.2 Legibility ratings

The experiment was interested in the main effects of light effect and background type on the subjective rating of legibility, as well as the interaction effect of light effect and background type. The experiment was a two-by-two analysis of variance, where there were two background types and two light conditions (Table 14).

Table 14 2x2 ANOVA design of the study.

	No light	Light
Plain background		
Wallpaper background		

There was a significant main effect of light on subjective legibility rating, Wilks' Lambda= .123, F(1,29)= 206.590, p= .000; Light on test images (M=3.5) led to significantly lower ratings than when no light was visible (M=4.7).

There was a significant main effect of background type on the subjective legibility rating, Wilks' Lambda= .165, F(1,29)=147.289, p=.000; ratings were significantly higher for plain background type (M=5.0) than for wallpaper type background (M=3.2).

There was a significant interaction effect of light effect and background type, Wilks' Lambda= .825, F(1,29)= 6.150, p= .019, reflecting that the decline in

subjective ratings from no light to light was more pronounced on plain background type than on wallpaper background type.

Table 15 Means and standard deviations for perceived legibility ratings.

	No Light	Light	Overall
Plain	5.7 (.867)	4.4 (.865)	5.0
Wallpaper	3.6 (1.096)	2.7 (.707)	3.2
Overall	4.7	3.5	4.1

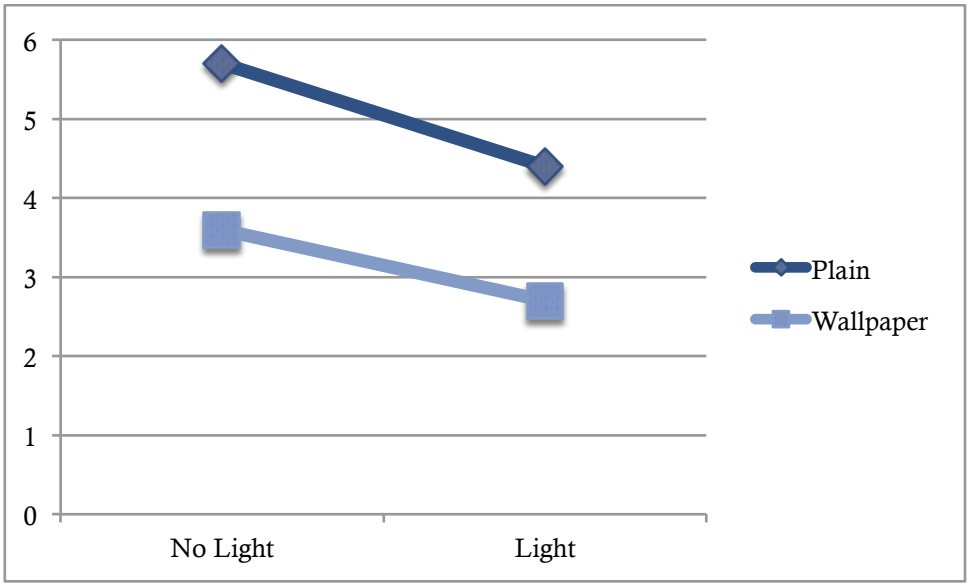


Figure 14 Plot diagram of means indicating that light has a similar effect under both conditions (plain background and wallpaper background), and test participants rated the images with no light higher than those with a light. The means for the legibility ratings are generally higher for plain background images than for wallpaper background images.

6.2.2 Participant results vs. OCR results

Figure 15 (line diagram) illustrates the performance of the OCR system, participant performance and participant perceived legibility rating (converted to percentage value to ease comparison) when grouped by test image type. Participant performance (the percentage of correctly counted occurrences of the searched-for word) remains close to the same level from group to group,

but the OCR performance levels and perceived legibility levels suggest a degree of positive correlation.

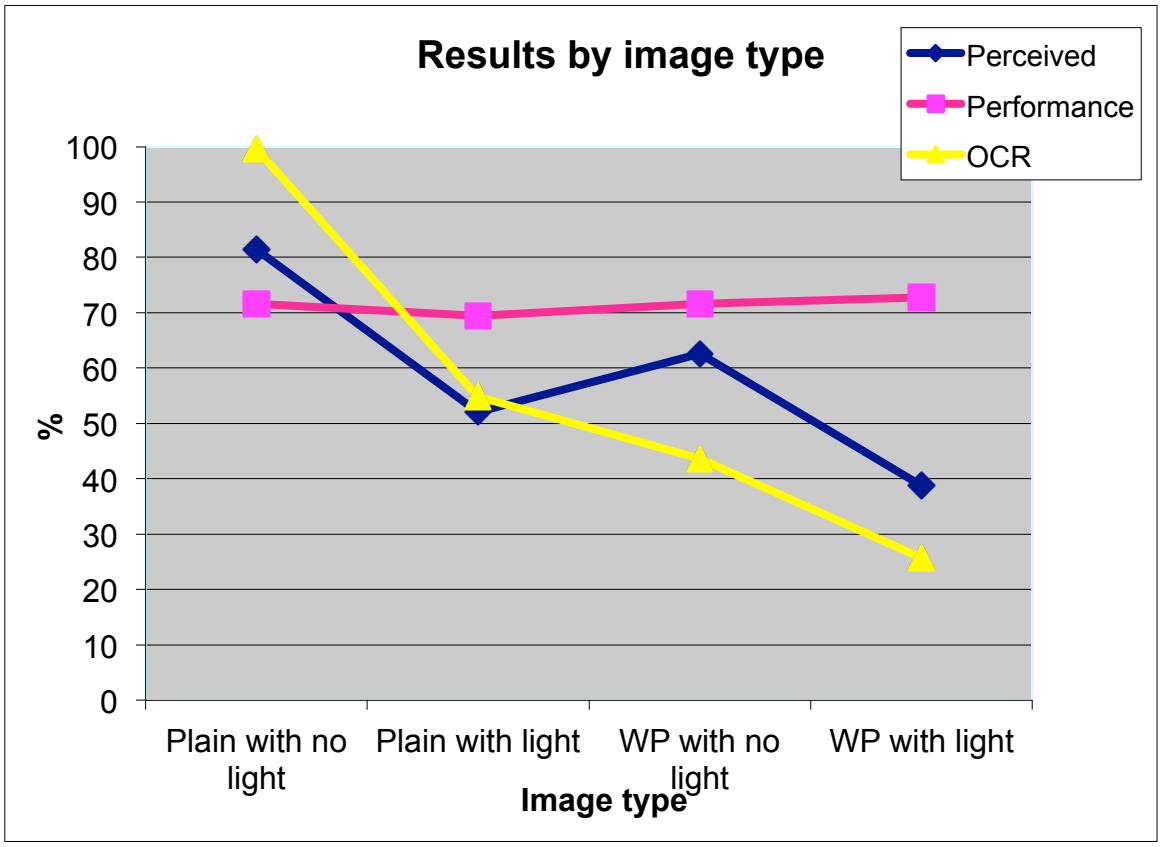


Figure 15 Line diagram of recognition results and perceived legibility per image type. 'Plain' refers to single-colour background, 'WP' refers to Wallpaper background, 'no effect' means no light effect was used, and 'effect' means light effect was used.

Plain background type with no light effect present produced a nearly 100% correctly recognized words result with the OCR system, and the same image type was rated as most legible by test participants. The perceived legibility means of the rest of the image types (plain background with light effect, wallpaper background without light effect, wallpaper background with light effect) reflect the performance level of the OCR system as well.

Participant performance compared to OCR performance level in the four categories pointed to a very small linear relationship (r= -0.235), whereas

comparison of perceived legibility ratings to OCR performance indicated a strong positive correlation (r= .913). Correlations were calculated using Pearson correlation coefficient. Colley et al. also considered the differences between images where the text was white and ones where the text was black, drawing conclusions in much the same direction: perceived legibility and OCR performance tended to correlate positively.

6.3 Discussion

Perceived quality of interfaces and interactions is still a very important axis in the overall evaluation of user experience. Even when performance results do not indicate the relative quality of interfaces or systems, there are still differences in the perceived quality. There are times when task performance (speed, accuracy, efficiency, etc.) and perceived quality of an interface do not seem to match, and this perceived quality often remains an important factor when consumers are deciding which devices or services they are willing to use. (Bernard et al., 2003; Grzeschick et al., 2011, Laarni, 2002; Dillon et al., 1990).

Information accessibility, through any medium, involves a number of factors affecting its overall fluency and efficiency. Legibility is one such factor in the reading process, not least in the mobile device context because of challenges placed upon it by the small interface size. The development process of content on small screens involves the interface design of a device as well as the information design of the content. These processes also contain assessment of quality and user experience through, for example, usability testing. The testing can focus on performance, on end-use feedback or, in best cases, on both. What is important, however, is to understand how far and in which aspects the performance and the perception support each other: is performance data fine-grained enough in itself to guide device, user interface and content design iterations; does end-user feedback (perceived experience) add value to the performance data?

The experiment by Colley et al. (2012) and the variance analysis repeated in the present analysis both showed that the simulated light conditions affected human participant performance very little. The OCR system's word recognition levels varied according to image type, as did the subjective ratings on legibility (Colley et al., 2012). The small effect on human participant performance shows that the human eye is superior to an OCR system in recognition even under difficult conditions. The perceived levels can differ considerably and concluding from the perceived legibility ratings, merely being able to see accurately and clearly enough does not mean that the text presentation would be deemed legible or optimal by human readers.

Being able to evaluate quality of interfaces or content even when measuring performance objectively is of the highest importance to interface and content design organisations. The results from the experiment presented in this paper as well as the research into and use of CAPTCHA images strongly suggest that human perception is highly capable of seeing and reading text even under difficult circumstances. Yet, as the present experiment indicates, the perceived user experience of these circumstances shows that the differences in quality are obvious to the end-users. Furthermore, the perceived quality tended to go hand in hand with the performance level of an OCR system. Based on these observations, it would appear feasible to use OCR systems as a fast evaluation tool for gauging perceived legibility for example in early design iteration phases. Further research on the topic should include more ecologically valid environments and contexts of use in terms of learning more about the visual distraction limits within which the human eye can read text on small screens efficiently in various light conditions and screen graphics contrasts. Because of the challenges of arranging such testing (controlling naturally occurring variables such as light conditions and context of use for each participant), design and device industries would benefit from methods where the conditions can be simulated at a reasonable proximity to ecologically valid situations so that early testing can be done at fast and relatively inexpensive cycles.

In the previous two chapters visibility and legibility elements of reading on small screens have been clearly present in the experiment results: particularly in terms of experience. The following chapter explores the visibility issues through observing end-user behaviour when test participants were asked to adjust visibility settings on a mobile device to their own preferences.

CHAPTER 7

Where text and layout adjustments made to electronic text on a small screen are analysed and discussed.

Appearance and layout adjustments: optimizing reading on a mobile device

The present chapter describes the data and analysis of adjustments that participants made to a mobile device text appearance and layout during a reading comprehension experiment (Chapter 5). In order to obtain better insight into what parts in a reading process most influence the subjective experience – what are the elements that readers take notice of – it was necessary to observe how the viewing parameters of an interface were adjusted by test participants. It was also important to ask how these adjustments changed the reading experience from when the device was in the default settings. By adjusting the settings the test participants indicated what parts – if any – of the interface most influenced their experience, thus providing help for interface and content designers in their quest to improve their designs.

In the said experiment reading comprehension was tested under three conditions: paper, mobile device and mobile device after participants had adjusted text appearance and layout to their personal preferences. The reading comprehension results and participants' perceived experiences with the three conditions for ease and fluency of reading are handled in detail in Section 5.2 while the present chapter focuses on what adjustments were

made and what reasons participants gave or how they described the experience of reading compared to reading on paper and reading on the mobile device in its default settings.

Earlier studies into the effects of layout and appearance on readability have largely concentrated on reading performance. Readability has been measured through comprehension and reading speed, whereas perceived fluency and preferred appearance have taken more of a back seat. The literature points to items such as typeface size and style (Bernard et al., 2003), column width and the number of columns (Dyson and Haselgrove, 2009; Dyson and Kipping, 1997), structural markers (Samuels and Eisenberg, 1981; Frase and Schwartz, 1979) being important factors in readability and reading comprehension as well as reading speed. In terms of perceived fluency and satisfaction, internal motivation (an individual's need or wish to read), and opportunity (time and place) make a difference in how positive or negative a reading experience is. Further to that, interaction and navigation come up as factors in perceived reading experience: how much text is visible at a given time so that there is no great need to turn or scroll pages and so that navigating within the text is fluent and easy. (Wilson, Landoni and Gibb, 2002).

Bernard et al. (2003), looking into the effect of typefaces and their sizes on electronic screens, found that while performance measures do not necessarily reveal differences in reading performance as such, test participants were still able to point to differences in their perceived experiences and indicate a preferred size and style. Also Yi et al. (2011) were able to indicate an optimal combination of column width and line spacing in terms of performance, which was then in slight conflict with their participant feedback on preferred column width.

The present study examines the observation data from the reading comprehension test reported in Chapter 5 in order to compare the actual adjustments participants made to the text appearance, what participants said

about the changes, and what factors have been presented in both existing research and in the performance results from the said comprehension test.

7.1 Design

The overall experiment design for the reading comprehension test during which the present data was collected is described in Chapter 5. The present section focuses on examining and observing trends and frequencies in the adjustments that test participants made to the layout and appearance of text on the mobile phone in the last phase of the said reading comprehension test.

On the whole, the test was a repeated measures mixed methods experiment where performance scores on the Cloze type comprehension test, time on task and perceived ratings on ease and open-ended questions produced dependent variables. The text sets on three different presentation formats (paper, mobile phone set appearance, and mobile phone where appearance was adjusted by the participants) were the independent variables in the experiment. The data in the present chapter is handled qualitatively throughout.

7.1.1 Participants

The reading comprehension experiment which provided the present data had 45 participants, between 18 and 55 in age (majority, some 76%, between 18 – 25) and predominantly female (82%). The sample came mostly from psychology undergraduates at Northumbria University. The sample is also described in Section 5.1.2.

7.1.2 Materials

The text to be adjusted was the third and last part of a reading comprehension test that compared reading comprehension performance and perceived experience between reading on paper and reading on a mobile phone. The mobile phone reading was completed in two ways: first with layout and

appearance that mimicked the presentation of the text on the paper format used, and, in the last part of the test, allowing participants to set the layout and appearance to their personal preference. The details of the reading comprehension test, text sets, etc. are described in Chapter 5.

When the last part of the reading comprehension test started, the mobile phone used in the test was presented with the default settings that had already been used in the first mobile reading comprehension test part: column width was approximately 4-7 words with a small margin so that the appearance of the text was visually as close as possible to how the text sets had been presented in the paper format. Figure 16 illustrates the appearance of the text on the mobile device screen; the paper printouts are included as Appendix 12.1. A free mobile e-reader software Stanza was used in presenting the test texts on the mobile. Text was presented as black on white, using Stanza application's default typeface, Georgia, which is a serif typeface. Typeface size control was not set by pixel size in the software, and so it had to be done on visual estimate. All typefaces on the device were anti-aliased. Device (iPhone 3G) screen resolution was 320x640 pixels (at 163 ppi) on a 3.5"screen. Brightness was set high so that the contrast was crisp and clear.

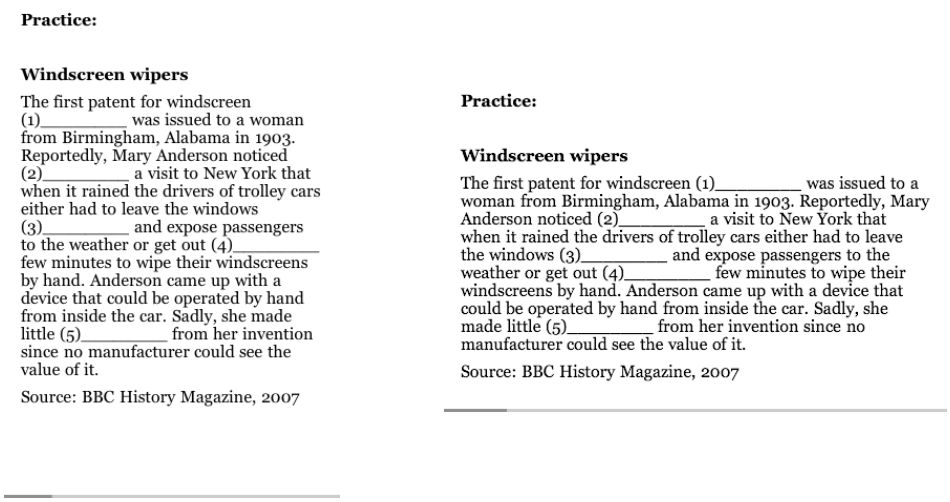


Figure 16 Reading comprehension test text sample as presented on the test device; portrait and landscape.

By the time the third and last test started, participants had completed a similar text set Cloze test on paper and on the mobile (in alternating order).

Participants were given guidance on what adjustments were available in Stanza application settings and on how these adjustments could be made.

Appearance adjustment options on Stanza application:

- theme (pre-defined list of colour schemes and background textures),
- font (selection of serif and sans serif fonts),
- font size,
- background colour,
- text colour,
- link colour,
- background image and
- (background) image opacity

Layout adjustments:

- alignment (justified, left, centre, right)
- hyphenation
- margins (width)
- line spacing
- paragraph spacing
- paragraph indent

As the software used on the mobile device was selected on the basis of its not only offering a good range of adjustments but also being freeware that allowed privately compiled .pub files to be used easily (not always the case with content provider specific software), the concerns for measuring adjustments exactly took a lower priority in the selection process. For the participants, however, this did allow for a subtle, intuition-based approach to changing any appearance parameters as the controls simply showed the difference between default state and any changes made by participants in a preview.

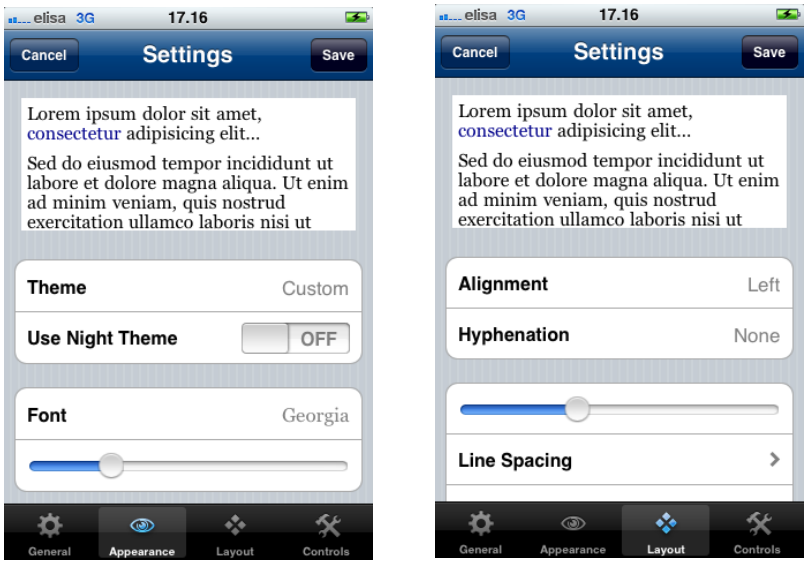


Figure 17 Stanza e-reader settings (example).

7.1.3 Procedure

Participants took part in one-on-one sessions where after some initial background questions (answered on a computer onto an electronic questionnaire) the moderator explained the procedure and provided them with the participation information regarding the test content and arrangements, as instructed in Northumbria University Life Science ethics guidance. The participants were asked to complete a reading comprehension test in three parts.

After the first two parts were completed (reading on paper and reading on a mobile phone where the text had been set to look like it did on paper, alternating the starting order of these formats), the participants were invited to make adjustments in the appearance and layout of the mobile device e-reader application to suit their own personal preference. The adjustment possibilities and how they could be manipulated on the interface were explained to the participants, and they were also told they did not have to necessarily alter anything if they were happy with the text appearance as it stood. The list of adjustments is described in more detail in Sections 5.1.3 and 7.1.2.

After trying some adjustments and confirming that participants were happy to carry on with their present settings, the final text set was displayed on the screen and participants completed this last comprehension test. Their

performance was timed discreetly to the closest minute of time on task, and after completing the test participants filled in a questionnaire regarding their experiences with all three parts of the test. The questionnaire (presented in Appendix 12.3) included Likert ratings on perceived ease and fluency as well as open-ended questions on the participants' reasons for the said ratings.

7.2 Results

The grounded analysis of the participants' open-ended reasons for their perceived ratings regarding the three reading formats they encountered in the test focuses on the generalized statements as they describe the impact of making any changes to the interface. The participants described the experience of reading text and completing the comprehension test after they adjusted the appearance settings and compared the experience to what it was like to read and complete the test on paper and on the mobile where the text was set to look as much like the paper layout as possible.

The present section is a cross-case analysis (Patton, 1990) and looks at the same comments, but focuses on the actual adjustments that were made with the intention of reading the open-ended comments in the light of the participants' comments on those specific adjustments. The weight of the findings is on the frequency of specific adjustments and then projecting the observations over the core findings from the open-ended comments analysis.

Examples of various adjusted device states are given in Appendix 12.5.

7.2.1 Adjustment types and frequencies

The adjustments available on Stanza are described in Section 7.1.2. Table 16 is a matrix of the types of adjustments participants made in the test. Six participants made no adjustments at all, and the category 'other' is a collection of adjustments that occurred less frequently than twice and which for that reason are of no great consequence.

The highest frequency of adjustment fell on making the typeface larger (40%), followed by turning the device into landscape (29%) and adjusting the margins to allow for more text on the screen and changing typeface from the default serif to a sans serif, both done by 22% of the participants. The only other category in addition to the ones mentioned already that was adjusted by more than 10% of the sample was the adjustment to increase line spacing. The rest of the adjustments occurred so rarely that they have little significance in themselves.

Examining the adjustments further, the matrix readily shows the combinations of adjustments where participants made changes to more than one parameter: the two biggest individually counted frequencies, larger typeface and turning the device to landscape, are also the most frequent pairing of adjustments. It should be noted, however, that the frequency is still only 8 participants out of 45, but other combinations only occur no more than three times in the sample.

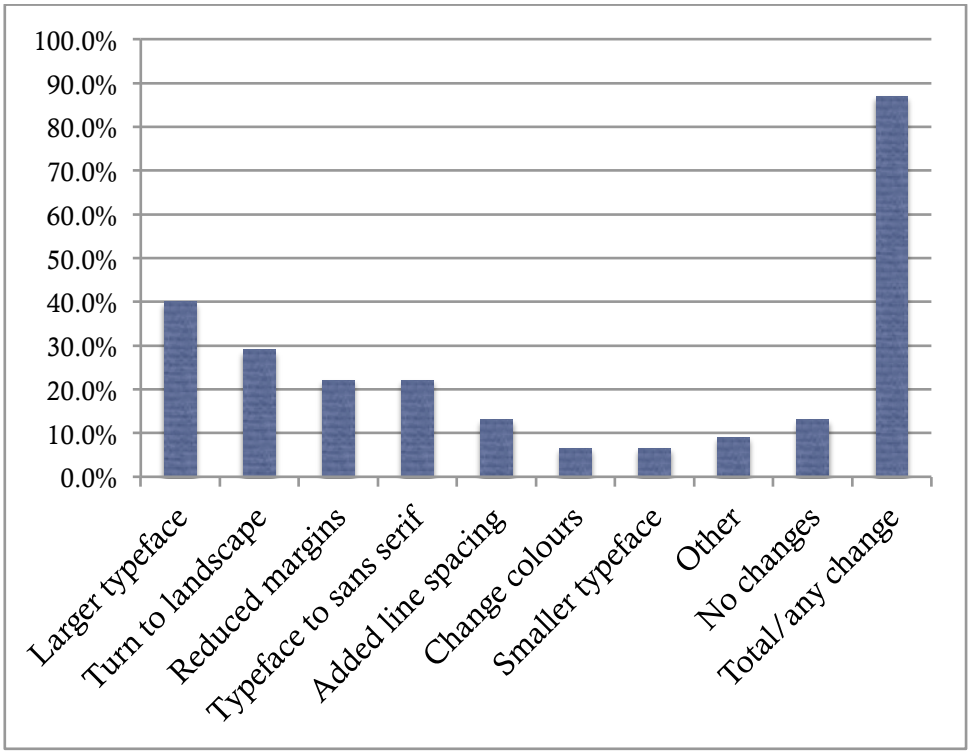


Figure 18 Frequencies of adjustments per identified category.

Figure 18 illustrates the frequencies of adjustments in percentages per identified category. From the bars it is easy to see that although a great

majority of participants made adjustments to the appearance, these adjustments were spread fairly thinly across the categories and only the first four have actually gained any mass at all: larger typeface, turning the device to landscape, reducing line spacing and switching to a sans serif typeface.

Table 16 Matrix of adjustments made by participants. The 6 participants who made no adjustments have been omitted from the table.

Participants 1 -39.	Larger typeface	Turn to landscape	Reduced margins	Sans serif	Add line spacing	Change colours	Smaller typeface	Other
	■							
	■							
	■							
	■							
	■							
	■	■						
	■	■						
	■	■						
	■	■						
	■	■	■					
	■	■		■				
	■	■			■			
	■	■				■		
	■		■	■				
	■			■				
	■			■	■			
	■							■
	■				■			
		■	■					
		■		■				
		■						
		■		■		■		
		■						■
			■					
			■					
			■					
			■					
			■					
			■				■	
			■				■	
				■				
				■				
				■	■			
				■		■	■	
					■			
					■			
								■
								■
	18	13	10	10	6	3	3	4
	40%	29%	22%	22%	13%	6.50 %	6.50 %	9%
	39/45 participants made some adjustments (above), 6/45 did not make any adjustments to the viewing settings.							

7.2.2 Open-ended responses: adjustments and their perceived effects

The present section draws together the main themes from participant comments showing the themes in the context of some descriptive quotations from the participants' answer forms. Note that the quotations are as they were typed by the participants: the laptop keyboard used in the test situation was possibly not very familiar to most participants because of the Scandinavian keyboard layout.

7.2.2.1 Seeking improved legibility and matching line length to typeface size

As regards those comments from participants who wanted a larger typeface, the reasons almost uniformly mention better legibility: easier to see in general, easier to distinguish (recognize) words. Participants who also turned the device to landscape were looking for a good compromise between legibility and fluent line length.

> *[Reading was] a bit easier because text was easier to see.*
>
> *I made the font size bigger which made it easier for me to read this is because the words looked more spaced out and a lot less closer together - was easier to distinguish the words.*
>
> *Once I had changed the font to a size that was easy to read, it was automatically easier ti complete the task. The change from portrait to landscape also helped because it felt like a more natural way to read and the lines were longer so it was more fluent.*
>
> *I think it improved my ability to read it more easily. But it didnt really help with the fluency issues as i still only had a limited amount of text available at any one time. It was hard to find a balance between font size being readable enough to comfotably hold it at a reasonable distance and the amount of text available on one page.*

The comments on turning the device to landscape were mainly concerned with legibility and fluency of reading that comes from being accustomed to a

certain style or simply preferring longer lines to the narrow column style offered by portrait orientation of the device:

> *I found it was quicker to read when the text was slightly larger as I felt I made less errors and it was easier to read as more text appeared horizontally.*
>
> *I turned the phone sideways to make it easier to read*
>
> *With the writing being a bit bigger and the sentences spaced more widley it was easier to process the words smoothly and find an appropriate word to fill the gaps*

In the case of landscape, particularly with larger font size, the comments refer to fluency. Given that such settings reduce the amount of text visible on the screen, it should be noted that the fluency in these cases seems to refer to localized fluency rather than a more overview-oriented fluency that allows skimming forward and backtracking and which demands the least amount of disruptive page turning.

7.2.2.2 Optimizing the amount of content on screen

Reducing margins was the third most frequent category of adjustments. The comments in this category tended to focus on maximising the amount of content on each page and thus reducing the need for page turning:

> *… meant that I could read more on each page and the reading wasnt broken up with page turning. It seemed the best comprimise between book and phone.*
>
> *I could view more of the text at a time, which increased fluency, with fewer distractions turning pages. But there's more page turing than the paper format.*

Changing typeface was also most frequently justified on the grounds of improved legibility:

> *The text/font changed helped, but the size limitations of the screen meant I still couldn't read the text as a whole. Kept moving back and forth.*
>
> *i felt when i had modified teh phone it became more easier due the fonat begining slightly more of an easier font to ready and*

increasing the spacing between the lines made it more clearly to distunguish between the rows of text.

The format i chose had bigger font and was more spaced out which made it easier to read, although i found this comphrehension harder than the other 2.

7.3 Discussion

Yi et al. (2011) reported from their anecdotal material that participants possibly found longer line widths somewhat daunting. In these tests by Yi et al. participants were subjected to a combination of two column designs (one column vs. two columns) and three levels of line spacing (single, 1.5 and double). The experiment excluded typeface size and type as items that are normally user-set on e-readers and therefore, in their opinion, not so crucial to the understanding of better design for e-books. Despite the voiced dislike of longer lines, Yi et al. argued that the best overall result in terms of readability, comprehension and satisfaction came from a combination of single column set on 1.5 line spacing. Learning was affected more by line spacing with single column design than with multiple column design. (Yi et al., 2011).

Comparing the results to the adjustments actions taken by the participants in the present experiment, it is interesting to note that turning the device to landscape for longer lines was the second most popular adjustment, done by just under one third of the participants. However, more than half of these also opted for a slightly larger typeface, thus negating some of the line width effect in terms of how many words would be presented per line. A more detailed experiment setup would be needed to observe how much, exactly, the typeface size was altered and how great an effect it had on the words per line count.

Typeface size was something that was ruled out from the experiment variables by Yi et al (2011), but in light of the frequency of the size being adjusted in the present reading comprehension experiment (and also when paired with landscape orientation of the device) it would seem that default typeface size

should indeed be something for device and content designers to think about: specifically, what sizes of typeface and line spacing they offer as a default. These decisions might well have a significant effect on end-users' first impression of content they see on their device, whether e-book or a mobile or a text-heavy website.

The analysis of participant comments on ease and fluency of reading and completing the comprehension test after making individual adjustments to the appearance and layout fell into the following categories:

- Identifying more text per screen as a source of improvement in the user experience.
- Identifying that more text per line with landscape orientation of the device made for a better user experience (often coinciding with slightly larger text).
- Obtained sense of improvement in general in the reading experience and performance owing to the changes participants had made.
- Conclusion that adjustments improved the experience compared to a mobile device that had not been adjusted, but stating that the experience was still not quite good enough or not as good as reading on paper.
- Perception that alterations made no observable difference in the (perceived) user experience.

In the light of these categories, it is interesting to note that the most frequent adjustments were to make the typeface larger and to turn the device to landscape orientation – the former in particular. It looks as if participants who adjusted the text size and orientation parameters were more concerned with legibility than overview, which was the overall dominant category in the analysis of the full body of comments. Earlier studies into optimal line lengths have ended up recommending a moderate length that allows enough information to be gathered within a saccade and also maintains a short enough distance between the end of one line and the beginning of another so

that the sweep backwards to fix the gaze back at the start of a new line would not have too long a span to make this difficult (Dyson and Haselgrove, 2001). It would be interesting to study the proportional and layout related factors affecting the optimal line length: what variation tolerance to font size do the character-based line length recommendations have and would increased line spacing allow for longer lines? Is there a physical mm distance that can be defined where the backsweep to the beginning of a new line becomes difficult?

Typeface size was increased by 40% of the participants, which would seem to speak volumes about the compromises that are necessary in finding a fluent level of readability. Participants are able to identify the constant interruptions that a small screen causes as an irritant and a factor that affects readability negatively, but an uncomfortable text size hinders fluent reading even more. Some participants in the present experiment put special emphasis on maximising content and minimizing the need for turning pages and did so mainly through reducing margin width.

The performance results in the comprehension test showed significant differences between paper, mobile and adjusted mobile in time on task between paper and adjusted mobile, and on perceived fluency of reading where both paper and adjusted mobile were deemed significantly more fluent than the mobile without any participant-made adjustments. Participant comments on larger font, landscape orientation and switching the typeface to a sans serif put emphasis on ease and fluency: it was important to these participants to be able to see clearly enough and in that way achieve a sense of fluency and ease in reading.

The one other area where the comprehension test showed any significant difference was in time on task between adjusted mobile and mobile in the default state. In the comments, however, speed of reading or completing the task does not seem to come up. This could be a result of a number of reasons in the phrasing of the question, namely that participants were not specifically asked whether they thought the adjustments affected their time on task, and also that the terms “easy” and “fluent” can in themselves contain the concept

of speed: the more fluent the reading the faster it probably is as well. Further research into the area would be required in order to draw conclusions on whether or not improved speed was experienced by readers after adjusting the mobile screen or not.

The experiment by Bernard et al. (2003) into typefaces (style and size) concluded that although a probable optimal combination, based on perceived experience, was found from the ecologically viable combination of typefaces and their sizes, for the most part it is difficult to measure accurately enough to see what difference there is exactly between 10pt and 12pt font, or between Arial (sans serif) and Times New Roman (serif) typefaces. Yet their participants were able to point to a combination that suited them best – and, indeed, to one specific combination rather than all participants having their own, individual preference as such.

On the hands-on front, design guides for web typography such as Webitect (2013) provide simple heuristics (rules of thumb) for pleasing-looking typographic design for electronic viewing. For example, Webitect's top ten tips include the recommendation to decide on font size on the basis of column width (the wider the column, the larger the text). Webitect also suggests that although generally sans serif fonts should be more legible on electronic devices, serifs should work equally well so long as their setting leaves enough white space around them.

The participants in the present experiment opted for a larger font size basically regardless of line width: 29% (13 participants out of 45) completed the last part of the comprehension test in landscape, and of them some eight also increased font size. This means that of the 60% (32 participants) who opted to complete the task in portrait orientation, and thus with a shorter line width, 10 participants (22% of total, 31% of those who used portrait orientation) still wanted the larger text. Bernard et al. (2003) concluded that the participant preference had fallen on the larger version of the sans serif typeface (12pt Arial), also interestingly in dot matrix rather than anti-aliased even though anti-aliasing has been brought in specifically to make text smoother and more

legible. In the present experiment, a larger typeface was accompanied by a switch to a sans serif a total of five times (sans serif was selected by ten participants, 22%, in total). On the basis of these figures it is not possible either to support or to reject the results by Bernard et al. (2003) Instead it would seem that there is a call for further research to find a rule of thumb for ideal typeface size and style for mobile device screens that would serve as a functional default. In the present experiment participants were given a free hand in making adjustments, but the role of inertia (Thaler and Sunstein, 2008) should not be ignored when leaving active changes in the hands of consumers or, in this case, test participants.

In terms of user experience, the direction of the adjustments made by the test participants emphasizes the importance of immediately clear legibility. Regardless of the difficulty of pointing to a universally applicable default text size, it is still clear that end-user testing can be used to iterate (and even validate) a product's final default settings and to provide valid feedback on functional text size on a given device. Generic e-text rules of thumb as the one mentioned earlier (Webitect, 2013) will need to be reviewed in the context of small displays: there is only so much room to play with in terms of column width, and as was seen from the adjustments in the present experiment, the users were often trying to find a compromise between legibility and how much content the set width (both portrait and landscape) allowed in terms of words per line. Interface designers should aim at establishing a functional compromise for default text size to words per line setting instead of relying on end-users to adjust the levels themselves. These default levels can be established through even small scale user testing, especially if the aim is to produce those settings for specific display resolutions and aspect ratios.

To summarize, the participant comments and the adjustments they made to the layout and appearance of text on the mobile phone that was used in the reading comprehension test focused mostly on making the text more legible and on assisting the flow of reading in a localised sense (reading the immediate sentence or phrase as opposed to the overall flow of getting a more collected view of the text so that skimming and backtracking would be

enabled). Although the localised fluency was mentioned as the reason for larger text and landscape orientation (and sometimes increased line spacing), a clear majority of the participants maintained portrait orientation and in the overall open-ended answers the importance of an overview (wider context rather than localized fluency) was emphasized. Of all the participants, larger text was the most frequently selected adjustment, with 40% of the participants making the adjustment, followed by landscape orientation (29%) and tighter margins and sans serif typeface (both at 22%).

In the three experiments described in the present thesis so far (chapters 5, 6 and 7), a clear trend in the results shows that while performance (measurable efficiency) may remain at a similar level regardless of reading media or even visual distractions, the user experience differs and end-users' responses to various media and distractions has significance. The following three studies (chapters 8, 9, and 10) explore the subjective experience of the reading process.

CHAPTER 8

Which presents a modified critical incident study into reasons why end users might not carry out a reading task on a mobile device.

To read or not to read: some real life examples of when reading on a small screen has not been the best solution

The previous chapters established how performance of reading remained similar whatever the reading medium (paper vs. mobile device), and in the case of mobile devices even changes in visibility conditions and settings had little or no impact on the actual reading performance. Yet in each experiment it was established that subjective experience regarding the media, legibility and layout varied significantly. The experience of reading and accessing content would appear to play a significant role when end-users decide to use devices with small screens for reading. This chapter presents a modified critical incident study carried out with a number of experienced interface design professionals. In the study these expert users were asked to share their experience of an incident where they had started to access text content on their mobile devices but had, for any reason, stopped the task.

The studies presented in chapters 5 and 7 in particular pointed to visibility elements such as overview of content and basic legibility playing a role in the end-users deciding whether to read text on a mobile. The reading comprehension test (Chapter 5) showed that although reading performance did not differ much between various formats (paper or mobile), participants in

the test were able to indicate elements that contributed to a positive or a negative reading experience between those formats. Overview of the content and the interaction required to navigate within the content took prominence in the findings of the reading comprehension study.

Factors that could be anticipated emerging from the responses in the present study include external distractions affecting attention and concentration, layout and appearance preferences, excessive need for content navigation and device interaction due to the screen size, irrelevance of content at the time of accessing it, and problems with data transfer (such as a page or e-mail not loading onto a mobile device). Of these anticipated factors, the distractions affecting attention and concentration cover a multitude of physical contexts of use from social interaction to using public transport or walking on the street. Adding mobility and social demands to mobile device use has a significant impact on the attentional resources a user can focus on interaction and content, as reported by Oulasvirta et al. (2005).

Overall design of the content could also, potentially, be off-putting for an end-user – especially if combined with less-than-ideal context of use. A number of studies into e-reading point out that whilst functionally end-users may be able to read and complete reading-related tasks with e-texts at a satisfactory level, the perceived experience does differ enough for end-users to be able to form clear opinions on what styles and appearances they prefer (Dillon et al., 1990; Bernard et al., 2003; Yi et al., 2011). Participants in the reading comprehension text drew attention to the problems they had in navigating within the content and their irritation at having to interact with the device too frequently (having to turn a page often). Then again, both the afore-mentioned interview study as well as the survey study suggest that if the content is interesting enough or in demand at a given moment, a number of visibility and interaction difficulties would be disregarded and content would be accessed.

Would a small sample of real-life experiences with interface design experts touch on the same themes as the results from the experiments in the earlier studies and give the findings added depth and definition? Would accounts of

specific events in real context of use help us better understand the nature of the acceptance of reading on a mobile screen? When the expert users were asked to describe and verbalise their experiences in free form and in their own words, what aspects of mobile reading emerged as prominent ones?

8.1 Method

The study method was based on critical incident technique (Flanagan, 1954), in which observations of human behaviour are collected for example by asking participants to describe some specific experiences. The aim, generally, is to be able to draw inferences and make predictions about behaviour, and learn about behaviours that result in success or failure in organisations or by individuals. For the present study, a more hands-on usability engineering approach was adopted, much as described at usabilitynet.org. Overall, the method comprises collection of data (typically one or more incidents from a number of system users, typically by means of an interview or questionnaire). The questions define the activity that is studied, and ask respondents to describe the context and outcome of the situation. Unlike the method described at usabilitynet.org, the present study did not start by defining an activity from which, retrospectively, the incidents would then be recognized. Instead, with a very focused theme and problem already defined (potential reluctance to read continuous text on small screens), the respondents were asked about such a specific situation, under what context and situation the incident took place, and what was the outcome of the situation.

At the analysis phase, the aim was not so much to produce a list of incidents categorized by severity and as such to produce a solution to each observed problem, but rather to categorize the observations and incidents in accordance with the issues and problematic areas identified in other studies in the present thesis. In typical industry usability research the number of responses in the present study is considered sufficient for iterative research on a specific problem: 5 – 6 participants, which according to Nielsen's usability website (www.useit.com), is enough to reveal some 80% of usability problems in a typical usability test. In the present study the respondents were professionals

in the field of user experience and thus possessed specific expertise through which to describe their experiences. While it would be easy to side-track at this point and criticize Nielsen's advocating only five users in usability testing, it should be noted that the premise seems to suit the present study in that the present study is only looking into one particular problem and is asking the respondents to describe a specific incident instead of an activity from which those incidents would be drawn.

Flanagan (1954) outlines the analysis process in two steps: categorization of the critical incidents and drawing inferences regarding practical procedures for improving performance, based on the observed incidents. Almost of necessity the first step is relatively subjective, but the categorization is ideally done by a qualified observer. In the present study the researcher has nearly a decade of hands-on experience from mobile communications user experience work in a research and development environment, paired with the research into small screen readability in the present thesis. Step two can lay claim to greater objectivity as the incidents are distributed into the established categories. In the present study the analysis aims at identifying what elements in each incident have caused the problem (respondent not willing to complete reading task) and discuss these elements against the findings from the other studies in the present thesis.

8.2 Procedure

The study was carried out by sending the incident question by e-mail and through a social networking site to a number of user interface professionals. Participants were asked to provide an open-ended answer to the following question using a web response collection tool provided:

> *Have you had, recently or otherwise, a situation where you have needed, wanted or been requested to view and read text format content on your mobile, but you decided not to read the content? Describe the situation and why you did not proceed with reading the content? Describe the following: Was the content something you were looking for yourself, or had someone else asked you to read it? Did you choose the time for viewing the content yourself?*

Why did you not wish to read the content there and then? Did you return to the content later, and if yes, was it on your mobile phone or on some other device?

8.3 Respondents

A group of user experience professionals (several years of experience in interface as well as information design, mostly on mobile devices) were asked to think of an event where they were expected or they wanted to access some text format content on their mobile phone, but for any reason then decided not to do so. The respondents were, as said, user experience professionals who were colleagues and acquaintances of the researcher in Finland, the UK, the USA, India and Japan.

8.4 Responses

Overall, ten people responded, one of whom simply replied simply confessing to not having encountered the kind of situation the study asked about, while three left the open-ended answer form empty either because they could not think of an incident or because they decided they did not want to answer after all. The responses from the six participants who did leave an account of a situation where they did not read some content on their mobile varied from over-long content (too much text in proportion to the size of the interface) to web design issues and simply deciding on the basis of content whether it was worth reading or not. One response was collected in person (interview). The responses are included in the present section as typed in by the respondents.

Knowing what the content is before seeing it

Two responses directly indicated that filtering based on some early indication of what the content was would result in the respondent ignoring the content. Another response (listed under another category later) touches on the same topic. Content was filtered by identifying its source (e-mail sender) or by an initial scan of the content of a message (see response later in the category “Content/text too long or too complicated to view on a mobile”). Content was

also filtered by its position and early identifiers (heading) when presented on a website. In all these cases, identifying the most likely content of a communication or text content resulted in the respondents performing a quick prioritisation of the communication, deciding either to return to it later (whether on the mobile or on PC) or to ignore it completely (website with 'terms and conditions').

> *"I rarely read it [terms and conditions] as I assume that the provider has a reasonable duty of care for my best interests and is governed by ethics."*
>
> *"If it's an e-mail I often know what it'll be about by just seeing who sent it, and I'll know it'll be about something I'll not want to or need to respond to straight away."*

In both cases quoted above the motivation for not continuing with the content is internal to the respondent: the message does not require immediate attention or the content is irrelevant/unnecessary/redundant. In the third case that was mentioned (under the category 'Content/text too long or too complicated to view on a mobile') the motivation for discontinuing with reading the communication was a combination of internal elements together with issues that are related to the interface: the respondent would have been happy to read the communication, but the action that was expected based on the communication was deemed too complicated to complete on a mobile device.

Layout and appearance / interface design

The following responses highlight elements in content design that make reading the content too complicated or otherwise irritating on mobile interface. In the first response, a web interface that was familiar to the respondent had been changed in a way that affected both legibility of the text as well as the interaction with the website: default text size was too small, and even the zoom interaction the respondent was used to (double-tap on the screen) to bring text size up to an acceptable level was not sufficient to provide a readable website. Any further zooming resulted in the column width of the text changing so much that reading a line would require horizontal panning, which the respondent considered a major inconvenience.

The second response points to advertisements and their design as the reason for not continuing with the task the respondent had initiated: reading the news. The text content itself may have been readable and acceptable, but the added parts from the website (advertisements) were using distracting technology (Flash animation) and also covered some of the content the respondent specifically wanted to see.

> *I have been regularly following a blog for about six months now (http://www.lily.fi/palsta/kaikki-mita-rakastin). I usually read the blog during my commute. I download the page with latest blog entries on my mobile when walking to the subway station so that I have it ready for reading before I lose internet connection. The blog recently moved to a new address and something happened with the sizing of the text (column width I think) and the text became unreadable (too small). Before, when I tapped the screen where the text was twice, it would adjust the text to fill the screen so that I would only have to scroll down, not sideways too. I guess it still does the same thing, but the text is so small that I would have to further zoom in and scroll not only up-down but also sideways and I will not do that. So, to answer your specific questions - I was looking content for myself - I chose the time for viewing the content myself - I chose to read it on the train because there is usually enough new text on the blog to keep me entertained the entire trip. There is no internet connection, so I can't browse (which would be my preference). - I did. Although I'm busy during the evenings and would rather use the train time to random entertainment things and my time in the evenings to something else than sitting in front on the computer. - I used my laptop to view the blog. And I do it now regularly a couple of times a week. I also occasionally go back on my mobile phone if I've read all the other bookmarked blogs and have nothing else to view when I'm bored. But it's rather painful so I'm not happy to do that.*
>
> *Wanted to read the news in the morning while in the bus, browsed thru the topics and opened a news item, but big flash ads blocked some of the text and made sliding down the page hard with the touch screen. Read the news later with my laptop.*

In both cases the respondents were accessing content on their own initiative and choosing, and in both cases the content provider had not managed to provide a mobile web design that was functional and usable. The content may have been perfectly fine for these respondents in terms of language, colours, contrasts and a number of other parameters, but a single oversight was enough to deter the respondents from continuing with the task. For the first

respondent here it was the change in the text size and the lack of flexibility in the website construction that did not allow functional zooming, and for the second respondent distraction and secondary content covering the primary content were enough to discontinue reading.

Content/text too long or complicated to view on a mobile

The first response in this category quite simply states that the text he or she accessed was just too long – whether for the moment or for a mobile device is not entirely clear from the response. The text in question was accessed on a laptop, but it is not possible to tell how much later this happened and, as said, whether the text was too long to be read at a particular moment or if it was simply a question of the interface size.

The second response is the one that straddles the first category presented ('Knowing what the content is before seeing it'). The message reading is interrupted, as discussed earlier, when the respondent recognises what the rest of the message content will be and therefore does not need to read any further. The motivation to stop is more complex than simply knowing what the content will be: it is also to do with the recognition of the resulting actions that would follow from reading the message in full.

In the third response in the present category attention should be paid to the effect even a small difference in display size can have on the experienced fluency of reading: the respondent did not have access to a laptop/PC with internet connection, and so used a mobile phone to make sure he or she would know when an important message arrived. However, when the message did arrive, reading the message was postponed until the respondent's other phone (with a larger display) was available.

> *-Was looking for myself -Yes -Too long textual content, several pages -Yes, used laptop to read the content*
>
> *Yes. Here's the situation in question: A friend sent me a private message on Facebook, asking help in deciding what kind of laptop she should buy. I read the first couple of lines of her message, and decided not to read it further on my mobile, because I don't like*

reading messages or emails if I am not able to reply to them immediately, and I knew that answering to her message would require some heavy duty googling and comparison of laptop details - something that I cannot do properly on a small mobile device screen. The message content was not something I was looking for myself, as it was something my friend sent to me. I check Facebook messages regularly, and quite often while going home (when I'm the passenger, obviously). I chose the time to check my Facebook messages myself (I do not have push notifications on), and I also chose the time to read the message properly through on my computer later that same day.

I was waiting for an important e-mail that I needed to react as soon as I got it. So I really did not choose the time, but had no Internet access but only 2 mobile phones. I saw that I got the mail, but on my smaller mobile so I decided not to view it from that device, but later read the mail with my bigger mobile.

8.5 Discussion

Earlier in this section a number of factors were hypothesized to emerge from the responses in the study, namely external distractions affecting attention and concentration, layout and appearance preferences, excessive need for content navigation and device interaction due to the screen size, irrelevance of content at the time of accessing it, and problems with data transfer (such as a page or e-mail not loading onto a mobile device). Generalizing from the responses in the study, almost all of the expected factors do seem to make an appearance with the exception of data transfer problems and external distractions. Also, ‘irrelevant content’ did not seem to come up as content that respondents browsed by following links or suggestions from other people (such as in a message from a friend): respondents in the present study seemed to filter push notifications or otherwise keep on top of who or what initiated the decision to seek out certain information or content. Another item emerging from the responses that touches on the idea of ‘irrelevant content’ is content where the respondents recognized the content or at least knew they would not be interested in it: one respondent identified ‘Terms and Conditions’ on some service or website as something that would not be read at all.

Distractions that affected attention were expected to manifest themselves in the responses in the form of external and context of use distractions, but the present sample did not point to external disruptions. Rather, distraction to reading content came from within the device, service or interface: Flash adverts and additional interaction leading to text size level that then failed in terms of layout. A subtext of the responses suggests that respondents were used to selecting convenient times and situations for accessing content instead of permitting push content – or at least the respondents did not allow themselves to be distracted by it.

Excessive content navigation and interaction came up in the responses in several places: layout (design) problems with websites and content that was too long or too complex would lead to additional interaction and extra effort in integrating content spread over several “screenfuls” on the mobile interface. In all the categories with, perhaps, the exception of ‘Knowing what the content is before seeing it’ the outcome was, eventually, to access the content with a device that was better equipped to handle the content than a mobile phone, regardless of whether the respondent had initiated the interaction with the content or not. One of the responses suggests that the “better” device still need not be “optimal” (i.e. computer or other device with a bigger screen), so long as the situation is improved.

In the response where the respondent does not read a message through because it becomes obvious that he or she would have to follow up with web searches and price comparisons, etc. an important aspect of ubiquitous computing and communications is highlighted: communications and social networking take place in time and space, and ideal conditions cannot always be guaranteed. Messaging is not simply a question of sending and receiving messages, but being able to read and to compose a response may require specific conditions (context of use), and very often a message will require additional actions before a response is possible: in this case the respondent knew the actions would be too complicated to complete on his or her mobile, given the immediate context of use.

The present study and its results suggest that many of the concerns discussed in the other sections of the present thesis are, indeed, identifiable in open-ended and freely-framed accounts by mobile phone users. Issues regarding visibility and legibility, layout and design, personal or time-critical interest, and simply the size and interaction design of the devices used all feature in the responses from the participants in the present study.

The respondent group in the present study was made up of user experience professionals whose work and experience will have guided their attention when thinking about mobile interfaces and user behaviour (even their own user behaviour). The expert evaluation presented in this chapter points to similar core topics as the subjective feedback in the earlier chapters (5, 6 and 7): experienced legibility, layout, overview and navigation all come up in the expert responses. In view of these results, it is specifically the experience paradigm of reading on small screens that deserves attention. The following chapter describes a qualitative study into end-users' expectations and attitudes regarding the prospect of reading various everyday materials on their mobiles (Chapter 9): what elements in the presentation of text are most central in their decision to access (read) text content on a small screen.

CHAPTER 9

Which presents an interview study into how mobile phone users perceived reading on a small screen.

Perceived comfort and fluency factors in reading, and their bearing on the acceptability of reading on small mobile displays

As seen across the preceding chapters, the reading process has been explored in terms of performance and experience side by side. The previous studies in the present thesis show how experience and subjective perception of the reading experience do not always go hand in hand with the measured efficiency of reading. The participants' perception of the reading process determined the adjustments they made to the reading interface during the reading comprehension experiment, showing what visual elements took priority when end-users aimed at improving the quality of the reading interface to better match their liking. The present chapter describes an interview study into the perceived attitudes and expectation as regards reading on small display devices. Ordinary mobile phone users were interviewed about their present reading habits and preferences and then about what they expected reading would be like on a mobile device.

Informal feedback from participants and colleagues involved in the process of usability testing with mobile devices, as well as the results in the earlier chapters, suggest that people experience significant difficulties with screen and text size and find the process of navigating through large content on a very small interface particularly challenging. These observations are supported by

studies from the mobile industry, where different means of maximising the comprehension of text on small-screen displays have been explored, including the use of different scrolling methods (Laarni, 2002) or different forms of text presentation (Öquist and Lundin, 2007).

The present study aims at contributing to this body of work by focusing on the perceived attitudes and expectations that may obstruct information acquisition in text format on small displays before a mobile device user even gets to the point of trying out these various methods. By examining these attitudes and expectations it may be possible to gain an insight into which factors in the reading process affect willingness to use a new system (a mobile interface) for reading, and which factors are perceived by the users to be most important in terms of fluent and comfortable use. By aiming at improving these elements in content and interface design it may then be possible to lower the hurdle and provide a reading interface that better caters for the expectations mobile devices users may have.

The purpose of the present study, then, is to construct a valid description of actual attitudes and perceived impressions as regards what constitutes a comfortable and fluent reading experience, and to discover what other factors affect the reading experience (whether based on experience or anticipated) when it comes to using mobile devices with small display real estate. Is reading a new format merely a question of “getting used to it”, or are there factors that encourage or discourage the adoption of a new medium? In sum, the present study aims at understanding what makes the perceived reading experience fluent and comfortable, and how these elements can influence the reader’s view of reading continuous text on small display devices.

The methodological approach for the study is a qualitative interview study with a sample of participants who represent a consumer category likely to purchase and use mobile phones (and other electronics) capable of web browsing and running e-reading software.

9.1 Method

The aim was to find what elements contribute to perceived reading fluency and which factors are the most influential ones when a reader decides whether or not a small display device is suitable for reading continuous text. One angle for approaching the topic was to find which factors generally defined a comfortable or fluent reading experience and to establish whether those experiences could be transferred to small displays. Grounded theory approach was selected, as the study was exploratory, based on participants' subjective experiences and attitudes, and the method aims at building a theory through inductive analysis of the observed processes and phenomena. The findings from the qualitative study can be used later as input for constructing further quantitative studies on the topic.

9.1.1 Design

Since the focus of the initial stage of studying perceived fluency and comfort in reading is on people's own experiences with reading, a qualitative approach was selected. The intention was also to learn from participants' own experiences, and therefore one-on-one interviews were chosen as the method for collecting data. Grounded theory approach is well suited to building up a theory based on the explorative material, such as interviews. The interviews were open-ended with a questions outline instead of an exact question list so that the data collection could focus more flexibly on the overall views and attitudes from a group of people (Patton, 1990).

9.1.1.1 Grounded theory

Grounded theory is a theory that is derived inductively from a phenomenon that is being studied. The theory is identified, developed and given preliminary verification through data collection and analysis. Thus the relationship between data collection, its analysis and the theory is interactive. Generally, a grounded theory approach is adopted when the target of study and observation is what people do, or do not do. (Järvinen and Järvinen, 2000).

One of the emphases in a grounded theory approach is that grounded theory describes phenomena through abstract concepts and their relationships (Strauss and Corbin, 1990). The analysis of a grounded theory study in the Strauss and Corbin model consists of three steps: open coding, axial coding and selective coding. Open coding is the process in which the researcher reduces the data to conceptual statements (sentences, phrases, or utterances). In the axial coding part the statements are further organized under categories so that the relationships and interaction can be analyzed. (Järvinen and Järvinen, 2000). In axial coding the categories are mapped together by observing the causal conditions, context, intervening conditions, action and interaction, and consequences of the given phenomenon (Strauss and Corbin, 1990). Finally, selective coding aims at identifying a core category that integrates the other categories around it (Järvinen and Järvinen, 2000).

9.1.2 Participants

Twelve participants were interviewed one at a time for approximately 30-50 minutes. The participants were pre-screened to be representative of a population whose consumer behaviour and characteristics suggest that they would be potential users of the kind of technology required to access text data on small mobile devices. Consumer profiling was necessary in order to be able to interview participants who had the potential of being familiar with electronic reading and advanced multimedia and internet capable mobile phones. These consumers would thus be in a position to compare their current reading habits with the fresh demands placed upon them by small mobile terminals. Along with the consumer behaviour characteristics, the participants were selected so that there would be a close-to-even distribution of men and women (7 male, 5 female), and ages ranging from 18 to 45 years. The participants were recruited from Newcastle upon Tyne, England. Five out of the twelve participants were students, the rest came from various professional backgrounds.

9.1.3 Materials

The open-ended questions and notes on possible prompts that were used to encourage participants are included as Appendix 12.4. Two mobile phones (Figure 19) were presented during the interviews as a prompt in order to help participant visualize the topics they were discussing in slightly more concrete terms. In addition, physical prompts included several UK newspapers (both tabloid and broadsheet format), magazines, a novel and a typical text-book.

Figure 19 Left: Apple iPhone (115.5mm x 62.1mm, 4" touch screen). Right: Nokia N78 (113mm x 49mm, 2.4" non-touch screen and hardware keyboard).

9.1.4 Procedure

The interview technique used was open-ended and semi-structured. The participants were interviewed one at a time on university premises and they were compensated for their time and effort. The participants were asked to give their consent to the interviews being digitally recorded (video and audio).

Owing to the multifaceted nature of the topic, in that all aspects of reading experiences were discussed, the interviewer used a number of set questions

for the purpose of keeping the discussion to the matter in hand. A brief introduction to the topic and an outline of the interview session was given to the participants at the beginning of the session. The aim was to have the participants talking freely about their experiences regarding reading in general.

During the interview the concepts of e-reading and reading off a mobile phone screen were introduced and participants were asked to reflect on what it would be like for them to read some of their regular material on a mobile phone. Some participants had previous experience of electronic reading and reading on mobile phone screens and they were encouraged to talk about their experiences. The participants were shown physical prompts of some typical reading materials (newspapers, novels, text books, magazines) as well as two different models of mobile phones currently on the market, a Nokia N78 and an Apple iPhone (Figure 19). In other words, the participants were encouraged to approach the topic from three perspectives: their experiences, including experiences of reading print material, their expectations and assumptions regarding technology (namely, mobile phones) and the way they have experienced reading, and finally spontaneous responses to a sample of text on actual high-end mobile phones.

Though the aim was to have the participants talk freely about their experiences, it was not easy for all of them to verbalise their experiences or to consider the full range of potential experiences around the topic. Accordingly, a number of questions were asked when the participant had trouble approaching the topic. The direct questions included:

a) What sort of things do you read regularly, in day-to-day life?
b) Do you have a favourite read?
c) How long do you read at a time?
d) When and where is a good time to read?
e) Do you ever pay attention to layouts and language (columns, text size, etc.)?
f) Do you read things on the computer?
g) Could you read your personal favourites on the computer screen?
h) Describe what it’s like to look at [an example] text on this mobile screen?
i) Could you read your typical material on the mobile screen?

The full set of direct questions was not needed with every participant, and additional encouragement was given through follow-up questions such as “Can you tell me more?”, “Yes?” and by asking the participant to elaborate on what he or she might have just said. The pre-defined interview frame and additional questions or prompts are listed in Appendix 12.4.

9.2 Results

The interview recordings were transcribed and the transcripts were used as the analysed material. The open-coding phase of the analysis focused on extracting conceptual statements from the material that could be then put through the axial-coding phase. In the axial-coding phase the open-coding results – the statements – formed a set of categories. Relationships and interactions between the categories were then identified through evaluating how the categories affected each other, asking what causal relationships there were and what interaction was taking place between the categories. In the model ‘mobile reading acceptance’ refers to the experienced or anticipated items that influence the reading experience on a small display, such as a mobile phone.

Further analysis of the categories revealed a number of higher level elements that could be seen as a major influence on what participants regarded as a comfortable and fluent read and what issues they might perceive either as enablers or disablers for using small display devices in the same way as they used the presently conventional formats. These elements then formed a basic reading model that describes key reading factors contributing to the experiences of fluency and comfort and eventually the readiness to use small display devices for reading. The reading factors, in turn, are affected by what was identified as device factors: how and on what sort of a device the text content is provided. On the whole, the high-level organization of the responses fell into two main categories:

- Elements whose design and presentation directly affect perception of reading fluency, labelled in the analysis as “Reading factors”.

- Reading media elements, such as what type and size devices were used for reading and how they were used to present contents. For the analysis these were labelled as “Device factors”.

The present section introduces and illustrates factors in perceived reading experience as they emerged from the interview material. The cognitive processes of reading and comprehension are not discussed in the present context as such. Instead, the study focuses on the subjective, perceived experiences of reading, comfort factors in reading and fluency, and finally how the perceived elements interact and affect attitudes towards reading continuous text on small displays.

9.2.1 Reading factors

The factors identified as reading factors have causal relationships with each other, affecting the perceived overall reading experience. On the whole, the degree of positive or negative experience with each factor goes towards building a holistic experience that then affects the eventual attitude towards small display reading, or mobile reading acceptance as it is referred to in the model. Overall, the participant statements pointed to four main structures: how well content could be viewed (basic legibility, how content was presented in general), how easy or difficult it was to form an overview of the content, how fluent it would be to move from one part of the content/text to another, and what the actual interaction on the device was like (selecting elements like links, for example). These responses, then, took the form of four overall factors:

- Visibility
- Overview
- Navigation
- Interaction.

These four factors and how they emerged from the participant responses are discussed in more detail in the following sub-sections.

9.2.1.1 Visibility

Visibility refers to the ability to perceive text in terms of text size, typeface, colours, contrast and, especially in the case of electronic interfaces, graphics quality. The experience of seeing clearly is related to the external interface factors (type and size) and the design of the text (text format). Visibility issues have a direct influence on the perceived need to interact with the viewing options of a given text, with electronic interfaces in particular. For example, if the text size is not big enough, a zooming interaction is required. Size, style and legibility of the text visualization have an effect on reading speed as well (Samuels and Eisenberg, 1981).

There is also a direct effect on perceived comfort when poor visibility makes the reader feel that he or she is having to struggle to see, as illustrated by one comment when a participant was looking at an example web page on a mobile phone screen:

> *"Oh, I have to squint already to see that."*

When visibility is not at an adequate level, reading fluency is not perceived to be very high either. The extra effort required to make out the text itself slows down the reading. The interview participants voiced observations regarding the actual size of text in books and on screens:

> *"With smaller fonts it's more difficult to find the next line. But larger text it's easier to manage, easier to follow on."*
>
> *"Some text are a lot easier to read than other text, you know. Arial, at work, is a lot easier to read than some of the other ones like Times New Roman. I think we normally use like 11 or 12 at work, depending on what it is. Which I think is not bad. You know, obviously if someone wants you to notice something they'll make it bigger or put it on bold."*
>
> *"My grandma used to have bad eyesight so we used to get her book with the big print, really big, At least she could read then. But obviously they were really big books because of the big print. My daughter's reading a book now, but… I don't know who the author is… where the pages were pink. Which really… when we went*

looking for a book to read I personally think she picked the book because it had paper in it. Because normally you have your white paper with the black writing on it, I think she picked that because it had pink paper in the book but it's not that easy to read. Black text, but it doesn't stand out like a normal book. But it got her attention. And I thought that looked nice, and it was keeping with the story, it was about ballet dancing and what ever…"

"Just average, regular font size is good, not too many words on a page."

Participants seemed to have identified an optimal font size for themselves, and being able to adjust the zoom level on electronic devices was a positive feature. Samuels and Eisenberg (1981) explain that there is a relationship between the size of text and how much information can be taken in within a single eye fixation, and that this relationship depends on the individual reader's reading competence. When participants in the interviews set zoom levels on a mobile device to a level that they found comfortable and fluent for themselves, the tendency was to match column width to display width. However, when rotating the device to landscape and the device responded by keeping to the same number of words per line and merely increasing the text size, participants tended to respond either by zooming back out to the text size they had previously selected or by commenting on having to scroll more frequently now that there was less content visible on the screen.

"It goes too big, 'cos I can't … I have to keep scrolling [participant turns device back to portrait orientation]. Yes, I can read a bit more. I get the title and read a bit of the intro."

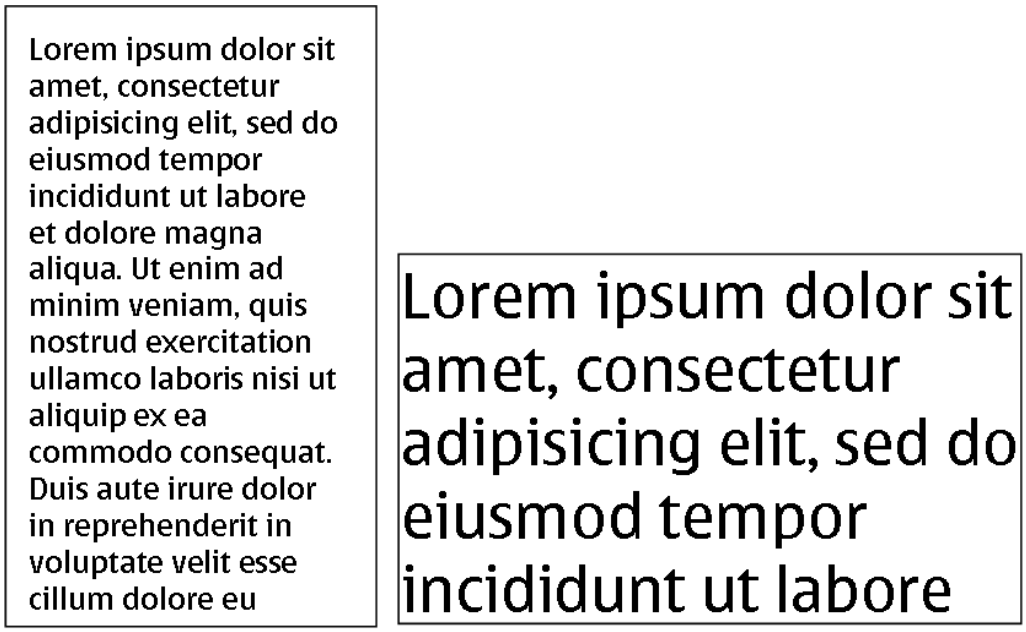


Figure 20 Device orientation change affecting text zoom level on screen. The text area in both orientations is the same.

9.2.1.2 Overview

The ability to obtain an overview of what is immediately before, after or around the piece of text that is being read is an important factor in establishing a context and employing reading tactics with regard to a given text. A good overview of what a page or a website contains will have an immediate positive influence on the perceived need to interact with the interface, as can be seen in the following statements when participants were looking at newspapers in paper format and at the newspaper website on an Apple iPhone:

> *"I think it would be hard to get something big like that front page onto a mobile phone, on a small screen, unless it's the way iPhone is laid out. A sort of small version of the front page, then you can sort of touch and bring up the article."*
>
> *"The text can be made bigger, that's not a problem, but it's just having a smaller section on the screen at one time. Whereas if you have the sheet in front of you, you don't have to scroll down to see, it's all there."*

A sense of immediate context seemed to help participants make decisions about which part of a page or a website they were going to focus on or to

quickly assess if the page had the content they were looking for or enablers for finding specific content from some other page or website. In other words, with overview, decisions about what to read and navigation to and from other parts of text were perceived to be more fluent and efficient. With electronic interfaces and small displays in particular, a poor overview resulted in a negatively perceived need for interaction:

> *"I'd rather it was all in front of me and didn't need to scroll."*

9.2.1.3 Navigation

Navigation within text and content overview are very closely linked and at times almost describe the same phenomenon. However, despite the obvious interdependency, each can be perceived as either positive or negative in its own right, so that overview is one factor and navigation is another. If content overview describes the available content to the reader, then text navigation focuses on text format items that guide the reading process.

> *"I think as long as you've got sort of fixed points on it, like photographs now, you can see where the photograph was when I've read half way down… Gaps between the lines there help you. Suppose it's subconscious you're looking for a two line statement or a part of a report or a one line one there I reading the one line one in the middle, I moved it and read there. Plus you have your finger on it [participant shows how finger helps keeping eye on the line, using an iPhone]."*
>
> *"It's easier if you've got a printed one, just to see where you are. Sometimes online you can scroll too fast or too slow, and can't remember which page you're on."*

Just as a lack of overview resulted in a negatively perceived need to interact with the interface, excessive scrolling or panning interaction is disliked when the text format does not support navigation at optimal visibility and readability settings:

> *"Yeah, I mean, it's…I'd say that'll be the maximum really [participant zooms text so that column width matches display width]. Without*

having to… any bigger and they you'd have to [participant gestures sideways scrolling]. And that's what I don't like."

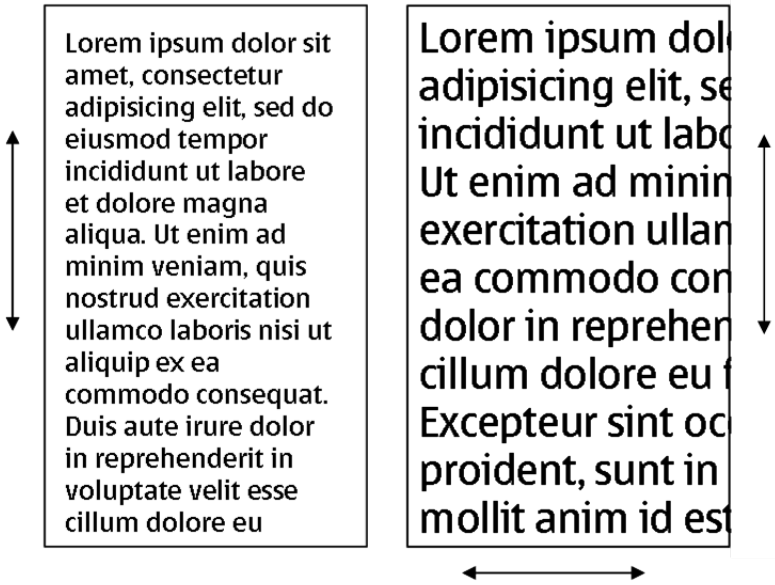


Figure 21 Zooming (magnifying) the text on screen without word wrapping (arrows indicating required scrolling and panning directions).

9.2.1.4 Interaction

A text interface may require interaction in order for fluent reading to be possible. Interaction here includes any actions from the reader for adjusting visibility elements, gaining a better overview or actions taken in navigating within a text. A book or a newspaper is also an interface that requires interaction: for example, pages need to be turned or viewing distance may need to be altered in order to gain optimum visibility. Interaction with conventional, i.e. paper, formats was perceived by interview participants as a *de facto* model to which electronic interfaces were compared.

Interviewer: "Can you think of reading your most typical texts on a mobile?"
Participant: "I doubt it. I'd be constantly having to scroll. Plus it's, the content's really small on the screen for looking at. When you look at the newspaper it's more relaxed and whereas that, you'll need to be sort of concentrating, switch off from all around you. And to get the pictures as well, it's going to take more of your scrolling time. … I might read reports and stuff like that if there was a report I was deeply interested in reading about, that had been again highlighted in the newspaper. But because it's a bit complicated I

wouldn't be particularly interested unless it was a really focused report. But for like normal stories, with the 24 hour news coverage on the telly and computer and papers. I mean between them they've got everything covered I think."

Interaction can also be something the reader may wish to be able to do, or it can be something that is required of the reader. When an interface requires an interaction in order for the reader to be able to even see the text comfortably, the need to interact is perceived to be an irritant:

"There's too much need to scroll."

"On my mobile, I'd have to have a bigger screen. Because it would be a pain, like, having to scroll down all the time."

At times, interaction may be something required by the reader:

"If I need to annotate it I'll print it out, but otherwise I'll just scan through, maybe jot down a few bullet points and read off the screen."

"I just like writing notes on the side."

"[…] in a way I prefer paper, because if you need to show someone something, it's easier to show it, like, the way you've highlighted on the paper or with your notes or whatever next to it, whereas if you're online you have to go down and point it out on the line, like, it's from that line to that line, and then compare your notes that you have here. Where, if it's on paper it's all together, easier to see."

Being unable to make personal notes easily on to a text left many participants wanting to print out the text on paper first so that they could then jot down any comments. The motivation factor in these cases was typically learning-related. On the whole, motivation and the perceived need determined the positive or negative weight of the experience. Wanting to make notes meant that interacting with the interface – a paper page in these cases – was a positive experience (it helped the learning). On the other hand, an interface requiring the reader to adjust any parameters was perceived as a demand and was therefore a negative experience. Thaler and Sunstein (2008) highlight inertia as an element in human decision-making. The tendency has been studied by,

for example, Samuelson and Zeckhauser (1988, in Thaler and Sunstein 2008) when observing human decision making; they found a tendency toward always maintaining the status quo. Many people simply do not change any default options or settings in various systems or interfaces even when doing so would lead to obvious benefits. One such adjustment would be to set the default text zoom level to an optimal size for oneself on a web browser or mobile phone.

> *"Do you ever adjust how the text is on the screen when reading things on websites?"*
> *"No, not really, on the websites no. If I was writing something and the font was like 11 I might put it up to like 13 just to see it better."*

Electronic interfaces have various attributes that determine how comfortable, easy and fluent the interaction is perceived to be. These attributes include interface responsiveness to user actions in general, a device's ability to produce requested content at reasonable speed, interaction style for increasing or decreasing zoom levels, typefaces, contrast, colours and other visibility factors on screens, and interaction style for moving content on screens, typically scrolling or panning. The interaction fluency can typically be tested with various usability engineering methods.

9.2.2 Device factors

The device factors describe factors that provide input to the subjective level of reading and the reading and perceptual processes involved. Interface type refers to the media of text, most typically paper (various subcategories) or electronic (PC or mobile, for example). Interface size refers to the physical size of a page or a display where text is shown. More typically size is an issue with electronic texts rather than with paper. Text format concerns the formatting of text from layout to use of headings, paragraphs or columns. Text format is something that depends on the reading media (paper, computer, mobile device), as seen in section 2.2: content that has originally been designed for one reading medium may change radically in its layout and visual design when it is shown in another medium.

9.2.2.1 Interface type and size

Interface type makes a distinction between paper and electronic, mobile or stationary, printout or paperback book, etc. The interface type affects how, when, where, how long and for what purpose at a time a text is accessed.

Interface size refers to the physical size of the media, be it papers, books, something accessed on the computer or on a mobile device. One dimension of the size definition is the proportions of text and the area it is displayed on: how many words or phrases fit in the given area. The physical size of the whole device is important as well. For example, tabloid size newspapers are sometimes perceived to be more convenient than broadsheets and a big text book is perceived to be more cumbersome than a smaller paperback novel. With mobile devices, typical mobile phone screen sizes (approximately 1.5 - 2 inches) were considered too small for reading, but when participants were shown a larger screen mobile device (3 inch display) with text content, the response and anticipated attitude became more positive.

> *Participant: "Say, the size of some of the newspapers, I like the size of like Sun and that I prefer that to those [participant points to a broadsheet newspaper], they're clumsy and that kind of stuff. Obviously magazine size, that sort of thing, I'm quite happy with that sort of size."*
> *Interviewer: "How about the columns and such? Does that make difference?"*
> *Participant: "Not really, I'm not really fussed on it, I more or less read either. You see that's totally different from work where it's all A4 and it's from one side to another all the way down."*
> *Interviewer: "How is that different from a narrow column?*
> *Participant: "I probably find reading that [participant points to magazines] easier. I don't know why, I just , I don't know why I just do because sometimes, it's not so much to take in each sort of paragraph, do you see what I'm getting at?"*
>
> *"After a while your eyes, like, well with a book it doesn't matter where you are you sit down, but with my laptop you like rest it on your knee, it's more awkward, whereas with a book you can lie down or anything you want and still read."*
>
> *"I think it's just easier to read a hard copy you sort of see everything that's there, it's not all sort of computer options around the side, things like that."*

Interviewer: "Do you use your mobile for web?"
Participant: "Well, it's not the best, you have to scroll down a lot. And it doesn't come up with the pictures, it just comes down with the story headlines, and you don't see pictures, you don't see what the website would look like, which isn't the best, but it's generally only if I'm looking for something or if I want a quick overview of the top three stories. And very rarely, well sometimes I go more often, whether I'm walking the dog or he's on a field and I'm looking at it for ten minutes, but, apart from that I don't really go on the phone a lot cos it takes a long time to get the information and load the pages and read through and keep scrolling down cos the screen's small as well, but it is handy to have it though."

Interviewer: "Do you use your mobile for web?"
Participant: "I wouldn't use internet on my mobile phone. Never. Cos, just never have. Don't see the point of that. Just go down to my laptop."

Interviewer: "Do you use your mobile for web?"
Participant: "No, and probably the graphics. I mean I've got quite a good phone at the moment, but… Well, I just like, the size. Images, if there's any sort of blurriness or if things aren't as bold – I'm not interested, it just bores me. I like to have things plain simple, there, easy to view."

Interviewer: "Do you think you could read things on the mobile?"
Participant: "On my mobile, I'd have to have a bigger screen. 'Cos, it would be a pain, like, having to scroll down all the time."

Interviewer: "Why would you not read on your mobile?"
Participant: "Just not big enough, just too small. The way I'm thinking, not used to it." Interviewer: Is it the amount of text?"
Participant: "Just so small, it's … I'd rather have a bigger screen or a book. But I've never used it so I don't know."

Interviewer: "Would you use the web if you didn't have to worry about the price?"
Participant: "Oh yes, if I had one. And I must say, if I was sitting in a bus, or a train, or I was sitting in a waiting room probably would be easy it would come out of the pocket, and you could catch up with what was going on. And I suppose if it was of interest what you've read you could get the paper and read more on the paper, but you could think well I don't need the paper cos I've read all the headlines. So yes. If I could... if I felt that confident on it, and it gave you the main headlines on it, I could buy it, I could be changed."

"I'd imagine it would be horrendous, unless you had a huge screen. […] Well, yeah, again, I would think it would be quite hard to follow. I know sometimes if you get a long text for example, it takes ages to

scroll down and it's easy to lose track of the last word. Sometimes I have to move back up again, and then move back down again just to make sure I'm on the right stage."

"I doubt it. I'd be constantly having to scroll. Plus it's, the content's really small on the screen for looking at. When you look at the newspaper it's more relaxed and whereas that, you'll need to be sort of concentrating, switch off from all around you. And to get the pictures as well, it's going to take more of your scrolling time. … I might read reports and stuff like that if there was a report I was deeply interested in reading about, that had been again highlighted in the newspaper or … and that [participant refers to a mobile device] was available, I might look at something like that. But because it's a bit complicated I wouldn't be particularly interested unless it was are really focused report. But for like normal stories, with the 24 hour news coverage on the telly and computer and papers. I mean between them they've got everything covered I think. I saw one company started using BBC on their mobile phone."

"I think I just feel more in tune with a paper than I would; probably because of the difference, with newspapers you're just used to the paper. The paper's just nice the way it's there and everything's what you want, and on the phone you have to dig a lot and sometimes you feel like you've missed things, paper's just what you see is what you get."

"I think mobile phone screens are too small, but that's part and parcel cos you don't want to be carrying something… I think it would be hard to get something big like that front page onto a mobile phone, on a small screen, unless sort of the way iPhone is laid out, a sort of small version of the small version of the front page, then you can sort of touch and bring up the article."

In the reading model, then, interface type and size have a high-level effect on how the reading factors are perceived.

9.2.2.2 Text format

How a piece of text is set out on a page – line spacing, column division, use of structure markers such as headings and paragraph division and the existence of any visual elements that allow for faster and more fluent perception of how the information in the text has been organized – are all considered under text format. According to Samuels and Eisenberg (1981) these physical elements allow readers to employ reading strategies based on their existing experience of various types of text. Text format affects perceived fluency in general, but

indirectly also the mobile reading acceptance. The combined effect of text format, interface size and interface type all affect the reading processes at varying magnitudes, depending on the reader's existing knowledge of text types and the physical ability to identify the formats on the interface.

> *"I think there are different structure for different purposes, I would say. The lecture notes maybe tell you like the [main points], and move on to a different point, but obviously a book has to try and keep all together instead of being fragmented cos they wouldn't read as well maybe … not to stop and start again and want everything to flow together like a story. The form should support the purpose."*
>
> *Participant: "I don't really notice. I don't like pictures or anything in books. I don't need that to visualize. I mean biographies, yes I like the pictures, the snaps, the photographs, but I don't like images in a novel 'cos I can visualize from the text.*
> *Interviewer: "Do you have any layout preferences, like with column width and such?"*
> *Participant: "No, I don't have any preference on that. I don't like reading six or seven columns, I prefer to read two or three columns."*
>
> *"You just pick up a book and expect it to be the same really. Definitely. If I go to the library to get a book out, I always look to see what the size of the text is and if it's too small I put the book down. I like big text. I mean not blind text, you know what I mean, eye difficulties, fairly larger text. Not to the full expanse of the page, I like to have a little white, erm, columns [participant points to a book's margins]. I don't know if it's a kind of a mind over matter, you like feel like you're flicking through the pages quicker whereas you wouldn't feel that with smaller text."*
>
> *[Sun reader shown a Guardian]: "This is more difficult there. That layout there, I don't know, a lot writing there, more difficult. That's why I wouldn't buy this, see, the layout [participant selects a page at random]; two main pictures and the rest is text. It's the first thing you see. Like, smaller sections of writing and then you can read it. Instead reading a massive long section.*
> *Interviewer: "So you like it when it's paced out in smaller chunks?"*
> *Participant: "Yes."*
> *Interviewer: "Have you noticed any difference in how fluently you read?"*
> *Participant: "No."*
>
> *"Probably because it's got lots of pictures in it as well. So it breaks up the text quite a bit, there's lots of white gaps which make things easier to read. If it was all just lots of text in one big space. Like the*

Times usually is there's lot's and lots of text it gets a bit tedious and boring and you just tend to skim and not thoroughly read whereas that [participant points to a magazine] you find easier to read."

"Wouldn't bother me. If I wanted to read the book I'll read it. Probably because I type I'd think why have they wasted all this space, but then sometimes it can look neater or tidier for the reasons of doing that but it wouldn't affect me in choosing something I wouldn't pick it because of how it was laid out."

9.2.3 Perceived comfort and fluency

Perceived comfort and fluency were structures that were identified as both instinctive for the participants and also as collective factors that are essentially formed causatively from reading and device factors. The comfort factors were expressed in the participant comments as an immediate preference for something they identified as being used to and familiar with, or as something they had motivation for reading.

Typically perceived comfort would be a strong positive or negative intuitive sense about text on screen with participants not necessarily able to explain clearly why their response was so positive or so negative. The perceived comfort factor affects reading motivation and mobile reading acceptance either by reinforcing existing motivation and attitude towards mobile reading when the perceived experience is positive, or by decreasing motivation (both in general and in terms of attitude towards mobile reading) when the experience is perceived to be negative.

The amount of effort required to achieve an overview and navigate text, ease of use (interaction) and the perceived fluency of reading itself all contribute to perceived comfort in reading. In addition, text format which enables use of reading strategies contributes towards the experience of comfort in reading.

"I wouldn't think simple form, I wouldn't enjoy that as much… I'd rather…well, not simply written though, you still need big words. You can come across a word where you go what's that word and that might make you look up that word. I'd rather it went like that, I

wouldn't enjoy if it was simply written for ease. No. But then I'm probably used to not reading particularly simple things. I'd rather more detail. And more often when you come across the word you don't understand, how it's written in the story line or the newspaper makes you understand what it means."

"I much prefer the, well all things being equal I'd pick that one [more difficult language], it gets your mind turned, rationalized debate. You can either agree or disagree with something but at least you know why you disagree or agree where as The Sun is pretty…basic. … But headlines are good and the breaking news like, but like I say, the heavier reading is more enjoyable."

Participant: "The straightforward language, definitely.
Interviewer: "Even if it takes a couple of more paragraphs to get the point across?"
Participant: "Yes, definitely, it sinks in better."

Interview participants identified a number of elements that for them affected their experience of reading fluency. In the present context, fluency is understood to include both perceived reading speed and comprehension.

"If it's more difficult to comprehend, it slows the reading down."

"[Participant is looking at Guardian newspaper] That uses bigger words, and it makes harder to read, naturally."

On the whole, elements that seemed to affect participants' experience of fluency most were to do with their individual and internal abilities, experience and disposition. The following subsection discusses these individual elements in more detail.

9.2.4 Individual variables in the equation

The interview responses refer to a number of elements in reading that are highly individual – or rather: how large a role these individual elements play in the experience of reading is something that can vary from reader to reader. Participants identified and described what kind of language worked best for them in terms of feeling that they were reading fluently. Individual reading

ability, experiences of text difficulty and a reader's ability to employ reading strategies to match text formats are all factors that have an impact on perceived reading fluency. How motivation is built up, maintained or reinforced, what the actual reading fluency of each interview participant is in terms of reading comprehension and reading speed, how they perceive text difficulty, and how capable they are of employing reading strategies were not issues discussed or described by the participants as such; indeed the impact of these elements would vary from reader to reader.

Motivation expresses itself most strongly in the initial decision to read. When discussing comfortable and fluent texts, the function of reading and the subject matter of the text ranked very highly among the responses. A comfortable and fluent read is something that satisfies personal, internal interest. External motivators are needed as well, such as the need to pass exams or to obtain information for work purposes. Motivation is considered to have a direct effect on perceived reading fluency and on how comfortable a reading experience is deemed.

> *"A good topic is obviously most important. If I see a headline and think "oh, that looks quite interesting" I'll read more, but if it's just a big boring page full of writing and the headline or the content doesn't really interest me then I'll not really read it. "*

On the whole, elements in the general reading experience can loop back to the initial motivation, either reinforcing it or diminishing it, but in the proposed model that loop has not been explored and motivation is discussed in the context of mobile reading acceptance.

Interview participants made some references to comprehension and text difficulty when describing fluent or not very fluent reading experiences.

> *"Some of the newspapers are like, basic wording and you can fly through, whereas, if you get more like the broadsheets they become more like the journals with intellectual content, you pay more attention."*

"Like, the more interpretative stuff is easier to read than the more factual stuff with numbers and stuff. So the stuff that gives you the idea or an opinion and stuff is easier to read and understand."

"As long as it's not too technical, I'm not a very technical person. I mean some of the books I read the crime fiction have quite scientific technical terms, but the terms that they use are pretty much same in all the books so you know what they mean, really. Sometimes there are a few words that come up that I don't know what they mean and I have to go check, sometimes it's just plain English and not technical terms, but generally the things I read are quite similar in the descriptive text so I can understand and get an idea of exactly what it means. Plain English is just perfect 'cos I'm kind of fast reader."

"Definitely, because there's lots of jargon in legal material, lots of long words that you haven't got a clue of what they mean, and so occasionally I might have to stop and look up in the law dictionary. The words that are in legal materials are not necessarily words that are in like Oxford English dictionary but they're strictly legal words so yeah, if you're not kind of fully up to speed with those kind of words you have to stop and look in the dictionary."

Some participants also referred to a sense of progress contributing to perceived fluency. When reading a book, there would be an optimal pace for turning the pages that created a feeling of good progress, and the main contributors to being able to read through pages at this fluent pace were text format and visibility.

"…not as many words on a page, like not, like some books are like lot of words in just two solid blocks of words on the page with text with like small print, makes cumbersome reading. If the text size's bigger it you're rushing through your pages quicker even though it's thicker, mm, and every day sort of reading it I find the way it's broken up, so if you're not like choosing to read it and don't have as much time maybe so you can grab the points and go on looking at something else."

Interviewer: "Do you know why the narrower columns would be better for you?"
Participant: "Maybe it's just to trick people into thinking they're doing more, like you're reading more lines quicker and doing more pages. I think it's just to make you feel more comfortable like reading quicker, obviously the same with like small text takes you a long time to read it; people sort like an average, what you sort of think how many pages you go through."

9.2.4.1 Mobile reading acceptance

Mobile reading acceptance, that is, how likely or unlikely using a small display mobile device for reading continuous text would be, is dependent (based on the interview response analysis above) first of all on a reader's motivation, then on a combination of reading factors. Before these reading factors are even considered, it seems the need for using a mobile device has to be present. Typically, there is such easy access to the more familiar interfaces of paper and regular computers that occasions are rare when text access is required but none of the more conventional interfaces are available. However, it is not difficult to find situations where reading and viewing content on a mobile – mainly websites – is perceived to be useful.

In the following comments participants were asked to consider whether they could think of situations where they might find it handy to use a mobile for text content, presuming the mobile was one they thought was advanced enough in terms of visibility and interactions:

> *"I suppose, if you have like time to kill like you're on a train so you'd do it for as long as you're on the train, just looking around the internet. Better than just sitting there. I could use it for ages. Like, I think eventually you'll get used to it and you'll read quicker."*
>
> *"If I was sitting in a bus, or a train, or I was sitting in a waiting room… probably would be easy. It would come out of the pocket, and you could catch up with what was going on. And I suppose if it was of interest what you've read you could get the paper and read more on the paper, but you could think well I don't need the paper cos I've read all the headlines."*
>
> *"I would say, I don't think that [mobile phone] would replace the paper at the minute, but if I had one available, see, if I had a coffee break and there wasn't a paper and I had something like that, I would quite happily just flick through, even if it was the front page that had half a dozen stories I'd read one or two of them."*

How the various factors grouped under the reading factors category are perceived by readers is influenced by external factors and then by interrelationships within the category itself (visibility affecting overview, interaction affecting navigation, etc.). Any one particular element can be

perceived to be positive or negative or even neutral. As often in usability engineering, if the end user does not notice some feature or interaction, it is usable. After all, we tend to pay more attention to interactions when they do not go right and do not pay them any attention when they are fluent.

In extreme cases, any one of the general reading experience factors can independently affect the mobile reading acceptance so strongly that it determines the attitude. For example, poor graphics quality on a device would, for some users, be enough of a deterrent in itself for them not to consider mobile devices for reading at all; on the other hand, extremely fluent and easy interaction may be enough to encourage some users to overlook overview or visibility problems. Where all general reading experience factors are close to neutral or sum up on the positive side of the experience scale, the mobile reading acceptance will tip to the positive side as well.

9.2.5 Perceived factors in reading fluency: a model

The proposed model aims at organising those factors that contribute to perceived experience of reading fluency and a perceived experience of comfortable reading, and then moves on to factors that contribute specifically to attitudes towards reading continuous text on small displays such as mobile telephones. The discovered factors fall into two main groups that are present in the reading process: reading factors and device factors. Reading factors are concerned with visibility, navigation, content overview and required interaction. Device factors consist of the physical aspects of the reading interface: size and type of the device as well as text format. Figure 22 illustrates the identified factor categories and their interactions.

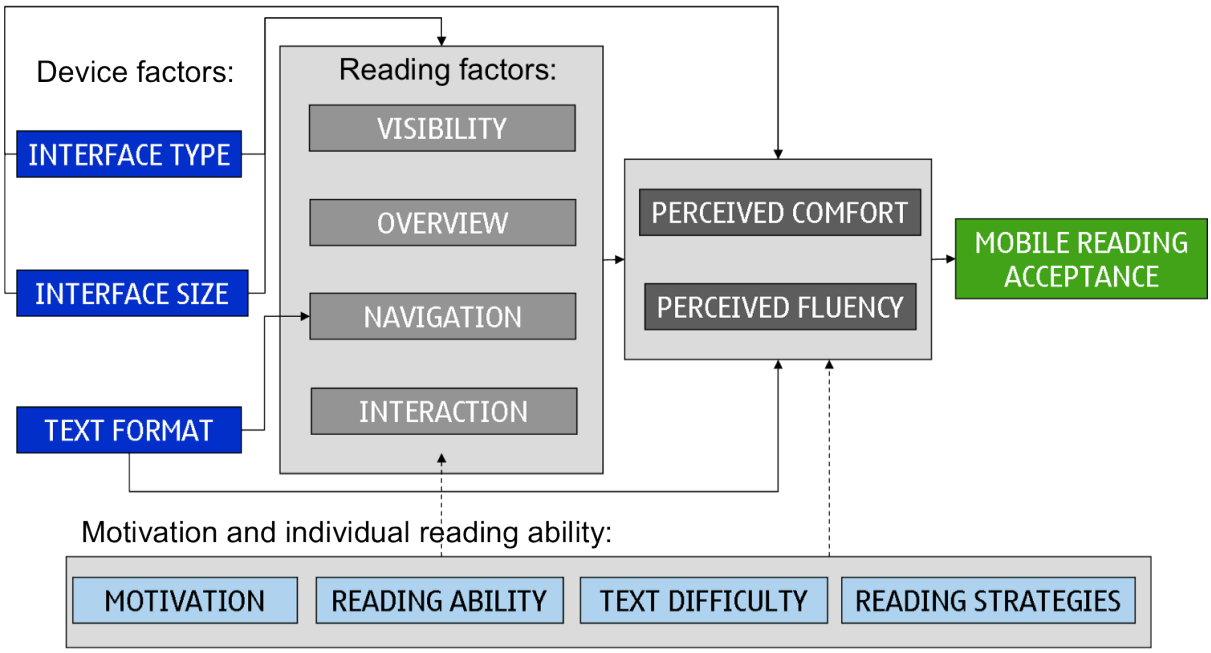


Figure 22 Mobile reading acceptance model.

The factors then contributed towards perceived reading fluency and perceived comfort in reading. In addition to these two main groups, motivation and individual reading abilities were identified as major contributing factors for all reading. These reader-specific factors affect the outcome more as enablers and disablers, affecting the magnitude of impact of some of the other factors. For example, motivation was identified as a major contributor to mobile reading acceptance even on its own. Regardless of the source of motivation, be it a need to revise for exams or to catch up on the latest sports news, motivation is a required element for any satisfactory reading experience.

Technology Acceptance Model (TAM) (Davies, 1989) is another model of factors more generally aiming at predicting end-user behaviour and attitudes towards new technology. The reading acceptance model presented in this study follows the same general themes of the TAM model in terms of external elements affecting the perceived factors (Figure 23), but as a model focusing on a very specific instance of technology use and use case (reading), the model presented here offers a finer definition of affecting factors.

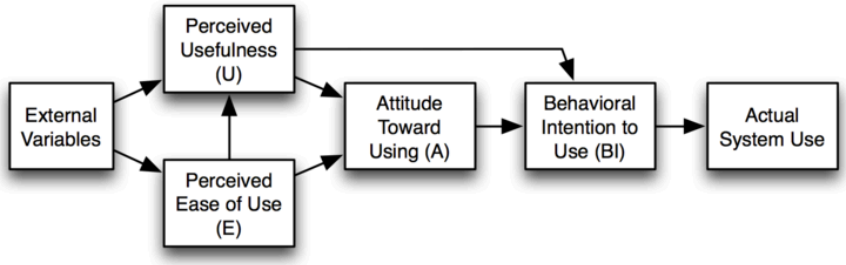


Figure 23 Technology Acceptance Model (TAM) (Davies, 1989).

The device factors appear to contribute to all or some of the reading factors, while reading factors have an impact on each other as well as on the perceived comfort and fluency. More precisely, what type of a device and what its size is affects the perception of the Reading factors, where Text format mainly affects the perception of how well navigation within the content will work for the end-user. Reading factors interact within their own group: for example, visibility has an immediate effect on overview (larger text means less content visible, fewer navigation options and possibly easier interaction in terms of item selection on screen). As the perception of the Reading factors forms, the end-user will build up expectations and opinions regarding the perceived comfort and fluency of reading. The effect the factors have on perceived comfort and fluency and hence mobile reading acceptance can be either positive or negative, depending on the nature and design of these factors and, of course, the state of motivation. Motivation and individual reading ability will play a role in forming perceptions regarding the Reading factors, but more directly these individual elements affect the perception of reading comfort and, particularly, reading fluency.

Linguistic aspects of information design and reading comprehension such as text difficulty, individual reading ability and individual ability to employ reading strategies are considered elements that impact upon perceived fluency at a general level. Comprehension, reading proficiency vs. text difficulty as well as ability to employ reading strategies are unique experiences or abilities for each reader and their impact on reading factors and perceived fluency will vary from

reader to reader. In the scope of the interviews, participants were not asked to evaluate their own individual reading proficiency.

9.3 The mobile reading acceptance model and perceived experience feedback

The mobile reading acceptance model (Section 9.2) indicated that factors affecting attitudes on viewing mobile reading positively or negatively initially fell into two categories: device factors (size and type of interface) and text format related factors. The former (interface size and type) covers the physical attributes of the reading media: size of display or printed item such as newspaper or book, and whether the content is presented in printed or electronic format and is mobile or stationary (mobile phone or desktop). The text format category included items that can be summarized as appearance and layout elements.

In the reading comprehension test (Chapter 5) participants were asked to rate the ease and fluency of reading as well as to provide free comments on these elements after they had completed the comprehension test. The participants were also asked to adjust the viewing conditions on the test device as part of the experiment and again to comment on the reasons for their adjustments. These adjustments and the feedback are analysed in Chapter 7. In the post-test feedback the core elements of the mobile reading acceptance model are clearly present: comments focused highly on the presentation of text (appearance and layout elements), but the limitations (and benefits) of interface size and type were mentioned regularly. For example, the participants noted a difference in their ability to form an overview of the content and its structure between paper and a small display. They also recognized that the smaller interface required more interaction effort (turning pages).

A group of factors within the ‘general reading experience’ category (Section 9.2) that most influenced the participants’ perception of reading ease and fluency, namely visibility, overview, navigation and interaction, also had a

strong presence in participant responses in the reading comprehension test. If anything, the post-test responses would appear to add detail regarding what precisely it is in these items that influences the overall user experience of reading.

Visibility very much concentrated on legibility in the participant comments after the comprehension test, and in many cases improved legibility was a trade-off of navigation and overview: improved legibility of words (larger font) resulted in less content on the mobile screen which in turn for many participants meant a decrease in reading fluency owing to loss of overview, i.e. inability to see full sentences and phrases at once and an increased need for interaction in the form of turning pages. The most frequent adjustment made to the mobile device settings during the reading comprehension test was to increase the size of the typeface. The second most frequent adjustment was to turn the device to landscape. The comments from participants pointed to aiming at a compromise of clearly legible text with a fluent line length (how many words per page) and also trying to maximise the amount of content on the screen. These settings indicated that localized legibility took priority over general overview: smaller typeface in portrait orientation would have offered more content per page, but it would have also demanded more effort in terms of legibility. One typical concern expressed by the interview participants in the present chapter regarded the effort they expected necessary with small screen text: having to squint or otherwise put extra effort into physically being able to read the text.

Prioritising visibility in the adjustments did not, however, mean that test participants would be happy to lose overview facility. Fragmented content in the mobile conditions also made navigation within text more difficult. In paper condition good overview was a highly positive element in the reading experience. The interaction element was also present in these comments where participants described how easy it was to navigate within the text on paper (skim forwards and backtrack to check detail). The need for and habit of skimming and backtracking was also something the interview participants

had identified in their own reading process and they expressed a concern over being able to use these techniques when reading text on a mobile device.

As said, in the mobile reading acceptance model perceived comfort and perceived fluency are bundled together with visibility, overview, navigation and interaction. The model suggests that text format has most influence over navigation and perceived comfort and perceived fluency, whereas interface type and interface size have an overall impact on all of these general reading experience categories. The more fine-grained post-test categorization coupled with the participants' self-assessment on reasons for their final selection of a preferred format is on the same track with the model, though, with some difference in the weight of the relationships. When participants are reflecting on these questions with fresh and immediate experience of both paper and mobile reading, navigation would seem to be more closely related to interface size and type rather than text format, and perceived comfort and fluency emerge in quite equal proportion from text format together with the other reading factors of visibility, overview, navigation and interaction.

In other words, when participants were interviewed, they were not subjected to set tasks and the discussion remained at a reasonably abstract level. Nevertheless, factors that emerged from those interviews are also there in the findings of the present experiment. However, compared with how participants were able to internalize and point to specific elements in the experience after a hands-on trial, the weight of specific categories can, perhaps, be seen in greater detail. The implications for practical interface testing in, say, a commercial setting would be to see that concept testing and early development cycle interviews do indeed point in the right direction, but to extract the detailed findings requires actual exposure to the system and content that is under scrutiny.

9.4 Discussion

The analysis of the interview data identified and organized factors and their relationships that most likely have an effect on consumer decision-making with

regards to reading continuous text on small screens. Visibility (an overall term for all items that affect end-users' ability to see content), overview effect (seeing enough context ahead and preceding the point that is being read at present), navigation (ability to see and understand the structure of the content) and interaction (physical interaction with the reading media) were identified as major elements in contributing to the sense of fluency and efficiency of reading. A positive impression and expectation of these collectively would go towards an end-user deciding that reading continuous text content would be more worth than trouble. However, an extremely negative or an extremely positive impression on any one factor would be enough to tip the scales either way, probably more readily to the negative than the positive.

Samuels and Eisenberg (1981) discuss factors that affect the reading process, as illustrated in Chapter 3. External factors include the physical characteristics of written materials from basic legibility to stylistic formatting of layout. There are also language factors that affect the reading process, such as complexity and communicativeness of language. The language factors are typically measured through word frequency, sentence construction and proposition density. (Samuels and Eisenberg, 1981). The findings from the interview study identified the same factors with the sole difference that visibility elements were categorized under general reading experience because of the personal and subjective nature of defining optimal visibility of text and how the magnitude of the impact on reading fluency is dependent on the reader's individual skill as a reader.

Such physical factors can determine reading strategies. For example, with text formats a reader will recognize a familiar narrow column width when looking at a typical newspaper and thus the reading process will take place within the contextual field typical of the newspaper format. The reading strategy on the identification of the text as a newspaper may be to quickly scan the whole of the front page to spot news items that are interesting, then focus on that section of the page. (Samuels and Eisenberg,1981). Ability to use reading

strategies is seen in the proposed reading model as part of an individual reader's skill set.

A breach of style conventions typically causes a breakdown in the reading strategy and thus will have a negative influence on comprehension and overall reading fluency. Readers will have an existing knowledge base for different text types based on their individual reading experiences, and this experience is a part of the reading process (ibid.). With mobile displays, limitations in screen size seem to compromise either visibility or available context and this, in turn, may prevent the reader from using reading strategies efficiently. Further to that, the limited display real estate can contribute to problems in content design to make content truly look and feel like the familiar stylistic cues that go with a particular type of text. Present smartphone screens are extremely clear, bright and sharp in terms of colours and resolutions but they are, all the same, limited in size.

Changing style to incorporate more efficient communication through higher syntactic complexity and propositional density would very likely break the familiarity with given text types and thus have the said negative effect on the reading experience (ibid.). Even if the denser, more efficient message delivery was to counterbalance the perceivable lack of context and overview, the cost of losing an existing reading strategy might be too high.

A small display will struggle to provide the optimal reading experience, since for maximum efficiency in delivering a message high propositional density would seem to be required at the expense of comprehension and fluency. How well test participants were able to foresee the trade-off between general fluency and the potentially added efficiency of high propositional density is not immediately clear from the interview material, but considering the frequency of statements expressing concern over the small screen being too irksome and awkward for genuinely fluent and smooth reading experience, participants were not spontaneously thinking of acquiring content and information in more condensed form. The participant response, when they were asked to think about various types of text and writing, was instinctively to wait until they had

access to an actual computer (PC) or to have printouts when the content they needed was complicated or important. Obviously, the effect of high propositional density would have to be studied specifically in the future to determine if content design at least in certain areas or directed at a certain readership could benefit from more condensed expression.

Readability is typically assisted by employing various construction methods to promote order and coherence, such as organisational markers (headings, use of outlines, etc.) (Samuels and Eisenberg, 1981). Such construction will group content and information into meaningful units and organise them in a coherent order. For mobile reading, the smaller screen may affect how clearly and how many of these structure markers are visible at any given time, thereby having a negative effect on choosing reading strategies or, overall, on being able to form any expectation of the content ahead. Thus, based on the interview feedback, such organisational markers play a key role in helping to create a positive reading experience: two of the central reading comfort factors coming out of the present interview study being overview and navigation. Participants identified the ability to skim forwards and backtrack in the text as crucial in terms of being able to read fluently and efficiently. They also appreciated the ability to understand the structures and relationships between pieces or segments of content (navigation) and typically organisational markers have been an efficient way of helping readers map out the content they are reading.

To judge by the experiences and expectations identified in the interview study as well as other discussed studies of reading process factors, the small display of a mobile phone seems to place great challenges upon interface and information design of text content, and consumers are able to identify many of the issues. In order for consumers to be able to form a positive first experience of reading on small screens, the perceived comfort and fluency must first add up to at least a neutral, though preferably a positive experience, collectively for all those factors that have been identified as contributing to how acceptable readers find the idea of reading on a mobile phone screen.

There is also the overpowering influence of any single extremely positive or negative factor to consider in the reading experience. An extremely positive or extremely negative experience of even one of the core factors may result in any shortcomings or even actual benefits in other factors being ignored, and the final mobile reading acceptance level may well be determined on the single factor alone. Further study is required into identifying exactly which factors are most likely to produce this effect.

As regards design implications on text content that is intended to be readable on small displays, the four reading factors are ones that a content designer can have an impact on: visibility, overview, navigation and interaction. The range of devices and interfaces is vast, and in many cases the viewed content is developed independently from particular devices. For example, generic web content will be accessed with a plethora of both PC and mobile web browsers. The industry has already recognized the problem of presenting content on multiple platforms and instead of offering content simply on a website, many content providers have now developed applications for the various mobile platforms. These applications, ranging from e-readers to internet banking are designed for specific interface platforms (Android, iOS, WindowsPhone, Symbian, etc.). Websites are increasingly adding mechanisms for identifying what device and even what browser is being used to access the content and often a mobile device specific website is available to mobile users. Thus, content can indeed be tailored for a specific device, but even then the device factors remain the property of the device manufacturer, and content design must then focus on the reading factors alone.

Individual reading ability and motivation are again elements a content designer has little control over: how motivated or interested a reader is in the topic, how familiar a reader is with the language conventions in the case of, say, professional content or the reader's overall level of reading fluency cannot be determined by the designer. The design can only aim at catering for a target audience. As for the device factors, the basic dilemma is the need for more screen real estate whilst keeping devices small and mobile. On the hardware side, quality of graphics and resolution of the screen have an important role. In

addition to this, interaction design (how to move content on the screen, how to zoom content in and out, etc.) contributes strongly to the overall experience of fluency.

The following chapter (Chapter 10) explores the mobile reading acceptance model further by means of a survey study on perceived elements of comfort and fluency in reading with the intention of validating the factor model through quantitative means.

CHAPTER 10

Which discusses a survey regarding what elements of text most influence the experience of fluency and comfort in reading.

Elements of comfort and fluency in the reading experience: a survey study

Chapter 9 describes and reports an interview study that looked into the attitudes and expectations regarding reading continuous text on a (small) mobile device screen. The study produced a model of factors that would appear to affect the perceived experience or the expected experience the most. The target of the interview study was to find factors that contribute to the experience of comfort and fluency in reading, and to build a model of mobile reading acceptance. The model prompts further questions: can the qualitatively studied and formulated model be quantified with a larger sample, and can the latent factors be predicted?

The present chapter describes the process of developing a web questionnaire based on the interview study results, iterating the questionnaire on the basis of expert feedback (DeMaio and Landreth, 2004, in Presser et al., 2004), deployment of the refined questionnaire and, finally, the factor and regression analysis of the results. Interview and survey respondents' views and attitudes can be deemed unstable at times (Foddy, 1994), but the present survey was based on an interview study (Chapter 9) and aimed at quantifying those interview findings. In other words, the present survey study worked as a second round when it came to asking mobile device end-users about their experiences with reading.

10.1 Questionnaire item development

The questionnaire was put together by closely following the mobile reading acceptance model (Chapter 9), basing the questions on those factors that had been identified from interview comments. The categories that emerged from the interviews fell into four main groups that built up to an overall perception of whether mobile device reading was acceptable or not. The four groups were Device factors, Reading factors, Perceived fluency and comfort of reading, and Reader-internal factors. Figure 24 shows the mobile reading acceptance model.

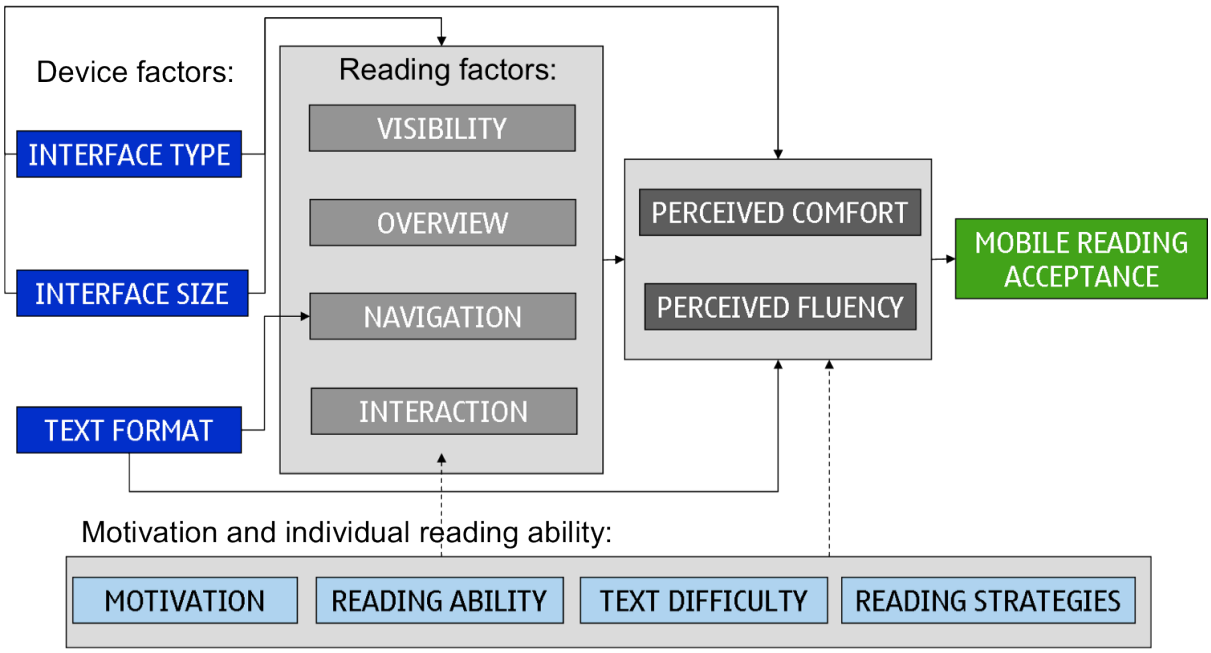


Figure 24 Mobile reading acceptance model.

After basic demographics, the questionnaire proceeded to probe some background information on respondents' familiarity with technology. The purpose of the background questions was to prime the respondents to think about the technologies they use so that they might find it easier to reflect upon those experiences when answering the rest of the questions. Additionally, individual experience and familiarity are potentially factors that affect expectations and assumptions of what reading on a small screen can be like, and could therefore be items that make up the experience and expectations the questionnaire was intended to capture.

After these background items, the questionnaire then focused on perceptions of importance and fluency of various concrete elements in the proposed mobile reading acceptance model: Device factors, Reading factors, Comfort and fluency, Reader-internal factors.

10.2 Factor categories in the questionnaire

The Device factor questions initially focused on simply probing respondents' impressions and experiences of what type of an interface provides the most fluent and comfortable reading experience in general: conventional paper, PC screen or a mobile device screen. Examples were listed with choices to help respondents visualise each of the media better. Further Device factor questions were focused more precisely on text format: layout, appearance and structure markers and their perceived importance. Table 17 presents an example of a Device factor question.

Table 17 An example of a Device factor question.

Question 10: How fluent do you find reading on the following formats? Answer scale: 1 = extremely poor fluency, 7 = extremely high fluency.								
		Poor fluency					Extremely high fluency	
10a	Reading text on paper (for example books, newspapers, magazines, etc.)	1	2	3	4	5	6	7
10b	Reading text on computer screens (desktop or laptop)	1	2	3	4	5	6	7
10c	Reading text on mobile phones	1	2	3	4	5	6	7

Reading factors, such as text format in device factors, were also asked about in more detail, focusing on specific and concrete text elements that affect visibility (including legibility), ability to gain an overview of a text in hand, items that affect navigation within a text and items that affect interaction with the

reading media. The items listed were ones that were discussed and mentioned in the interviews, and the question in the questionnaire was phrased in such a way that respondents estimated how important or unimportant each item was personally to their reading experience. Table 18 presents an example of a Reading factor question.

Table 18 An example of a Reading factor question.

Question 13:
How important or not important is the design and appearance of the following elements to your reading experience? Answer scale: 1 = the design or appearance does not matter at all, 7 = the design or appearance is highly important to the reading experience.

		Very low importance					Extremely high importance	
13a	Size of text	1	2	3	4	5	6	7
13b	Typeface used	1	2	3	4	5	6	7
13c	Use of colour	1	2	3	4	5	6	7
13d	Contrast	1	2	3	4	5	6	7
13e	Resolution (sharpness) of screen	1	2	3	4	5	6	7
13f	Ability to see an overview of the whole content	1	2	3	4	5	6	7
13g	Ability to estimate how much content is available in addition to what is visible at one time	1	2	3	4	5	6	7
13h	Clarity of links and directions to related or further sections of the content	1	2	3	4	5	6	7
13i	Ability to filter visible content (for example, expanding lists)	1	2	3	4	5	6	7
13j	Ability to adjust size of text	1	2	3	4	5	6	7
13k	Ability to adjust the overall zoom level of content	1	2	3	4	5	6	7

13l	Ability to adjust the orientation of a page (landscape or portrait)	1	2	3	4	5	6	7
13m	Ability to move content on screen	1	2	3	4	5	6	7
13n	Ease of selecting links, menus, expanding lists, dropdown menus, etc.	1	2	3	4	5	6	7

Riding on the momentum of considering the importance of the text elements in the questionnaire, the following questions asked directly how important the Reading factors were to the experienced fluency of reading. At this point a number of Reader-internal questions were introduced, such as the importance of personal interest in the topic, existing knowledge of the subject matter and the writing (language) style of a text. Table 19 illustrates a questionnaire item on Reader-internal factors.

Table 19 A questionnaire item on Reader-internal factors.

Question 15:
How much do you agree or disagree with the following statements on reading?
Answer scale: 1 = completely disagree, 7 = completely agree.

		Completely disagree					Completely agree	
15a	Personal interest in subject matter affects reading speed.	1	2	3	4	5	6	7
15b	Personal interest in subject matter affects comprehension of text.	1	2	3	4	5	6	7
15c	Generally, I am a fast reader.	1	2	3	4	5	6	7
15d	Generally, I am an attentive reader.	1	2	3	4	5	6	7
15e	I read frequently and regularly.	1	2	3	4	5	6	7
15f	Different writing styles for different publication types support reading comprehension.	1	2	3	4	5	6	7

15g	I have a broad vocabulary.	1	2	3	4	5	6	7
15h	I read professional specialist texts for my work or studies.	1	2	3	4	5	6	7

The interview results identified that Device factors, Reading factors and Reader-internal factors affected the Perceived fluency and comfort of reading. To approach the relationship from another angle, the questionnaire included a set of questions focusing directly on the relationship, asking simply about the importance of these various factors on the respondents' perceived fluency and comfort of reading. A set of questions focused on the Reader-internal aspects from the perceived influence of personal interest in the subject matter to questions about how the respondents saw themselves as readers: did they consider themselves to be fast, attentive or regular/frequent readers, etc? These questions were designed, like the technology proficiency questions earlier, to indicate if personal experience or skills would emerge as the predicting factors for perceived reading comfort and fluency.

10.3 Genres as factors

In addition to questions based directly on the categories in the mobile reading acceptance model, respondents were also asked to identify what types of texts they might be willing to read on a mobile device screen. These questions were designed to demonstrate if it was the genre or the type of content that determined willingness to read on a mobile device screen.

10.4 Response scale

The responses were collected on a progressive, symmetrical Likert rating scale (Likert, 1932) where respondents would indicate their answers on a scale from one to seven. The wording for the scale was defined per question, and varied between "completely disagree – disagree", "very low importance – extremely high importance", "poor fluency – extremely high fluency" and "not likely – very likely". The scale was set so that the negative, or disagreeing,

answer was one and the positive, or agreeing, statement was seven. A seven-item scale was selected in order to have an equidistant scale. Likert scale is commonly used in social sciences, psychology research and in usability and market research to measure social attitudes or experiences with products and interfaces.

10.5 Pilot study / questionnaire review

Instead of pretesting the questionnaire with a sample group (conventional pilot test), a more time-and-resource appropriate method was deployed: a small group of expert evaluators were asked to complete the questionnaire and return any feedback and observations on any aspects of the questionnaire they found challenging or problematic. The evaluators had expertise in language and communication, end-user and usability research (significant work experience in testing technology end-users), and academic research methods (including questionnaire development).

The feedback collection method was deemed adequate and appropriate on the basis of earlier research into questionnaire and interface evaluation methods. Interfaces (if one is allowed to extend the concept of interface to reach as far as questionnaire design) can be tested and iterated successfully through use of expert evaluators (Nielsen, 1993), and in questionnaire evaluation the use of experts has been seen as an effective revision method in finding problems (DeMaio and Landreth, 2004, in Presser et al., 2004).

10.6 Questionnaire feedback and resulting modifications

Feedback from the expert evaluators pointed to problems with language, where in places it was not possible to understand questions precisely or to relate the answers on the response scale to the question accurately. There were also some minor grammar and typing errors. Further comments were made about the number of questions, pointing out that some items were perhaps queried more than once or that at least the difference between some questions was too subtle. The length of the questionnaire was also

commented on in terms of both the number of questions and how long it took to answer all the questions.

As a result of the expert evaluator feedback, question setting was refined and simplified, leading to fewer questions, although the question groupings remained the same. The items that were redesigned (omitted and/or edited) were detailed questions regarding the importance of individual overview and navigation elements such as use of specific structure markers (headings, paragraphs, etc.). The edited version amalgamated such detail into question 12c, which asks “How important do you find the following items in being able to read text fluently and comfortably? c) Use of headings, paragraphs and other items that indicate structure of content.” Language issues were also addressed so that questions would be clearer and any obvious mistakes in the language were corrected.

10.7 Method

10.7.1 Materials

The revisions to the questionnaire were implemented in the web survey tool where the questionnaire was created in the first place. The web survey content included the ethics information (pre-test and post-test information and de-briefing, a statement that the study had ethics approval from Northumbria University, means of asking for feedback and information on how respondents could pull their data out of the study if they wanted to for any reason at all). The entire questionnaire, including the introductory information and ethics is attached to the present thesis as Appendix 12.6. Table 20 (below) is a summary of the main questions without the corresponding scale items.

Table 20 Question categories in the questionnaire. Each listed question consisted of scale items (marked a, b, c, etc. in the questionnaire) for Device, Reading or Reader-internal items. Demographic and background questions have been omitted from this list.

#	Question
8	In general, how interested are you are in new technologies and electronic devices? Answer on a scale from 1 to 7 (below), where 1 indicates very low interest and 7 extremely high interest.
9	In your opinion, how fluent and competent do you estimate you are in using computers and other digital devices such as mobile phones, games consoles, music players, etc? Answer on a scale from 1 to 7 (below), where 1 indicates extremely low level of fluency and 7 indicates an extremely high level of fluency.
10	How fluent do you find reading on the following formats? Answer scale: 1 = extremely poor fluency, 7 = extremely high fluency.
11	How comfortable do you find reading when the text is presented in the following formats? Answer scale: 1 = extremely uncomfortable, 7 = extremely comfortable.
12	How important do you find the following items in being able to read text fluently and comfortably? Answer scale: 1 = very low importance, 7 = extremely high importance.
13	How important or not important is the design and appearance of the following elements to your reading experience? Answer scale: 1 = the design or appearance does not matter at all, 7 = the design or appearance is highly important to the reading experience.
14	How important are the following elements to you in creating a fluent reading experience? Answer scale: 1 = very low importance, 7 = very high importance.
15	How much do you agree or disagree with the following statements on reading? Answer scale: 1 = completely disagree, 7 = completely agree.
16	How important are the following elements to you in deciding to read or not to read a piece of text on a mobile phone screen? Answer scale: 1 = very low importance, 7 = very high importance.
17	How likely would you estimate it would be for you to read the following types of text on a mobile device screen. Answer scale: 1 = not at all likely, 7 = very likely.

10.7.2 Participants

Participants were recruited to take part in the survey by snowballing a request through social networks (Facebook with the researcher's personal contacts and user experience special interest groups at the researcher's full-time employment) as well as by inviting psychology undergraduates at Northumbria University to take part in the survey.

10.8 Results

10.8.1 Data handling

All survey data was handled in SPSS. The data were examined for internal consistency by means of Cronbach's alpha, which was .881, indicating high internal consistency of the scale (Henson, 2001). Examining the item-total statistics to see how removal of any questionnaire item would affect the coefficient did not point to any of the 59 items lowering Cronbach's alpha (the alpha would remain between .875 and .884). Accordingly, none of the items were removed based on this reliability analysis. The web form did not allow for unclear responses on the Likert scale (responses between two scores).

10.8.2 Sample description (participants)

The survey received overall 174 responses. For the purpose of limiting language-related variance in how reading and text is experienced on small displays the participants were asked to identify if their first language was English or 'other'. From the total 174 responses all those who answered 'other' were filtered out, leaving 150 responses for analysis. It should be noted that the data will show that not all 150 participants always answered all the set questions.

The majority of respondents (46%) came from the age group 18-24 years of age. The 18-24 and 25-30 age groups made up over 50% of the respondents in terms of age. A substantial majority of respondents (69.6%) were female. Age, gender and first language were asked for in questions 2, 3 and 4 in the questionnaire.

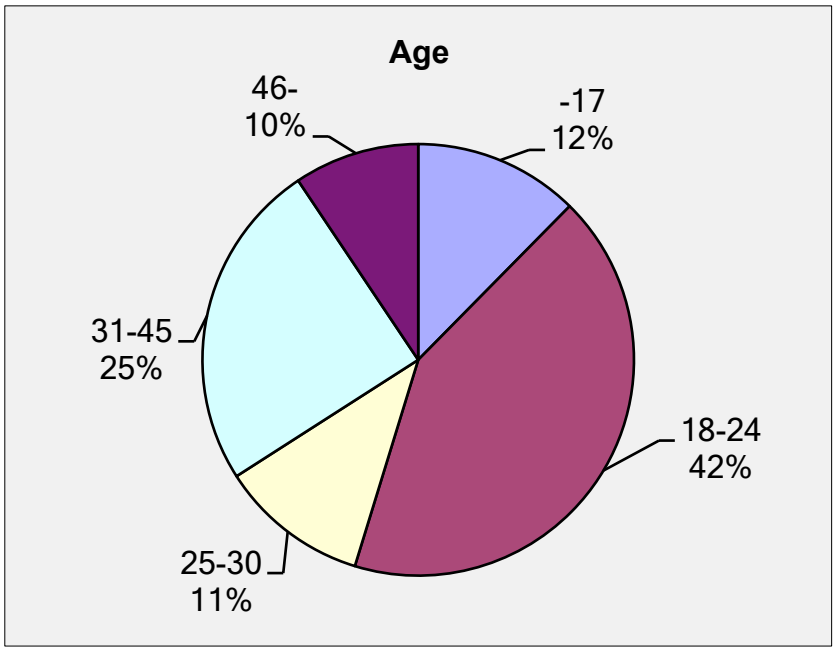


Figure 25 Distribution of age groups in the sample.

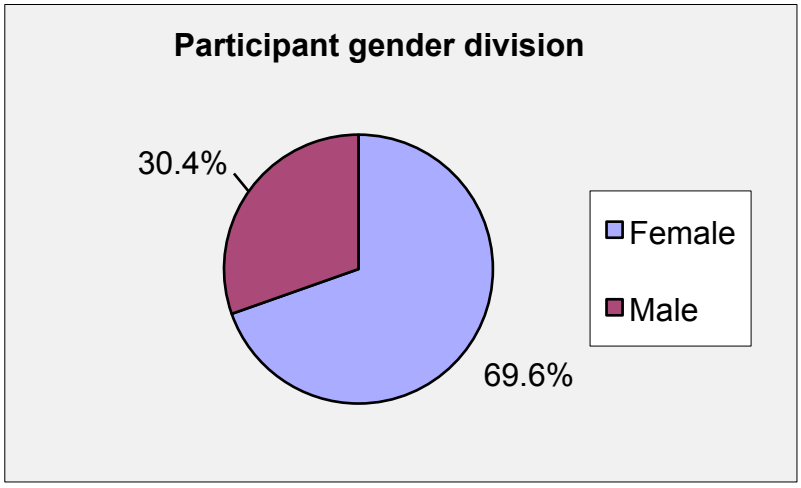


Figure 26 Participant gender division by percentage.

Questions 5 – 9 focused on respondents' technology use background and self-assessment of their computing and technology proficiency. As explained earlier, the purpose of these questions was to help respondents into thinking about their technology use and also to add experience as a factor in the data. Additionally, the responses are used here to form an overall picture of the sample build-up. The question allows some insight into the level of technology orientation within the sample: low technology involvement may affect

acceptance of new technologies and new interaction methods and the involvement level should therefore be taken into consideration when drawing conclusions from the data.

On the whole, the sample who answered the question (N=147) reported using core smartphone features quite broadly on their mobile devices: SMS, MMS, camera, games, e-mail and social networking applications were used by over 50% of the respondents. A clear majority (72%) of the sample also said they had had their current mobile phone for more than six months. Respondents were asked about their current phone in order to ensure that lack of experience with their current phone would not affect the assessments participants were asked to complete later on in the study. Respondents also reported using computers (desktop or laptop) daily (97%).

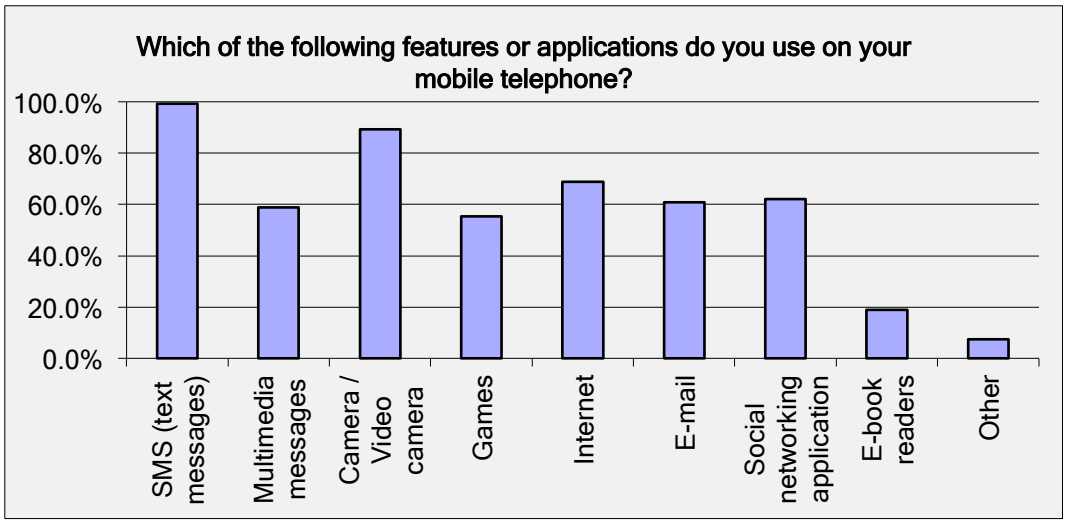


Figure 27 Use of typical mobile device applications.

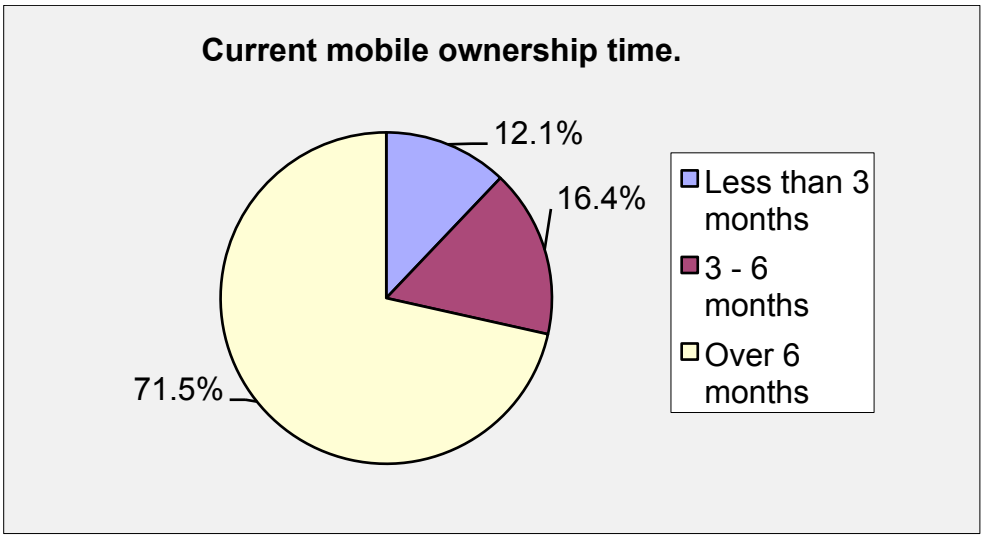


Figure 28 Rough division of how long participants have owned their current mobile phone.

On average the participants considered themselves to be above the mid point on the scale when it came to both interest in technology and computing fluency. The mean for technology interest, asked for on a scale from one to seven, was 5.04 (median was 5.0 and standard deviation 1.5) and the mean for computing fluency was 5.5 (median was 6.0 and standard deviation 1.26). The responses cannot be taken as an indication of participants' actual skills or of their level of knowledge regarding the latest innovations in consumer electronics, but they do illustrate the respondents' general attitude towards technology and computers: a positive response (reasonable interest and above average self-assessment on skills) is taken as an indication of a low level of apprehension towards technology and that respondents are likely to accept and try new innovations.

It would, therefore, be reasonable to summarize the survey respondents as young smartphone users who are also relatively confident with their computing and technology skills.

Table 21 Descriptive statistics on participants' self-assessed level of interest in technology and how fluent they consider themselves to be in computing.

	Level of interest in technology	Computing fluency
Mean	5.04	5.50
Median	5.00	6.00
Mode	5.00	6.00
Std. Deviation	1.50	1.26

10.8.2.1 Initial attitudes and expectations regarding reading on paper, on a computer (PC) screen and on a mobile device screen

The descriptive statistics for perceived fluency (Table 22) indicate that participants (N=140) found reading fluency on paper and on a computer screen very similar, but the mean score drops for mobile phone screens. For perceived comfort, however, the paper format received a similar mean as for fluency (close to 6.5), whereas the mean for the computer screen is just below 6 and for the mobile phone screen is close to 5.

Table 22 Descriptive statistics for perceived comfort in reading on paper, computer screen and mobile phone screen.

	Paper Format		Computer Screen		Mobile Phone	
	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median
Fluency	6.56 (0.82)	7.00	6.26 (0.95)	7.00	5.66 (1.50)	6.00
Comfort	6.46 (1.05)	7.00	5.69 (1.32)	6.00	5.09 (1.70)	5.00

10.8.2.2 Factor analysis

Principal component analysis with Varimax rotation initially suggested a 16-component structure. The initial analysis followed used the Kaiser Criterion, extracting all factors with eigenvalues greater than 1.0. However, the scree plot and the initial overview of the 16 components suggested a solution with fewer components to be a more valid solution. The initial analysis followed the convention of extracting all factors with eigenvalues greater than 1.0.

Costello and Osborne (2005) recommend considering all component loadings greater than .3, but Stevens (2002) points out that the significance of loadings is dependent on the sample size. Therefore, with a sample size of 150 a more conservative cut-off is appropriate, and as such loadings greater than .5 are considered here. Kaiser-Mayer-Olkin measure of sampling adequacy in the present study was .535, which can be considered acceptable (Kaiser, 1974) and Bartlett's test of sphericity was significant, X^2=3716.760,' df=1711, p<.001; together these measures suggest the data are appropriate for principal component analysis.

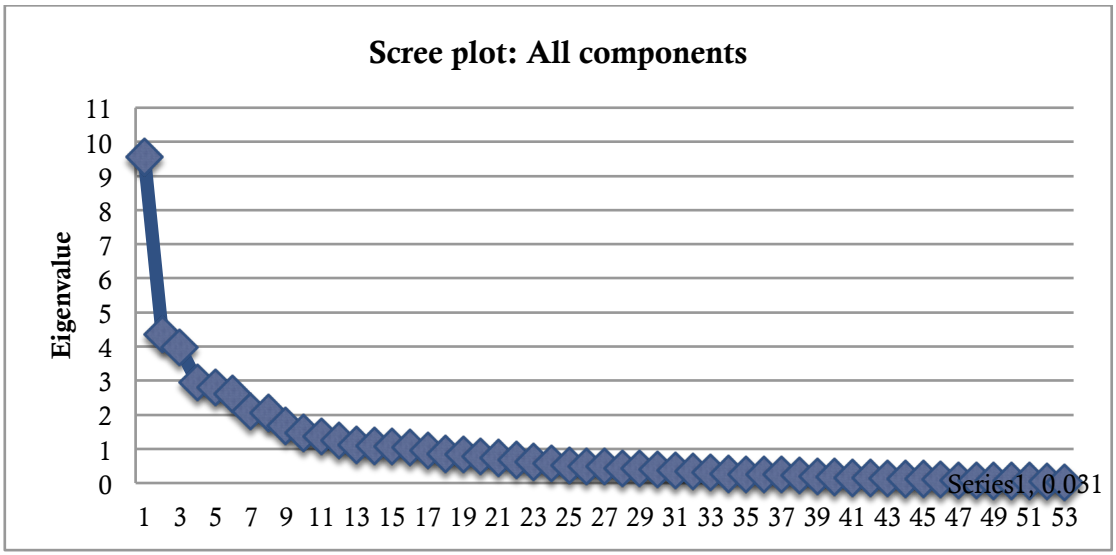


Figure 29 Principal component analysis scree plot for all components.

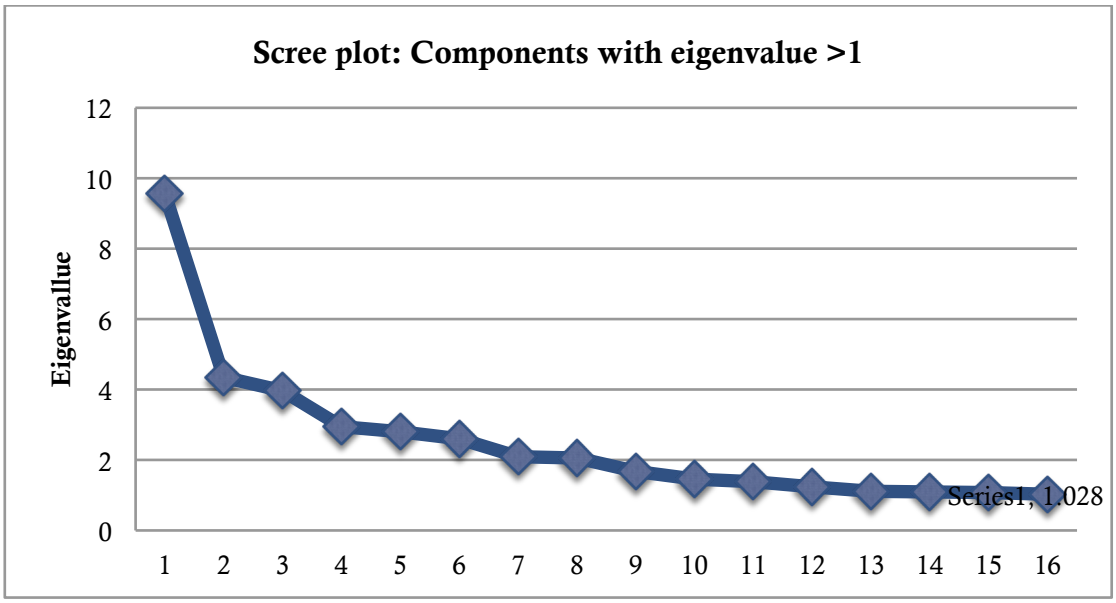


Figure 30 Principal component analysis scree plot for components with eigenvalue greater than 1.

Looking at the scree plot with all components (Figure 29), it is hard to see a clear breaking point in the curve. When the scree plot is limited to only the components that have eigenvalue greater than 1 (Figure 30), the elbow is still ambiguous. The most dramatic break in the data appears at the second component, but the trend of the curve is still clearly down until approximately the third or the fourth component. Costello and Osborne (2005) suggest running the analysis several times, manually extracting not just the number of components suggested by the scree plot but one above and below it as well, and comparing the rotated component loadings, finding the one with the clearest structure. A clear structure would be indicated by a component with item loadings above the designated cut-off and components explaining more than three items. In the present study, as said, loadings >.5 are considered significant.

The scree plot does not offer a very simple and obvious break point in the data, but the point where the components start to level out appears to start at approximately factors 3 or 4. Running the principal components analysis again twice by limiting the number of factors to be extracted to three and four already provides a better fit for the data in both cases than the 16 factors extracted earlier using the Kaiser criterion. The four factor solution explains 38% of total

variance, all extracted factors contain three or more items with a factor loading > .5, and each item loads uniquely onto a single factor. A three-factor solution explains 32.7% of the total variation and also yielded more than three items with loading s >.5 for all three factors, all loading uniquely onto a single factor. Examining the items in each factor thematically, the four-factor solution returned a more coherent set of items in each factor, whereas the three-factor solution bundled a more broad scope of items per factor – particularly for the first factor. Based on the amount of variance explained by the solution and the internal cohesion, the four-factor model was selected.

Checking the internal reliability of the items in each factor in the four-factor solution using Cronbach's alpha showed that removal of all but one item in factor 4 (question 13i) would lower the Cronbach's alpha. The said item's item-total correlation was .346 and as such the item was removed and factor analysis was repeated. The other three factors did not have items that would affect Cronbach's alpha either way if removed. A new factor analysis was performed, and the items in each factor were explored thematically. No further omissions seemed necessary.

The Kaiser-Mayer-Olkin measure of sampling adequacy was .552, which can be considered acceptable (Kaiser, 1974) and Bartlett's test of sphericity was significant, X^2=3601.815, df=1653, p<.001. This third factor analysis also returned meaningful cohesion thematically. Factor 1 contained the largest number of items relating to being able to see content and adjust the device to improve visibility, content navigation and perceived ease and fluency; the factor explained 16.2% of the variance in the data (see Table 23 for full factor loadings). Factor 2 items, explaining 8.4% of total variance, centred around respondents' perceived reading proficiency and activity, and their experience of affluent and comfortable text formats (paper and PC/laptop). Factor 3 contained 7 out of the 9 listed genres (types of content) that respondents would be willing to read on a mobile, and explained 7.8% of total variance in the data. In factor 4 (covering 5.6% of the total variance) the items focus very strongly on legibility and control over legibility. The factors were labelled *Visibility, Interaction and Navigation* (factor 1), *Perceived reading proficiency*

and format familiarity (factor 2), *Willingness to read on a mobile* (factor 3), and finally *Perceived legibility* (factor 4).

Table 23 Four factors on the comfort and fluency in reading questionnaire, and the factor loadings on the relevant factors.

Item number	**Items**	**Factor loading**
	Factor 1: *Visibility, Interaction and Navigation*	
13m	How important or not important is the design and appearance of the following elements to your reading experience? • Ability to move content on screen	0.558
14b	How important are the following elements to you in creating a fluent reading experience? • Amount of content clearly visible on the screen	0.547
14c	• An overview of content (even if details are not fully visible)	0.679
14e	• Ease of moving the contents on the screen	0.649
14g	• Existing knowledge about the topic / subject matter	0.603
14h	• Style of writing (how complex or simple the language is)	0.546
14i	• Length of sentences and phrases	0.502
14k	• Electronic content looking the same as if it was in print format (for example, newspaper style on a website is similar to what the actual paper newspaper looks)	0.549
16c	How important are the following elements to you in deciding to read or not to read a piece of text on a mobile phone screen? • The feeling that reading is fluent.	0.694
16d	• The feeling that reading is a comfortable experience.	0.688
16f	• Familiarity with the topic/subject matter.	0.641
	Factor 2: *Perceived reading proficiency and format familiarity*	
10a	How fluent do you find reading on the following formats? (paper, PC screen, mobile screen) • Reading text on paper (for example books, newspapers, magazines, etc.)	0.734
10b	• Reading text on computer screens (desktop or laptop)	0.616
11a	How comfortable do you find reading when the text is presented in the following formats? (paper, PC screen, mobile screen). • Reading text on paper (for example books,	0.683

	newspapers, magazines, etc.)	
15c	How much do you agree or disagree with the following statements on reading? • Generally, I am a fast reader.	0.557
15d	• Generally, I am an attentive reader.	0.729
15e	• I read frequently and regularly.	0.644
15g	• I have a broad vocabulary.	0.627
	Factor 3: *Willingness to read on a mobile*	
17a	How likely would you estimate it would be for you to read the following types of text on a mobile device screen. • Latest news headlines.	0.565
17b	• Newspaper website.	0.606
17c	• Novels.	0.56
17d	• Newsgroups or blogs.	0.71
17g	• User guides.	0.612
17h	• Reports or other professional communications.	0.757
17i	• Learning materials.	0.613
	Factor 4: *Perceived legibility*	
13e	How important or not important is the design and appearance of the following elements to your reading experience? • Resolution (sharpness) of screen	0.503
13j	• Ability to adjust size of text	0.556
14a	How important are the following elements to you in creating a fluent reading experience? • Visibility of text	0.601

The item number in Table 23 refers to the question title and the specific answer item respondents rated on the questionnaire. The full questionnaire is attached as Appendix 12.6.

Table 24 Summary of factor labels.

Factor #	Name
1	*Visibility, interaction, navigation*
2	*Perceived reading proficiency and format familiarity*
3	*Willingness to read on a mobile*
4	*Perceived legibility*

10.8.2.3 Regression

A hierarchical multiple regression was used to determine if the factors 1, 2 and 3 (Table 24) could be used in predicting Perceived legibility (factor 4 in Table

24). Initial bivariate correlations showed that *Visibility, interaction, navigation* had a significant positive correlation with *Perceived legibility* (r=.391, df = 93, p<.05), *Perceived reading proficiency and format familiarity* had a significant positive correlation with *Perceived legibility* (r=.248, df = 93, p<.05), and also *Willingness to read on a mobile* had a significant positive correlation with *Perceived legibility* (r=.187, df = 93, p<.05). The results of the regression analysis are reported in Table 25. R^2 for Model 1, predicting legibility from *Visibility, interaction, navigation,* was .153, accounting for 15.3% of the variance in *Perceived legibility*. The addition of the second factor, *Perceived reading proficiency and format familiarity*, was .031, and the addition of the third factor, *Willingness to read*, was .020. Together, factors 1, 2 and 3 account for 20.4% of the variance in *Perceived legibility.*

Table 25 Hierarchical regression analysis predicting Perceived legibility. N=93

Dependent Variable		β	t	R^2	ΔR^2	ΔF
Model 1				.153	--	16.385*
Perceived legibility						
	Visibility, interaction, navigation	0.367	4.048*			
Model 2:				.184	.031	3.428*
Perceived legibility						
	Visibility, interaction, navigation	.356	3.672			
	Perceived reading proficiency and format familiarity	0.180	1.852			
Model 3:				.204	.020	2.264*
Perceived legibility						
	Visibility, interaction,	.337	3.472			

navigation		
Perceived reading proficiency and format familiarity	.183	1.897
Willingness to read	.144	1.505

Analysis of variance (ANOVA) shows that the regression of model 3 is significant: R = .452, F(3,92)=7.599, p < .001. The t-tests of the beta weights show that *Visibility, interaction, navigation*, make a significant contribution to regression t(92)= 4.048, p. < .01. However, *Perceived reading proficiency and format familiarity* and *Willingness to read* do not lead to any significant fall in R^2 if they are left out of the regression analysis, and so do not make a significant contribution in the presence of the other predictor, only resulting in a change in R^2 of 5.1%.

It therefore appears that *Visibility, interaction, navigation* can be seen as a modest predictor of *Perceived legibility.* The other factors did not make a significant contribution to the prediction of *Perceived Legibility*.

Table 26 Bivariate correlations between the four extracted factors.

	Perceived legibility	*Visibility, interaction, navigation*	*Perceived reading proficiency and format familiarity*	*Willingness to read*
Perceived legibility	---	0.391*	0.248*	.187*
Visibility, interaction, navigation	.319*	---	0.191*	.127
Perceived reading proficiency and format familiarity	.248*	.191*	---	.003
Willingness to read	.187*	.127	.003	---

* significant at p<.05

10.9 Discussion

The main aims of the survey were to quantify the mobile reading acceptance model created on the basis of a qualitative interview study, and to develop a questionnaire that would work in predicting acceptability of reading on mobile devices. The factor and regression analysis of the collected survey data provides a better understanding of what elements are in a central position when deciding whether or not to read content on a mobile phone.

Four meaningful constructs were extracted through factor analysis of the data. Formed from a relatively large question set the four factors focused on visibility and the ability to control it, familiarity with the format and respondents' perceived reading proficiency, indication of what type of content would be all right to read on a mobile, and the importance of clear legibility. The first factor included a vast array of items on information accessibility that can, on closer inspection, be further categorized into visibility (namely, legibility and overview), interaction (manipulating the visibility condition and the interaction required by content navigation) and (content) navigation. The number of items in the build-up of this factor emphasizes the interdependency of visual design and interaction elements in user interface design: it may not be possible to focus on just one element and improve that without paying for it in reduced quality in another; for example, improving overview creation at the cost of text legibility, or optimizing text legibility at the cost of overview and navigation.

The second factor, format familiarity and perceived reading proficiency, suggests that internal and reader-specific variables will always have a high influence on the overall acceptance of reading on a mobile device: perceived competence and recognizing the text format. Paper and desktop computer screen (PC/laptop) formats were grouped in this factor together with items that described respondents' perception of their reading proficiency.

The third factor seemed simply to reflect on a selection of content types that perhaps matched the sample group's interest at the time. The factor items were fully collected from one set of questions that listed various content

examples from newspapers to user guides. No particular item type stood out and the conclusion to be drawn here may be that content types (genres) cannot perhaps be used as predictors as such, but that content and interface designers need to understand their target audience and appreciate their content preferences when producing interfaces and services. The fourth factor had the lowest number of items, which were tightly focused on visibility or, even more specifically, on legibility: screen resolution, general visibility of text and ability to adjust the size of the text on the screen.

The four factors correspond to the mobile reading acceptance model: Factors 1 and 4 (*Visibility, interaction, navigation* and *Perceived legibility*) straddle the Device and Reading factors categories, and Factors 2 and 3 (*Perceived reading proficiency and format familiarity*) correspond with the reader-internal factors in the acceptance model. A closer look, however, indicates that some factors inside the Device, Reading and Reader-internal categories seem to have a more central role than others: for example, size or type of the interface does not seem to stand out from the factor analysis as much as the more generally functional requirements of clear visibility, possibility of gaining an overview of the content and easy interaction so that these elements are under the end-user's control. Further research into the weight of the various identified factors would be required in order to identify and define a core set of factors that have the strongest influence on mobile reading acceptance.

Regression analysis took all four factors and aimed at finding out if *Perceived legibility* could be predicted from the other three factors. The results suggest that *Visibility, interaction, navigation, Perceived reading proficiency and format familiarity* and *Willingness to read on a mobile* had significant correlations with and the R^2 showed that 20.4% of the variance in *Perceived legibility* is explained by the three factors together. However, *Visibility, interaction, navigation* was the only factor that had a significant contribution to *Perceived legibility,* which in itself accounts for 15.3% of the variance in *Perceived legibility*.

On this analysis alone it is not feasible to propose that the developed questionnaire is functional as an industry evaluation tool as such, but with further analysis and development it should be possible to iterate the various items and produce a finer-grained factor model that would pinpoint a functional level of predictors for estimating further perceived factors that affect end-users' decisions on whether to read continuous text on small screens. The present study proposes that a questionnaire focusing on appearance and visual design of content and interfaces could be developed for the purpose of estimating end-user experience on interface visibility and clarity. Across the experiments and their results in the present thesis the prominent themes have been those of visibility, overview, navigation and interaction as well as ability to control the viewing settings of a mobile device. The results from quantifying the qualitative factor model in the present chapter do validate the said model and confirm which elements of the reading process most affect the perceived experience of reading.

CHAPTER 11

Which presents a summary of the studies in the present thesis and discusses their contributions to both industry and research.

General discussion

The present discussion chapter of the thesis draws on the findings from the six studies on reading on small screens and on the corresponding perceived experiences, highlighting the core contributions these studies make to our knowledge and understanding of reading on small screens. The present chapter discusses these findings in the context of the research questions and aims in relation to both the reading performance and also the perceived experience of reading on a small screen. It also includes a discussion section on the findings, specifically from the perspective of content creation and mobile interface design: how can the knowledge gained be used to improve or help improve creating, handling and presenting text content on a mobile device? Finally, the present chapter also discusses the findings in the context of the literature considered in chapters 1 - 4 with the intention of reviewing how the output of the present thesis adds and contributes to knowledge of reading on small screens.

11.1 Research aims

With more and more of our communications and information handling is taking place on mobile devices, the presentation of that information and of those communications is a relevant topic for HCI research in general. The present research studied a small sector in this mobile device interface HCI field by

looking into what reading is like on a mobile device and what the perceived (subjective) experience is like. The goal was to increase knowledge and understanding of reading performance and reading experience on a small screen, to compare these two aspects, and to attempt to offer useful tools and heuristics to the mobile interface and content design community.

The following list is a summary of research aims and research questions:

Aim: to gain further understanding and knowledge of how the limited screen size of mobile devices affects reading performance and perceived experience of fluent and comfortable reading, by means of the following questions:

1) Does good reading performance also mean satisfaction in terms of experienced reading fluency and comfort, or do the two differ?
2) Does reading format (mobile device vs. paper) affect reading performance?
3) Will the knowledge gained about perceived experience offer practical quality evaluation support and tools for (mobile) user interface and content design community for design iteration purposes?

Reading performance was measured in two ways: in a reading comprehension test and in an experiment on legibility on a small screen. In addition, the presentation preferences for achieving a fluent and comfortable reading interface were observed by allowing the participants in the reading comprehension test to adjust the interface settings to their personal preferences. Perceived reading experience was observed and measured through, initially, interviewing mobile phone users about their experiences, expectations and attitudes regarding reading in general and on a mobile device. The interviews were complemented with a web survey that aimed at quantifying the interview results and at also exploring whether such user feedback could present latent models and structures that would in any way make it possible to predict user experience in mobile reading. Qualitative methods were employed to further understand what elements most affect the perceived reading experience. The reading comprehension test and the

legibility experiment also included a questionnaire section for the perceived experiences on reading comfort and fluency. Finally, as one last look at reading on a mobile in a real environment, critical incident technique was applied to a sample of user experience professionals to collect actual experiences (cases) with mobile reading and potential problems associated with it.

11.2 Summary of studies and main findings

The experiments in the present thesis measured reading performance and the user experience of reading on small screens. The studies employed both quantitative and qualitative methods. The experiments have been organized into six chapters in the present thesis.

1) Chapter 5: *Reading comprehension test on paper and two mobile device conditions*. The experiment was a modified Cloze procedure with 45 participants.
2) Chapter 6: *Word search on small display: legibility in mobile context*. In the experiment 30 participants searched for a target word from a group of similar-looking words.
3) Chapter 7: *Appearance and layout adjustments: optimizing reading on a mobile device*. The 45 participants in the comprehension test adjusted the viewing conditions of the mobile device as part of the final text condition, and a grounded analysis was performed on the adjustments.
4) Chapter 8: *To read or not to read: some real-life examples of when reading on a small screen has not been the best solution*. A modified critical incident study with a small group of user interface and interaction design professionals on situations where they had stopped a reading task in real life. Their descriptions of the situations were then analysed to establish whether these real-life experiences from actual situations mirrored the experiences, expectations and attitudes that respondents and participants in the other studies had expressed.

5) Chapter 9: *Perceived comfort and fluency factors in reading, and their bearing on the acceptability of reading on small mobile displays*. The study was an interview study with 12 participants, discussing their present reading habits and their expectations and attitudes towards reading their typical materials on a moblle phone screen.
6) Chapter 10: *Elements of comfort and fluency in the reading experience: a survey study*. 150 respondents answered a web survey on comfort and fluency factors in reading in various formats. The survey was based on the interview study presented in the previous chapter. The study then aimed at identifying any latent mobile reading factors, and also at identifying the possibility of predicting perceived reading experiences or attitudes from these factors so that the questionnaire or parts of it could be used at interface and content design iteration as an evaluation tool.

The following sections summarise the experiments that focused on reading performance. The results provided answers to the research questions regarding performance vs. perceived fluency and reading comfort, differences in reading performance between different text formats (paper or mobile) as well as offering further insight into the practical factors that influence the perceived reading experience.

11.2.1 Comprehension test

45 participants were tested in a one-on-one test in which passages of text were presented to them on paper and on a mobile device. Participants were also allowed to adjust the viewing settings on the mobile device in the final part of the experiment. The comprehension test was a modified Cloze procedure and the results were compared for significant differences in the number of correct answers in each condition. In addition, time on task was monitored and compared between the conditions as well. After the comprehension test part of the experiment, the participants were asked to rate their reading experience on a Likert type response scale and eventually the

performance results (correct answers and time on task) were compared with the experience ratings.

11.2.1.1 Summary of results: comprehension test and experience ratings

Table 27 below summarises the significant differences in the reading comprehension test scores and in the subjective ratings on the experience, as well as indicating the preferred format.

Table 27 Summary of test scores and perceived experience ratings.

	Paper	Mobile	Adjusted mobile
Correct answers	No significant difference		
Time on task			Significantly better than paper and default settings mobile.
Subjective ratings: ease of task	Significantly better than mobile or adjusted mobile.		
Subjective ratings: fluency of reading	Significantly better than default settings mobile		Significantly better than default settings mobile.
Preferred format	(paper)		

The comprehension test results (correct answers in the Cloze test) did not point to any significant difference between paper and mobile device. Time on task showed a significant difference between the two mobile formats, but the result is open to criticism in terms of learning effect. At the same time the perceived assessments of ease and fluency showed there was a significant difference between paper and mobile formats. Participants' preferred format in the test was paper. In open-ended comments participants were also able to point to elements that made a difference for them in terms of reading experience. The results of the study, then, echoed those of Bernard et al. (2003) and Dillon, Richardson and McKnight (1990) insofar as they showed

that while basic performance within certain parameters may not differ significantly, participants have nevertheless been in a position to identify differences and internalise what elements make a difference in their user experience.

What can be seen from the summary of the analysis of variance for the various scores collected in the experiment is that more factors than a single performance score alone can determine subjective response to a system or an interface. Assuming that the better time-on-task with an adjusted mobile reflects the effort required to complete the test, and with reference to the open-ended comments collected with the subjective ratings stating that paper was considered familiar and provided a better overview of the content, it is probably not surprising that participants showed a preference for these two formats in the subjective ratings on ease and fluency. Having been able to complete the comprehension part of the test at the same level on the default settings mobile as on the other two formats did not help the default settings mobile in the preference ratings.

11.2.1.2 Comprehension test: Modified Cloze procedure

The focus of the reading comprehension test was to establish if there was a difference in comprehension between reading on a mobile device and on paper and, further to that, if there was a significant difference in the comprehension result between the two mobile device text formats used in the test.

The analysis of results showed that text format (paper or mobile device) did not have a significant effect on the scores even when participants had adjusted the mobile device viewing settings to their own preferences. The analysis of time on task showed a significant difference between the mobile device that had the default viewing settings and the mobile device that the participants had been allowed to make adjustments to. The adjusted mobile condition was faster than the default settings one. Although there was no

significant difference observed between the paper condition and either of the mobile device conditions the time on task of paper condition fell midway between the two mobile conditions.

The discussion on the reported difference takes into account the format rotation matrix: making adjustments to the mobile device was, throughout, the final part of the comprehension test and it is possible that the improved time on task might be the result of increased overall familiarity with the test protocol and the device.

11.2.1.3 Comprehension test: Perceived experience ratings

The reading comprehension test included a perceived assessment regarding ease and fluency with each format used, participants also being invited to provide open-ended comments to justify the scores they had given to each text format. The perceived experience data was collected in order to compare the task performance with perceived experience.

Ratings for ease of completing the test and fluency of reading showed that participants found the test significantly easier to complete in paper format than on either of the mobile formats. Fluency of reading was rated significantly higher for both paper and adjusted mobile formats than for the plain default mobile format. When asked to pick a preferred format, paper was (as was perhaps to be expected) the most preferred option with adjusted mobile second and the default unadjusted mobile third.

11.2.1.4 Comprehension test: Open-ended comments to subjective ratings

The open-ended comments led to a number of general observation categories in elements that most affected user experience: a preference for paper format had much to do with its familiarity and the convention of “having always” used paper. More interestingly, a better overview with more efficient preview and backtracking opportunity as well as easier content navigation were all benefits

of the paper format. Mobile devices were criticized, correspondingly, for their poor overview as well as for a lack of proper preview and difficulty in tracking back to something that had been read before. In both paper and mobile formats, then, participants recognized the importance of being able to gain an overview of the content and how it has been ordered, as well as being able to scan the text ahead and to return back to earlier parts to re-read or check an earlier item. An additional difficulty mentioned with the mobile phone was the extra effort required by physical interaction: less content on a page required more frequent page turning. With the adjusted mobile format a number of elements that improved the experience compared to the unadjusted mobile format were identified: more text per page is better, as is more text per line. In addition, simply being able to adjust the appearance and layout details seemed to produce a better user experience for many participants.

Looking at the overall results from the present thesis, the participant comment categories overlapped with the reading factors discovered in the interview study into experiences, expectations and attitudes (Chapter 9) that contributed to a model of mobile reading acceptance. Appearance and layout contained many of the same elements that made up the text format factors in the mobile reading acceptance model, and the limitations participants found with the mobile device echoed the items that made up the device factors in the said model. The most prominent overlap between the comment categories in the comprehension study and the factor model in the interview study was with the importance of overview, navigation, interaction and visibility. In the mobile acceptance model these comprise a factor category called ‘Reading factors’.

11.2.2 Word search

30 participants were asked to perform a word search task on a mobile phone. In the test combinations of background graphics, text colour and foreground effect graphic (simulating reflections on the screen and other light effects) were presented to the participants (one picture on a phone screen at a time), and the participants were then asked to count how many times a specific word occurred in the list of words. After each image, participants were asked to rate

the perceived legibility on a Likert scale with reference to their idea of optimal and perfect legibility and the poorest legibility possible.

Small displays typically also mean mobile context – which in turn means varying conditions in terms of attention and light. The experiment on legibility on a mobile device was devised for the purpose of comparing a computer-based character recognition system (Optical Character Recognition, OCR) and human subjects in conditions where different light conditions were simulated graphically on a mobile phone screen. The experiment's original purpose was to determine whether an OCR system could be used as an evaluation tool in interface design iteration for evaluating perceived legibility of text (Colley et al., 2012). For the present thesis, the data from the experiment were used to compare the word recognition scores by human participants in the experiment against their perceived assessment of legibility in the test and to study if typical legibility restricting distractions (reflections, graphical design of the background of the text) affected word recognition on a mobile phone screen.

The results showed that word recognition score with human participants remained fairly constant throughout the rotation of background and foreground graphics and text colour, whereas the perceived legibility varied significantly. Both light effect (simulated reflection) and background graphic resulted in a lower legibility rating. Moreover, the interaction effect was significant, indicating that the decline in experienced legibility was more pronounced on plain background than on wallpaper graphic background. A matter of additional interest in the analysis was that the recognition scores by the OCR system used in the experiment showed that the OCR scores and perceived legibility scores had correlated positively.

With regard to the relationship between performance and perceived experience, then, it is clear that while human performance remains at a level where at least the present experiment was unable to detect enough variance between tested conditions, participants are nevertheless able to differentiate between optimal and less than optimal viewing conditions. Furthermore, OCR performance (in any case with the system used in the experiment) could

possibly be used in design iteration if not for estimating actual legibility then at least for gauging end-user satisfaction.

11.2.3 Appearance and layout

As described earlier, the reading comprehension study (Chapter 5) included a section in which participants adjusted the appearance and the layout of the text on a mobile phone screen to better match their personal preference. The results of the comprehension test are discussed in section 5.2 but the analysis of participants' adjustments and their open-ended answers explaining why the said settings were selected are presented in Chapter 7.

Existing literature on effects of appearance and layout tends to focus on the direct effect of a specific set-up on some measurable reading performance indicator, such as reading speed or reading comprehension (Bernard et al., 2003; Dyson and Haselgrove, 2009; Dyson and Kipping, 1997; Samuels and Eisenberg, 1981; Fraze and Schwartz, 1979). With the present experiment in Chapter 7 the interest lay in observing the actual changes to appearance and layout if participants were invited to make these themselves with the objective of rendering the screen as readable and comfortable as possible.

In the comprehension test, participants first completed test sets on paper and on a mobile which had default appearance and layout settings that replicated the look and feel of the paper text as closely as possible. For one more set the participants were asked to adjust these settings to their personal preference. The settings included general appearance ones such as typeface, font size, colours (text, background, etc.) and layout (margin width, spacing, alignment, etc.). Participants were simply asked to adjust any items they wished, or told not to change anything at all if they thought the default they had tried earlier was fine.

The single most adjusted item of the possible adjustments was font size: 40% of participants opted for a slightly larger typeface than the one that had been presented in the default settings. The second most frequent adjustment was to

turn the mobile device into landscape orientation from the default portrait orientation. Switching from a serif font to a sans serif one was done by 22% of the participants and line spacing was increased by 10% of the participants. Many participants did adjust more than one setting, but the spread of the combinations was very wide: only the combination of a larger font together with landscape orientation stood out as slightly more frequent than other combinations, and even in this case the adjustment was made by only eight out of the 45 participants.

Better legibility was one of the main reasons for selecting a larger font size. Despite the reasonably good resolution of the test device, text size that was close to the 9pt font used on the paper versions of the test texts was too small for many of the participants. On the whole, clarity (being able to see the text or words more clearly) was cited as a reason for the adjustments. Slightly longer lines also seemed to appeal to many of the participants as a factor that made reading easier on the small screen: many had complained about the lack of overview and how difficult it was to preview text while reading (skim ahead), and many expected landscape orientation would help. Interestingly, the eight who both increased the font size and also turned the device onto landscape orientation seemed to be looking for a compromise between a legible size font and the amount of text they were able to see on a single line or on the screen at once.

11.2.4 Modified critical incident study (to read or not to read?)

All the other studies in the present thesis have been conducted formally either as experiments in laboratory conditions or as set questions in an interview or survey questionnaire, using predefined materials in order to control the independent variables. To balance the pre-set experiments and studies with some understanding of what experiences have been in actual, ecologically valid, situations, the critical incident technique (Flanagan, 1954) was adapted in asking a small number of user experience professionals for examples of situations where reading some content on their mobile phones had not been feasible – for whatever reason.

Though the sample group was very small, adjustments to the method should have been enough to ensure that the output from the study would be of acceptable quality for, say, a design or concept evaluation as an informative check into a system in realistic context of use. The adjustments to the method were:

1) Applying the principle (Nielsen: useit.com) that 5-6 participants in a usability test is an adequate sample size for discovering up to 80% of critical findings,
2) Making that sample up from user experience professionals, thus employing principles of expert evaluations in the process, and
3) Defining the focus of the study directly to a specific issue instead of categorising issues from observations of general use of a given system (as done in the full critical incident process)

Focusing the question was done on the basis of other studies in the present thesis. The participants were asked to describe if they had had an experience with their own mobile phones where they were going to, were asked to, or were required to read some text content on their mobiles but for some reason had decided against doing so.

The answers from the respondents in the study fell into 3 rough categories. First, there was no need to carry on reading a given piece of text content because it quickly became clear to the respondents what the content was about. The respondents either recognized the sender (e-mail or other private message) and thus knew what the communication was about, or the content was otherwise recognized by type (terms and conditions statement on a website/service). In one case a private message was read to a point where the respondent saw that responding to the message properly would require additional actions (finding information). A second group of responses drew attention to problems with layout and appearance that were so disruptive to the reading process that despite high personal interest it was not possible to carry on reading the content. In the third category the respondents considered

the content to be too long or complicated to be viewed and read properly on a mobile phone screen.

In all cases the interrupted task (reading a specific text) was postponed until a more functional interface was available. Typically this was a PC/laptop, but in one case simply a different mobile phone with a slightly larger screen was improvement enough for the participant to carry on with the task. Mobile context of use (being out and about) appears one thing that will always influence mobile reading: the situation where content is accessed has to be one that caters for focusing attention on the content for long enough. Respondents mentioned commuting as such a time and place for reading a favourite blog or a news website; others did not read a message because their present situation did not allow them to pay enough attention or to comfortably respond to a message.

11.2.5 Interview on attitudes and expectations

12 participants were interviewed regarding their reading habits and preferences. Initially, the topic was approached from a general reading perspective to help the participants ground their answers to their everyday behaviour and experiences, but eventually the focus moved more to discussing reading on mobile phones. The responses were analysed using grounded theory to establish categories of answers and from the categories a model of elements (factors) that affected reading experience emerged. The model showed how factors regarding the physical aspects of reading media (size and type, for example display size of a mobile) and text formatting affected how easy or fluent it was to gain an overview of a text, or to navigate within content, what the interaction with the paper or device was like, or how easy it was to see the content clearly enough. It was also established that a number of factors were reader-internal in the sense that the quality of these factors was something individual for each reader: general reading proficiency, familiarity with given subject matters, interest in a given topic, etc.

The factors discovered in the study follow the reading factor model by Samuels and Eisenberg (1981), and the initial apprehension and concerns expressed by the interview participants echoed those of the students' concerns prior to a PDA reading experiment in Waycott and Kukulska-Hulme (2003): small screen, small text, quality of the screen, navigating in the content and interacting with the device (namely, annotating text) were raised both in the interview study in the present thesis and in the PDA reading evaluation.

The factors in the model indicated whether reading experience for some text would be comfortable and fluent or not, and based on that decision, mobile reading acceptability would follow. In other words, if the balance of the factors was favourable (the factors would be perceived to give a positive outcome for reading a given piece of text), reading on the mobile would be acceptable. The model does not directly pinpoint the relative influence of the factors on the outcome, but suggests that the sum total will need to be positive in order for mobile reading to be attractive or acceptable. The analysis does, however, establish that it is possible that a single factor in the model could outweigh the others (such as interest in the topic or need to access the information there and then), and likewise a single factor could outweigh the others negatively, resulting in the end-user abandoning a text on a mobile even when all or many of the other factors were seen to be at an acceptable (positive) level.

11.2.6 Survey

A web survey with 150 respondents was conducted on the basis of the mobile reading acceptance model that emerged from the interview study. The aim of the survey study was to quantify the interview responses and to establish any latent factors that affect the perceived mobile reading experience. Another goal was to see if mobile reading experience and attitudes could be predicted from the factors and thus determine whether the survey questionnaire or parts of it could be used as an evaluation tool in interface and content design.

Factor analysis of the answers, all ratings on a Likert scale from one to seven, pointed to a four-factor solution: *Visibility, Interaction and Navigation* (factor 1),

Perceived reading proficiency and format familiarity (factor 2), *Willingness to read on a mobile* (factor 3), and finally *Perceived legibility* (factor 4). These factors, when compared with the mobile reading acceptance model constructed from the interview responses in Chapter 9 confirm the factor model, but also seem to provide further insight into the relationships between the factors. The *Visibility, interaction and navigation* factor incorporates text format and reading factors from the mobile reading acceptance model, whereas *Perceived reading proficiency and format familiarity* and *Willingness to read on a mobile* point to the importance of reader-internal factors: perceived interest or need in accessing content on the mobile. *Perceived legibility* was also a factor that touched on text format and device factors, emphasizing the need for clarity in the presentation of the content and an appreciation of high quality screens with sharp resolution. Further regression analysis of the factors in the study suggested that *Visibility, interaction, navigation* can be seen as a predictor of *Perceived legibility,* accounting for 15.3% of the variance in *Perceived legibility*.

11.3 Perceived experience vs. performance: what we say and what we do

The experimental studies in the present thesis, namely the reading comprehension test (Chapter 5) and the legibility experiment (Chapter 6) indicated that measured test performance and perceived experience do not necessarily correlate: performance in the experiments did not differ significantly between paper and mobile device conditions or between various levels of distracted visibility on a mobile screen. In both experiments, however, participants were able to notice a difference in the ease and fluency of competing the tasks, and were also able to verbalise some of what they thought were the reasons for the difference in the experience. The lack of correlation has also been acknowledged early on in usability engineering (Bailey, 1993; MacLean, et al., 1985; Nielsen and Levy, 1994; Kissel, 1995), where studies show that although in a large set of materials (such as the meta-analysis by Nielsen and Levy (1994) the objective and subjective measures have a positive association, it is still clear that end-users may well

prefer a system or interface that is not deemed most efficient in the light of task performance. The literature also points to end-users' experience with computing as a factor in how well subjective ratings correspond with objective performance measures: experienced users tend to be more critical of the system being tested.

The interview and survey studies also pointed to the importance of factors that are closely related to any individual end-user/reader: motivation (both internal and external) appears to be (as perhaps common knowledge might suggest as well) a factor that can override many of the other factors that affect reading experience. Although the survey responses showed that respondents greatly preferred paper as a reading format, at the same time they indicated many genres of publications (news, social networking, etc.) as something they would be prepared to read on a mobile phone.

11.4 What affects perceived reading experience

The cornerstones of acceptable reading experience on a mobile device appear to be clarity of presentation (legibility, being able to estimate how the content is structured, identifiable structure markers, etc.), functional interface design (moving from page to page, scrolling / panning, adjusting size of text, minimised need to interact with the content), and motivation (personal interest in and need for the content). While listing these main points may well seem straightforward enough, the research in the present thesis shows that all of the elements in these main factors are made up of various details that interact with each other and may operate differently depending on context: a functional design may not guarantee reading if the moment (context of use) is wrong, or the appearance of the text may not be important if the topic of the text is interesting enough.

11.5 Interface and content design: contributions to interface and content design and user research

The present thesis aimed at looking at reading on small displays from as practical a perspective as possible and at complementing existing research and knowledge by focusing on the perceived comfort and sense of fluency of reading: the all-important UX (user experience). The emergence of larger display smartphones with up to 4 – 4.5” screens has not meant that small displays (2.8 – 3.5”) have become obsolete. Quite the reverse: the big manufacturers’ drive to provide affordable feature phones and smartphones in the emerging and growing markets means producing devices that have cheaper components (including lower resolution and smaller screens) which nevertheless provide the full smartphone experience with e-mail, chat, social networking and mobile web.

In this frame of reference, the present thesis offers mobile interface and content design communities confirmation of some ‘truisms’ regarding end-user motivation and functional design. The modular factor model presented in the present thesis also offers some new approaches to handling the creation and development of user experience less as an abstract bundle of features where everything affects everything else and more as a collection of experience-forming factors that interact and whose influence over UX in specific cases can be tested in a focused manner.

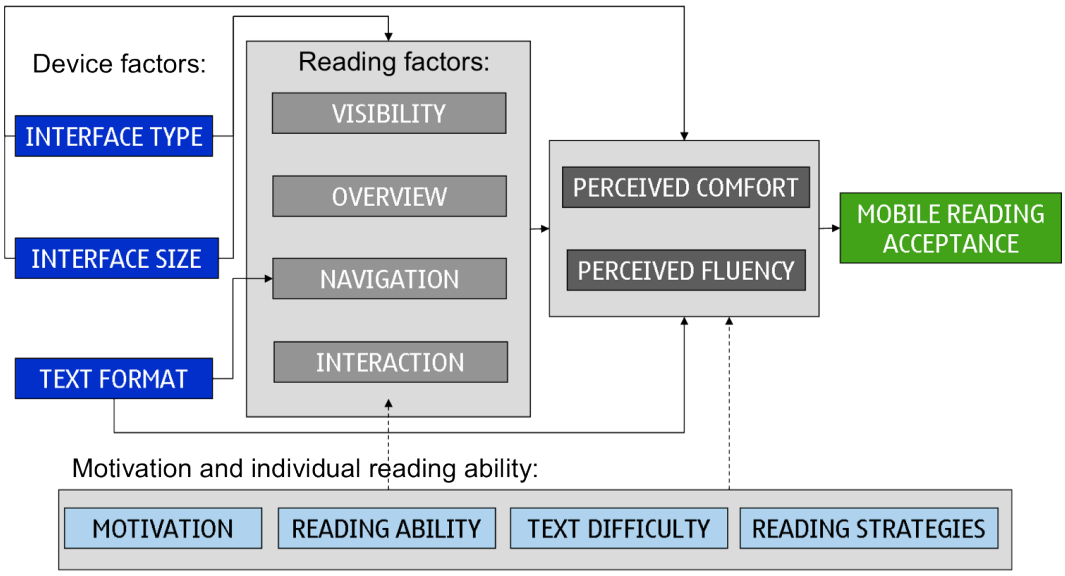


Figure 31 Mobile reading acceptance model.

The model (Figure 31) highlights the importance of visibility, overview, navigation and interaction at the core of experiences that form end-users' perception of a comfortable and fluent reading experience on a small screen device. A positive experience in these four factors (which are affected by the device factors on the left in Figure 31) is likely to lead to a positive outlook on what reading on a mobile device would be like. Samuels and Eisenberg (1981) identified much the same factors, but the present model offers additional clarification as regards the relationships between the factors and thus should help designers focus their effort on ensuring the four core factors are prioritized in their design: it is all well and good to improve display size and technological quality, but to do so must also serve to improve visibility, overview, navigation and interaction.

Highlighting visibility, overview, navigation and interaction as the core issues in forming a reading experience is supported by an earlier study by Waycott and Kukulska-Hulme (2003), in which students evaluated the reading experience of having their course texts on a PDA device. In that experiment size of the screen, clarity of text and issues with navigation were identified as the main problems in reading on a small PDA screen. The present thesis offers further insight into and a more detailed understanding of these concerns: from the

mobile reading acceptance model we see that “size of device” is a contributing factor to experience factors of overview and visibility, and the importance of overview and navigation is emphasized in the reading acceptance model.

The test methods presented in the present thesis may not be functional research and development in 1:1 as such, but the conclusions and models do provide grounds for adjusting these and other established methods in order to focus user research on relevant issues at each design iteration. For example, knowing how the adjusted critical incident technique used in the present thesis produced response categories that supported received reading experience conclusions from other (laboratory and survey) studies, it would then be feasible to use the same adapted technique as a cost-effective expert evaluation with added ecological validity for not only reading but other interaction issues where context of use is seen as a factor.

The results from the studies in the present thesis also have direct relevance for mobile interface and content design. The experiments in reading comprehension and legibility on a mobile screen (word search) indicated that while not all performance is always measurable in the degree of detail that shows a significant difference, issues with user experience are still detectable by the participants. Such conclusions were referred to by other research as well, as listed in Chapter 5 (Comprehension test) and Chapter 6 (Word search study), and highlight something many user experience researchers have discovered informally through usability testing anyway: functionality and usability do not always directly equal good user experience. It is also useful to know that there may not be a big difference in reading performance between conventional formats like paper and mobile devices. The results from the comprehension test as well as the legibility experiment suggest that improvements can be made in user experience, but meanwhile a reasonable content design and presentation do not immediately lead to critical user errors.

The experiment with legibility also confirmed the principles behind CAPTCHA images: the human eye is superior to optical character recognition systems (OCR). The new knowledge obtained from the experiment, however, provides

a functional suggestion for evaluating user experience of text on screen without having to involve human participants in slower and costly user testing at least early on in a system's development cycle. The experiment showed that an OCR system's recognition rate correlated positively with human participants' perceived legibility ratings in the test, suggesting that OCR results could be used as a quick evaluation method for checking whether legibility levels in a system were at a satisfactory level in terms of user experience. In much the same way as using a readability formula like Flesch-Kincaid to establish that a text's readability remains at a level that is clear and easy to understand for most people, the OCR reading can establish early on in the design cycle that the typographic and other graphical design does not affect user experience (too) negatively.

When test participants were asked to adjust viewing conditions on a mobile phone screen to their liking, font size was an important choice, but the adjustments also aimed at displaying such an amount of text on the screen as to gain a better overview and enable previewing the text. In the critical incident technique test, however, an important aspect of text size adjustment was mentioned: default text size levels should be as functional as possible to begin with, and any shortcut interactions (double tap in the respondent comment) should offer a meaningful jump in the adjustment. Furthermore, text should always wrap to the width of the screen, as seen in the critical incident study response as well as in the reading experiences, expectations and attitudes interviews. Text size should not increase (or decrease) too excessively when a mobile device is turned from portrait to landscape orientation or vice versa: end-users may find this distracting and irritating (as seen in the interview results). Also, on the whole, anything the interface and content design can do to allow readers to gain an overview of how a piece of text has been constructed, how to move within the said content (clear navigation) and how to minimise interaction with the device is likely to result in better user experience.

The general heuristics listed above provide interface and content designers with some basic, practical cornerstones for design: to have these elements in place means that from earlier on in user testing the focus can be on the detail

and unique features of a system or body of content. To put it simply, the niceties of particularly fluent zooming interaction are rendered void unless word wrapping can be guaranteed, but if the system has wrapping sorted and functional, testing resources can be directed to the niceties of the zooming interaction straight away with the chance of obtaining meaningful results for just that: the zooming interaction.

11.6 Contribution to research

The present research aimed at gaining a better understanding of the realities of continuous reading on small display devices. The focus of the research was on understanding the perceived experience of reading: when is reading considered fluent and comfortable (easy), particularly on small display devices? The existing work on mobile devices and on reading on electronic screens has placed much emphasis on improving performance such as reading speed. The present research adds to the user experience dimension in the reading research and mobile user interface research field, not forgetting the needs of the content creation and interface design industries.

From previous research we have learned that reading performance may not always be measured but that end-users (test participants) are able to experience differences in the process all the same (Bernard et al., 2003, Grzeschick et al., 2011, Laarni, 2002; Dillon et al., 1990). The experiments in the present thesis confirm this conclusion, but also emphasize the importance of keeping an eye on perceived experience when striving to improve performance: experience of ease or fluency, or mere familiarity with an interface can outweigh the sense of improvement in performance that might come from innovative interface solutions. Furthermore, as seen from the experiments in the present thesis, it is vital to complement performance rates with perceived experience data in order to be able to understand how the various factors affect each other when performance remains unchanged but participants are still able to rate experience differently from one condition to another.

Research into learning materials (De Bruijn, et al., 1992, Oostendorp, 2003) and research into how content has been split on different size screens (Dillon, et al.,1990) identified that fragmentation of content can produce problems in both reading performance and in the reading experience. Again, the present research can confirm and emphasize the severity of the problem as an experienced one: while the reading comprehension test showed no significant difference in the comprehension score (Chapter 5), participants in the test complained of the difficulty of forming an overview of the texts in the tests and of using typical reading skills such as preview (skimming ahead) or backtracking to check on a segment of text earlier in the set. Based on the expectation of fragmentation, split attention and working memory load suggested by the literature, it was hypothesized that there would be a difference in the comprehension test scores between reading on paper and reading on a mobile device, the expectation being that paper would produce more correct answers. This was not the case, however, and it was down to the other variables (time on task and subjective ratings with open ended comments) to establish the reasons for the negative expectations mobile device users had voiced in the interviews (Chapter 9). A time limitation on a task like reading comprehension test might offer more variance in test scores from format to format, but such a time limit may not have great ecological validity.

Samuels and Eisenberg (1981) presented how the reading process can be categorised into smaller factors that design and development can address directly. In the present research some of the principles from Samuels and Eisenberg were used in studying mobile device users' experiences and attitudes, and then developed further to propose a model of how the experience of reading can be split into smaller factors and how these factors interact and eventually contribute to an acceptance model on reading on mobile devices. The main factors in the reading process that were identified in the present research (Chapters 9 and 10) were typically the same as identified by Samuels and Eisenberg (1981). However, present research augmented this existing model by proposing a clearer division of categories for the said factors and by illustrating a more causal relationship between the device-dependent

factors and a set of core experience factors (called Reading factors in the model). These experience/reading factors are ones that rely more on interface and content design.

Eye movement studies such as Balota and Rayner (1991, In Besner and Humphreys, 1991) and Hyönä (1993) have found that reading involves more than merely focusing the eye on the characters and words that are being processed (read) at the time. The reading process requires previewing and regression (backtracking). These little preview and regression movements are very quick and do not necessarily reach very far from the word or character that is being read. On a somewhat larger scale, Dillon, et al. (1990) showed that splitting sentences across a page break would seem to interfere with the comprehension process. In the present research (Chapters 5 and 7) it was again the perceived reading fluency and ease that showed the effect of having content fragmented across multiple pages: participants did not directly know how to verbalise any change at the small preview/regression level of eye fixations when reading, but they were able to recognize the increased need to move from page to page in order to properly integrate all parts of a sentence or an 'idea unit'.

The experiment described in Chapter 7 aimed at identifying patterns and preferences in how text content was ideally presented on a mobile device screen. In earlier literature Yi et al. (2011) pointed to a combination of 1.5pt line spacing in a single column as an optimal setting, although they noted that long lines were seen by end-users as perhaps a little daunting. The present research found that the most important adjustment and expectation for end-users is clear text legibility (participants most typically increased text size on the screen). Line length did not seem to be an issue with the <4.5" screen size, as turning the device to landscape orientation was the second most typical adjustment. The interesting connection between text size and device orientation was, then, that often turning the device to landscape was paired with increasing the text size. Such a pattern of adjustments suggests that it might be possible to aim at defining ideal line-lengths in terms of words per line at legible text size. Likewise the practical implication, as suggested in the

discussion in section 7.3, is for designers and device manufacturers to place an emphasis on the default text size of mobile content.

The importance of collecting and analysing perceived experience feedback was again justified and confirmed in the experiment presented in Chapter 6, when again performance scores in the task of finding set words from a group of similar-looking words on a mobile phone screen did not vary significantly. Again, participants' experience feedback indicated a clear preference for certain contrast and colour combinations over others. The same study further proposes that computerised OCR (optical recognition systems) could under certain conditions provide a low-cost means of early development UX testing on text and graphics: an OCR system's word search scores and test participants' experience scores showed a positive correlation.

11.7 About methodology and future research prospects

The research in the present thesis makes use of both qualitative and quantitative methods. Mixed methods have been used for studying both measurable task performance as well as participant experiences that have been freely expressed. The aim of the mixed methods approach was to gain a multifaceted view of the reading experience and particularly the relationship between performance and experience. Another aim of the research was to produce new knowledge that could be applied to further interface and content design and research, and so statistical methods for predicting behaviour were employed as well.

The reading comprehension test presented in Chapter 5 was a mixed methods study in which participants performed set reading comprehension tasks, answered a small number of qualitatively analysed questions and also made interface adjustments that were later categorised and analysed. The word search study in Chapter 6 again had participants perform controlled tasks with a mobile interface and then rate their experience. Analysis in Chapter 7 categorised and listed what adjustments were made to the mobile interface in

the reading comprehension test and how participants commented on or justified these adjustments.

The study in Chapter 8 followed a semi-structured interview protocol and the material was analysed using grounded theory method. The resulting mobile reading acceptance model from the interview study was used as the basis for creating a web survey (Chapter 10). Factor analysis and multiple regression were used in analysing the data from the web survey. In the final study that asked user experience professionals about their mobile reading experience, critical incident technique was modified slightly in order to adapt it more closely to the study context and situation.

There are, nonetheless, areas where further investigation would yield additional, useful knowledge about the relationship between what end-users do as opposed to what they say, as well as about establishing some tolerance thresholds in areas such as layout and appearance where performance and experience potentially correlate positively. In the first case, the interview results suggested that reading on the mobile phone was not perhaps very appealing to the sample; and yet the survey indicated that if the topic was interesting (reader-internal factors were right for an individual participant), it would be possible to read continuous text on a mobile. The user experience professionals, who are typically advanced mobile users, revealed in their responses that mobile phones are used for reading a variety of content. It would, therefore, be interesting and useful to explore the borderline between expressing a preference for conventional text formats and replacing the conventional formats with digital media, particularly mobile devices: do we say we would prefer paper and yet access our favourite content on a mobile? Or perhaps even vice versa: do we maintain that we use our smartphones in a very versatile manner, including reading web content and e-books, but in reality revert to print with any significant reading matter?

In the case of layout and appearance tolerance thresholds, the question would be to find out if performance can be measured so accurately that the point where test participants register (consciously or sub-consciously) a difference

on the experience scale could be seen also in the performance scores. With such measures it would be feasible to predict user experience from formally controlled performance tests and UX researchers would be able to use more measure-based data and so rely less on interpretative observation. Such an approach would benefit organisations that tend to replicate a test series globally, using contracted UX researchers for data collection. The less need there is for distributed (qualitative) interpretation the more comparable the results from various locations would be.

The web survey study presented in Chapter 10 used a questionnaire and then aimed at finding out if the questionnaire or parts of it could work as an evaluation tool in predicting user behaviour. The results did not indicate that the entire questionnaire would work as such, but it would appear that parts of it could be re-designed in order to predict likely responses to design: namely, the perceived legibility could to a degree be predicted from end-user responses to questions on overall visibility, interaction and content navigation.

11.8 Final conclusions

The present research has identified and confirmed a number of elements regarding reading on small screens that affect the experience of reading continuous text. A mobile reading acceptance model was devised and confirmed, proposing a set of factor categories with a degree of causality. The model emphasizes key experience factors (namely, visibility, overview navigation and interaction) that interact and can even outweigh each other. The model also includes motivation as a key factor in forming an end-user's attitude and willingness to access text on a small screen device.

The research also confirmed earlier research findings that performance and preference (or experience) do not always correlate positively, and the present research complemented these findings by identifying that it is beneficial to measure more than one performance variable and also to collect subjective feedback in order to augment the understanding of the performance data. These findings highlight the importance of perceived experience and its role in

content and technology use: pure usability (as often tested through functionality and ease-of-use) does not guarantee a positive user experience, and the mobile reading acceptance model shows how it is precisely experience that largely dictates the likelihood of a given text content being accessed and read on a mobile device.

The results of a word search experiment indicated a possibility of using a computer-based optical recognition system as an early design cycle low-cost method of identifying possible problems in visibility and legibility in terms of user acceptance. Regarding optimal presentation of text on small screens, (perceived) legibility appeared to be the most important parameter, followed by a (perceived) functional line length in terms of words per line. One important factor that emerged from observing the individual adjustments made by test participants was that in some cases merely the knowledge of having the freedom to adjust the interface seemed sufficient to improve some participants' reading UX. The contributions to interface design in the present thesis include the use of a modified critical incident technique as a way of conducting expert evaluations on a system. The thesis also offers a short list of heuristics to be considered and used when designing how text is displayed on small screen devices.

The present studies add to the body of knowledge that previous research has approached in the context of reading in general or in the context of reading on electronic displays that are larger than what is considered a mobile phone (feature phones or smartphones). The experience of fluency and ease does not necessarily follow clearly definable attributes or performance as such in a way that would allow us to conclude that simply offering a larger font size or improved display resolution will guarantee a positive perceived experience in reading. The present research should offer improved insights and knowledge about the components that build up the experience and about how those components interact. With that knowledge, further research can focus more precisely on what contributes to problems with reading on small displays, and what might be done to improve the situation.

Appendices

12.1 Comprehension test: text sets

Practice: Windscreen wipers

The first patent for windscreen (1)_____ was issued to a woman from Birmingham, Alabama in 1903. Reportedly, Mary Anderson noticed (2)_____ a visit to New York that when it rained the drivers of trolley cars either had to leave the windows (3)_____ and expose passengers to the weather or get out (4)_____ few minutes to wipe their windscreens by hand. Anderson came up with a device that could be operated by hand from inside the car. Sadly, she made little (6)_____ from her invention since no manufacturer could see the value of it

Source: BBC History Magazine, 2007

SET 1
How I poisoned my family

THE novelist Nicholas Evans once declared: "Guilt is my subject." That was (1)_____ the warm August evening in 2008 when he almost killed half his family. Since then, he says: "I've taken (2)_____ research to a rather extreme degree." Evans and his wife, Charlotte, were staying with her brother and sister-in-law (3)_____ Scotland, when Evans picked some wild mushrooms and fried them up for supper. Within 24 hours they were all in (4)_____ fighting for their lives, as one by one they went into kidney failure. They survived, but remain on dialysis. Evans's 29-year-old daughter, Lauren, recently persuaded (5)_____ to accept one of her kidneys, but (6)_____ else is still waiting for a donor.

Although Evans's wife has forgiven him, family relations are not good, he (7)_____ Decca Aitkenhead in The Guardian. It doesn't help that they can't agree on whose fault it was. "I did (8)_____ the mushrooms, but it was really two people, each thinking the other one knew what he or she was (9)_____... It was a complicated transaction, really, and it involved the two of us suspending our responsibility, assuming that the other (10)_____what they were doing." He pauses uncomfortably. "It's really too hard to talk about. I can't go into this, it's (11)_____ too much. It has caused us too much pain." Can he see things improving? "I don't know," he (12)_____. "I think maybe when we're all better. I think that will help enormously, when everybody's got a transplant."

Source: The Guardian/summarized in The Week, November 2011

For the Best Pick-Me-Up, Lie Down

SCIENTISTS (13)_____ a morning training 61 people in motor, perceptual and verbal tasks: tapping a keyboard in a specific sequence, discriminating (14)_____ shapes on a computer screen and memorizing a list of words. Then the scientists randomly divided (15)_____ subjects into three groups. The first took a nap from 1 to 3 p.m. At 3, the second (16)_____ took a 200-milligram caffeine pill, and the third took a placebo (like medicine but having no effect). The subjects (17)_____ the tasks they had been taught earlier and were scored by researchers who did not know which group they (18)_____ in. Those who had caffeine had worse motor skills than those who napped or had a placebo. In the perceptual (19)_____, the nappers did significantly better than either the caffeine (20)_____ placebo group. On the verbal test, nappers were best by a wide margin, and the caffeine consumers did no (21)_____ than those given a placebo. Despite their mediocre performance, caffeine takers consistently reported less sleepiness than the others. "People think they're smarter on (22)_____," said Sara C. Mednick, an assistant professor of psychiatry at the University of California, San Diego, and the lead author of the study. "But this study is a strong argument for taking a nap instead of having a cup of coffee."

Source: The New York Times, December 2, 2008

Mobiles Are Probably Safe

THE global economy may be imploding – but at least our (23)_____ phones aren't slowly killing us. Campaigners have long argued that exposure to the electro-magnetic radiation emitted (24)_____ phones could generate cancerous tumours. However, researchers at the Institute of Cancer Epidemiology in Copenhagen (25)_____data from nearly 360,000 mobile phone users (some of whom had owned phones for 13 years or more) (26)_____ non-users. Published on the British Medical Journal website, this is the second major study this year to (27)_____ no significant difference in the cancer rates between mobile-phone users and non-users. However, the Danish (28)_____ said that further studies were needed to (29)_____ out the risk of longer-term damage, and determine the (30)_____ of mobile usage on children.

Source: The British Medical Journal/summarized in The Week 5 November 2011.

Practice: Windscreen wipers

The first patent for windscreen (1)_____ was issued to a woman from Birmingham, Alabama in 1903. Reportedly, Mary Anderson noticed (2)_____ a visit to New York that when it rained the drivers of trolley cars either had to leave the windows (3)_____ and expose passengers to the weather or get out (4)_____ few minutes to wipe their windscreens by hand. Anderson came up with a device that could be operated by hand from inside the car. Sadly, she made little (6)_____ from her invention since no manufacturer could see the value of it.

Source: BBC History Magazine, 2007

SET 2
Cowell versus the punks

SIMON Cowell has always been a creature of the mainstream. In 1977, when he (1)_____ 18, the future pop mogul got his first job at a record label in London. The punk (2)_____ was exploding and it was a thrilling time for British music – but Cowell was not impressed. "I went to (3)_____ Stranglers gig, and thought: 'This is just the worst night I've ever had in my life,'" he told Chris Heath (4)_____ American GQ. "Everyone was gobbing on each other, literally. I was not into that. I just didn't get it. (5)_____ wasn't that angry. I mean, I genuinely wasn't. I wasn't anti-establishment or anti-royals. I wasn't really anti (6)_____." And deep down, he believes, the punks weren't either. These days Johnny Rotten, lead singer of the Sex Pistols, (7)_____ most often seen on TV flogging Country Life butter. "They were taking themselves so seriously." laughs Cowell – "and the great (8)_____ you want to tell people 40 years later is 'Put butter on your crumpets'? In the end, they all (9)_____ the cheque. That's the truth."

Source: American GQ / summarized in The Week, November 2011

Umbrellas

UMBRELLAS and bowler hats were (10)_____ twin hallmarks of the 1950s City gentleman. By contrast, 200 years earlier umbrellas had come from the Far East and (11)_____ feminine luxuries – men had to do with overcoats called "surtouts".

In the 1750s, on returning from an exotic odyssey (12)_____ took him to Persia, the philanthropist and social reformer Jonas Hanway (1712-86) started carrying an umbrella on the (13)_____ of London. Weak health motivated this unorthodox move – he was "extremely sensitive to cold" and wore flannel underwear and three (14)_____ of stockings. A contemporary biographer also hinted at vanity – the umbrella stopped dirty rain and waste water from chamber (15)_____ spoiling his clothes and large wig.

Hanway's pioneering use of a female accessory meant he was ridiculed for (16)_____ feminine. Coachmen, who were fearful that the spread of umbrellas would rob them of trade, jeered and jostled him. Hanway (17)_____ about being splashed with mud at every "accidental" jolt of the wheel in the street gutter. But the protests (18)_____ futile and within a few years men could buy umbrellas from dedicated suppliers In the 1780s, Samuel Lund sold (19)_____ to sufferers of gout and rheumatism Hanway's umbrella would have been difficult to carry, because it was made from oiled silk or cotton (20)_____ over a heavy ribbed wooden framework.

The first registered patent for a spring-action umbrella with a jointed handle was (21)_____ in 1786, the year Hanway breathed his last. In 1787, a poet wrote that "umbrellas ... are all the rage, For youth, for manhood, or for age." A Thomas Folgham marketed "pocket and portable" umbrellas in Cheapside. (22)_____ the 19th century the main colour was black so that industrial dirt did not stain them. They were known colloquially as "Hanways".

Source: BBC History Magazine, 2007

The weight of an e-book

E-readers enable (23)_____ to carry around their libraries – but adding extra books makes the devices heavier. They are, however, unlikely to get (24)_____ heavy: filling a 4GB Kindle to its full capacity, a scientist has calculated, would increase its weight by about (25)_____ billionth of a gram. Storing new data involves holding electrons in a fixed place in the device's memory, and (26)_____ them still rather than letting them float around takes up extra energy, Professor John Kubiatowicz of the University (27)_____ California, Berkeley, told The New York Times. Using Einstein's $E=mc^2$ (28)_____, which implies that energy and mass are related, he (29)_____ that each additional book weighs about the same as a molecule (39)_____ DNA.

Source: New York Times / summarized in The Week, November 2011.

Practice: Windscreen wipers

The first patent for windscreen (1)_____ was issued to a woman from Birmingham, Alabama in 1903. Reportedly, Mary Anderson noticed (2)_____ a visit to New York that when it rained the drivers of trolley cars either had to leave the windows (3)_____ and expose passengers to the weather or get out (4)_____ few minutes to wipe their windscreens by hand. Anderson came up with a device that could be operated by hand from inside the car. Sadly, she made little (6)_____ from her invention since no manufacturer could see the value of it.

Source: BBC History Magazine, 2007

SET 3
A Forgiving Hostage

PAUL Chandler is surprisingly understanding towards (1)_____ former captors. In 2009, he and his wife Rachel were captured by Somali pirates while on a sailing trip around (2)_____ world. They were held hostage for more than a year, with the pirates demanding millions of dollars in (3)_____. Terrifying though it was, Chandler never took it personally.

" I think we're all driven by greed," he told Decca Aitkenhead (4)_____ The Guardian. "But because we live in a society with rules, we live by them. We suppress our instinct (5)_____ fight for every last morsel. But for them, after decades of total lawlessness, there are no consequences for wrongdoing." The (6)_____ were not so much morally flawed, he says, as financially misinformed. "They were incredibly naïve and ignorant of the (7)_____ world. They couldn't understand why we couldn't raise a lot of money. They said: 'Britain's population is 60 (8)_____ – one man, one dollar.' There was no malice. Because if it had been the other way round, they thought, well everyone in (9)_____ clan would give money. That was their genuine belief."

Source: The Guardian / summarized in The Week, November 2011.

Learning from the woodpeckers

IF you pounded your head against (10)_____ tree trunk hundreds of times a minute, you'd suffer brain damage. Yet woodpeckers peck wood at speeds (11)_____ six metres a second, enduring a deceleration more than1, 000 times the force of gravity at each peck – and (12)_____ unharmed. Now researchers from Beijing and Hong Kong have examined how the bird's brains are protected, with a (13)_____ to designing more effective helmets for humans. Their report, published on Plos One, concludes that the woodpecker anatomy (14)_____ three key mechanisms to prevent concussion. First, the bird's lower beak is longer than the upper, so it takes (15)_____ of the impact when the beak hits the tree. Second, the brain is contained in a unique skull casing, (16)_____ from spongy plates that make it unusually shock-absorbent. Finally, (17)_____ have an unusual hyoid bone, which reaches from (18)_____ beak and loops around the skull, acting like a seat belt. In humans, the brain fits more loosely inside (19)_____ casing – after an impact, it's the movement of the brain against the skull that causes concussion. Lead author Yubo Fan (20)_____ the findings could be "applied to human protective devices such as sports-helmet designs".

Source: Plos One / summarized in The Week, November 2011.

"Plastic plane" takes to the sky

BOEING'S (21)_____ -awaited and much-hyped new 787 Dreamliner made its maiden commercial (22)_____ from Tokyo to Hong Kong (23)_____ week – three years later than planned – and was immediately hailed by industry experts as the most significant development in (24)_____ engineering since Concorde's maiden flight in 1976. Boeing says its aircraft, which is made largely from light carbon fibre, (25)_____ burn 20% less fuel than comparable aircraft, and cause less noise pollution. They also claim that it offers a far (26)_____ relaxing flight, thanks to the cabin's spaciousness (8ft high and 17ft wide), LED-coloured mood lighting and adjustable humidity.

(27)_____ the passengers on the All Nippon Airways 264-seat jet for the inaugural flight were around 100 Japanese aeroplane enthusiasts (28)_____ had won the chance to buy a ticket in a lottery. How many of them used the bidet loos that gently (29)_____the passengers' bottoms is unknown: the billions spent developing the (30)_____ have delivered an almost silent flush that is inaudible beyond the cubicle door.

Source: The Week, November 2011.

12.2 Comprehension test: participant answers and accepted answers

SET 1 answers

#	Original	From participants	Accepted	Note
1	before	- after before during evident in on said the	before evident	
2	my	experimental guilt onboard my new poisoning scientific some that the this up	experimental guilt onboard my poisoning scientific up	
3	in	in Melissa	in	
4	hospital	- agony bed hospital pain trouble	agony bed hospital pain trouble	
5	him	Charlotte father find him me Nicholas one someone	Charlotte him Nicholas	
6	everyone	- Charlotte Evans every everyone everybody government Nicholas someone unfortunately	everyone everybody	
7	told	- told and(?) tells contacted informed	- told tells informed	

		interviewed tells		
8	pick	- by(?my?) cook eat find fry get pick serve take test wash	cook find fry pick serve	
9	doing	doing eating thinking	doing thinking	
10	knew	asked finished knew know person people	knew	
11	just	- all became far hurt hurting/upsetting hurting just really	- all far just really	
12	admits	commented cries replies said says sobbed	cries replies says	
13	spent	- after attended began completed conduct had held in on organised report running say spend spent start started tested took	began held organised spent took	
14	between	- against	between	Had there been ‘distinguishing’

		ambiguous between certain colour different from some the various		instead of 'discriminating', then 'various, 'different', 'certain', etc. would have been acceptable. ('Discriminating' means 'seeing fine differences').
15	the	- 61 the those	- 61 the those	
16	group	- then group participants subject	- group	
17	repeated	attempted completed did in learned performed recalled repeated received recital retook thought undertook underwent who	attempted completed performed repeated undertook belonged were	
18	were	been belonged were	belonged were	
19	task	category condition group task tasks test tests	category condition task tasks test tests	
20	or	and group or pill	or	
21	better	better different better or worse worse	better	The context is all about defining better/worse and as such only 'better' is truly acceptable choice here.
22	caffeine	caffeine caffeine caffine coffee drugs	caffeine caffiene (sic) caffine (sic) coffee	

		it		
23	mobile	cell mobile	mobile cell	Cell here is accepted as referring to 'cell phone' (mostly US colloquial English for 'cellular phone').
24	by	by from means mobile through	by from means	
25	analysed	- collated collected compared found gathered reported reviewed showed studied tools	collated collected compared gathered reviewed studied	
26	and	- against and than to were with	- against and to with	'Against' could possibly work IF the previous one is 'compared'. Same with 'to'. 'With' would be acceptable IF it had been with 'compared' or 'reviewed'.
27	find	demonstrate find have report reveal see show state	demonstrate find have report reveal show	
28	researchers	- ambassador authorities government institute of cancer man researcher researchers scientist scientists spokesman team	- researchers scientists spokesman team	
29	rule	assess balance carry cancel check cut	rule	Using 'find' or 'figure' here would suggest there is a problem, but the context and the argument of the

		find figure rule iron see show work		text here is that there IS no problem.
30	effect	damage danger dangers effect effects impact implications outcome risks risk safety usage	effect effects impact	

SET2

#	Original	From participants	Accepted	Notes
1	was	began turned was	turned was	
2	scene	craze decade era genre movement music phase rock rocker scene star trend	craze decade era movement phase scene trend	
3	a	a one see the	a one the	
4	in	- an at editor for from North of the	at for from of	
5	I	I	I	
6	anything	- American anything government it punk social	anything government social	
7	is	- and are is now was who's who is	is was	As this is a case of reported speech, 'was' is acceptable.
8	message	- accomplishment advice is that message news part people stories story that thing things	message news story thing truth	

		thing is truth		
9	want	about accept banked cash cashed earnt got had know need shared sign signed spent take took want wanted	accept baked cashed take took want wanted	Accepting past tense here is supporting a misreading of “the end”: this is how the world is (in general), not simply what it was then.
10	the	- all both branded considered iconic major once popular the undoubtably	- both considered iconic popular the	
11	became	became become considered for numerous origins other were were viewed as	became were	
12	that	- and business chance he I she shoe (?) someone something that they travelling umbrella’s which work	that which	
13	streets	bridge east scene streets	streets	

		tour tower walk		
14	layers	layers pairs sets	layers pairs sets	
15	pots	- and from houses maids pans pots pipes roofs rooms water	pots	
16	being	appearing being looking seeming	appearing being looking	
17	worried	- cared complained joked just knew protested raced said spoilt spoke talked talks thought walked was went woried (sic) worried wrote	complained protested spoke talked walked went worried wrote	In this case, the text itself has an unfortunate lapse in it, which makes it impossible to determine who protested and to what.
18	were	became grew proved seemed were	proved were	
19	umbrellas	- hundreds items many medicine none sticks them thousands tights umbrella's umbrellas	hundreds items many them thousands umbrellas	
20	stretched	- and	- stitched	

		draped drapped fastened/attached fabric fibres fixed from wool material placed pulled sheet spread stitched stretched stuck tied wool wraped	stretched	
21	granted	- created developed established filed found given granted in into invented issued made out processed recorded registered registers released signed sold	granted issued recorded registered registers	
22	By	before by during however in until	by during in	
23	people	bookworms devices enthusiasts everyone individuals kindles people readers technology themselves us users you	bookworms enthusiasts everyone individuals people readers us users you	

24	too	- any much noticeably physically really that too very	noticeably really that too very	
25	a	one a 1 0.001	one a 1 0.001	
26	keeping	- accessing has have holding holds keep keeping keeps rather stopping storing then trapping	holding keeping storing	
27	of	in of North southern	of	
28	formula	- calculation equation formula formulae theory	equation formula theory	
29	calculated	added agrees argued believes calculated calculates claimed concluded estimated explained found noted predicted suggested said says stated states	argued believes calculated calculates claimed concluded estimated explained suggested said says stated states	
30	of	of	of	

SET3

#	Original	From participants	Accepted	Notes
1	his	helping his hostage other previous the	his	
2	the	the their	the	
3	ransom	- cash exchange funds gold ransom randsom(sic) ransome (sic) ransum (sic) return	cash exchange gold ransom randsom (sic) ransome (sic) ransum (sic)	Unlike with potential apostrophe errors, in these words it is obvious what the respondent is trying to offer as an answer and they are therefore accepted.
4	in	at for from in of of the	at from in of	
5	to	and however of to	to	The context/meaning of the text is precisely that it is the instinct that is suppressed and therefore only 'to' is acceptable.
6	pirates	captors offenders pirates people	captors pirates	
7	real	- British economic financial modern money normal outside real western whole wider working	- economic financial modern money normal outside real western wider	
8	million	billion million percent	million	
9	my	our my the their whole	our my their	

10	a	a solid the	a	
11	of	like of per test as/of(?) twenty upto	of	
12	are	are emerge is normally remain remaining they're were	are emerge remain	
13	view	- aim barrier desire focus goal helmet hope idea intent intention interest means mechanism motive mobile objective plan purpose query shield way view	view	
14	provides	comprises of contains has have includes is users	contains has includes	
15	more	alot brunt force less majority more most some	more most	'Brunt', 'force' and 'majority' would require the definite article 'the' with them. The text makes it clear that it is the upper beak that is longer and so, logically, will take the impact: therefore 'more' or 'most' are the acceptable options.

16	constructed	- and apart away coming constructed formed its made protected protection suspended that	constructed formed made	
17	woodpeckers	birds woodpeckers woodpecker's they	woodpeckers they	
18	the	its it's the the tip of their their top	its the their	
19	its	a it's its like skull the the skull this	its the	
20	says	explains claimed concluded reported said says stated suggested suggests thought wrote	explains claimed concluded reported said says stated suggested suggests	
21	long	- 700 747 aircraft had plane long much newly the was	long much	
22	flight	- debut flight flying journey take off trip	debut flight journey trip	

		voyage		
23	this	last next one this	last this	
24	civil	- aeronautical aerospace aeroplane aircraft airline airplane all aviation commercial electrical flight history modern plane recent technological the transport	aeronautical aerospace aircraft all aviation flight	
25	will	- and can material will would	will would	
26	more	more	more	
27	Among	- all all of also among amongst as most of of seats when	among amongst	
28	who	and most that they who whom would	that who	
29	rinse	-- clean cleaned from gently massage massaged messaged raised rinsed seated	clean cleaned rinsed sprayed wash washed	

		soothe splash sprayed that warmed wash washed washes wipe wiped		
30	plane	aircraft bidet loo loos plane system toilet toilets	aircraft plane	

12.3 Comprehension test: background information and perceived readability questionnaire (as set on SurveyMonkey.com)

PRIOR TO TEST:

These background questions are for the purpose of analysing and comparing data as well as for describing the test population sample. These questions do not ask anything that would allow an individual to be identified directly but a participant number is used as a record identifier. Participant number correspondence to individual participant is stored in a single file, stored separately from the data records in a locked facility.

1. **Participant number (filled by researcher)**
2. **2. Age**

- ○ 18 - 25
- ○ 26 - 35
- ○ 36 - 45
- ○ 46 - 55
- ○ 56 -

3. Gender

- ○ Female
- ○ Male

4. Computer and technology skills:
On a scale from one to seven, where one means "not fluent at all" and seven means "extremely fluent", how fluent would you say you are with computers and technology in general?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

5. Do you have a mobile phone?

- ○ Yes
- ○ No

If yes, what is it (make/model)

WHEN COMPREHENSION TEST HAS BEEN COMPLETED:

Perceived ease and fluency

Answer the following questions with reference to your experience today with the three different conditions of reading and completing the comprehension tests. The questions ask how fluent or easy you experienced reading to be, and also ask you to explain what elements made things fluent or not fluent, easy or difficult.

6. Agree / disagree with the following, using a scale from "completely disagree" to "completely agree".

	Completely disagree	-	-	-	-	-	Completely agree
I found the paper format of the test easy to complete.							
I found reading on paper format to be fluent.							

Why?

7. Agree/disagree with the following statemens with reference to the mobile device used when there were no modifications made to the viewing conditions:

	Completely disagree	-	-	-	-	-	Completely agree
I found the mobile phone format (without modifications) of the test to be easy to complete.							
I found reading on the mobile phone format (without modifications) to be fluent.							

Why?

8. Agree/disagree with the following statements with reference to the mobile phone used when the viewing conditions and layout were adjusted to your liking.

	Completely disagree	-	-	-	-	-	Completely agree
I found the mobile phone format of the test that I had modified myself to be easy to complete.							
I found reading on the mobile phone format that I had mysel modified to be fluent.							

Why?

9. Which of the three presented reading formats (paper, basic mobile, personalized mobile) offers the best reading experience for you?

- ○ Paper
- ○ Basic mobile
- ○ Personalized mobile
- ○ Other

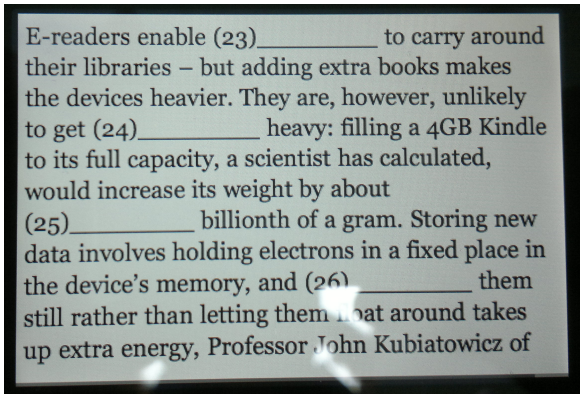
Why?

12.4 Perceived comfort and fluency in reading: Interview guide

#	Question / info to give	Additional prompts
1	What sort of things do you read in your everyday life?	**Printed:** • newspapers / magazines • books: novels (fiction or fact, hardback or soft?), other? • letters / cards **Electronic:** • web versions of newspapers or magazines? • other news sites (news, entertainment, etc.) • blogs • newsgroups • e-mails, etc.
2	Are there other things that you read daily that you may not think of as reading,?(such as …)	• adverts • time tables for trains/busses • your own shopping notes • ??
3	Can you say if you have a favourite format for reading?	• Books (hard bound or soft?), mags, papers…? • Describe the situation when and where you normally read different things?
4	What sort of reading is your favourite in general?	• Fiction, fact, short stories, little articles, long columns, political commentators…? => what do you turn to first thing when you go to a bookshop or get today's paper?
5	Have you ever thought about what sort of things / elements are a "fun" read?	

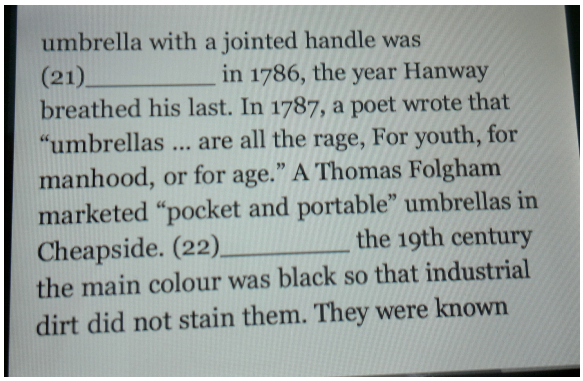
6	How long at a time?	
7	Are there situations when you read things just because you have to?	• How do you motivate yourself? • Do you use coping tactics like “skimming” the text or skipping bits of it?
8	Do you ever pay attention to things like size or colour of text, page layout, size of a page/book, etc.?	• What sort of elements do you associate with making reading more fluent, easy and comfortable? • What sort of elements, similarly, make things harder and slower to read?
9	Do you read things on the computer screen much?	• blogs, papers, mags, e-mail, messenger, anything at all? • e-books or articles?
10	How would you describe reading on the screen compared to reading printed material?	
11	If you think of your most regular type of reading, can you imagine reading it on	• computer screen • mobile phone (show example devices)
12	I have opened a newspaper article on a newspaper website onto this mobile. You can move the page with the scroll on the smaller phone and by dragging with your finger on the bigger one. Take a look, and tell me what sort of thoughts this brings up in terms of you reading it on this device?	• Can you think of doing this much? Why? • What things are there that you find familiar, and what sort of things make it feel unfamiliar?

12.5 Comprehension test: Participant text layout and adjustment examples



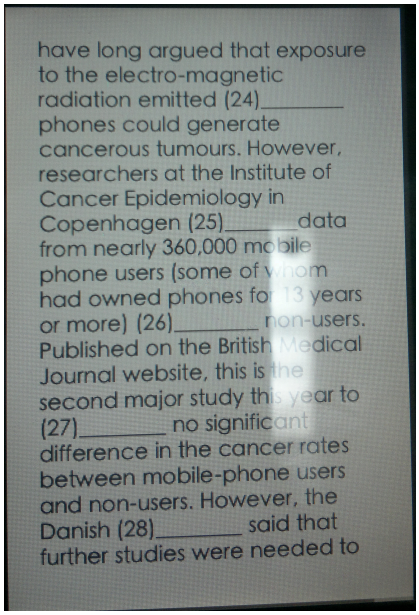
E-readers enable (23)_____ to carry around their libraries – but adding extra books makes the devices heavier. They are, however, unlikely to get (24)_____ heavy: filling a 4GB Kindle to its full capacity, a scientist has calculated, would increase its weight by about (25)_____ billionth of a gram. Storing new data involves holding electrons in a fixed place in the device's memory, and (26)_____ them still rather than letting them float around takes up extra energy, Professor John Kubiatowicz of

Larger font, landscape orientation, reduced margins.



umbrella with a jointed handle was (21)_____ in 1786, the year Hanway breathed his last. In 1787, a poet wrote that "umbrellas ... are all the rage, For youth, for manhood, or for age." A Thomas Folgham marketed "pocket and portable" umbrellas in Cheapside. (22)_____ the 19th century the main colour was black so that industrial dirt did not stain them. They were known

Larger font, landscape, increased line spacing.



have long argued that exposure to the electro-magnetic radiation emitted (24)_____ phones could generate cancerous tumours. However, researchers at the Institute of Cancer Epidemiology in Copenhagen (25)_____data from nearly 360,000 mobile phone users (some of whom had owned phones for 13 years or more) (26)_____ non-users. Published on the British Medical Journal website, this is the second major study this year to (27)_____ no significant difference in the cancer rates between mobile-phone users and non-users. However, the Danish (28)_____ said that further studies were needed to

Portrait, sans serif typeface.

12.6 Small screen readability survey: online questionnaire

Question 1:

The present survey is interested in your views and opinions regarding reading text on the screens of small mobile devices such as mobile phones and PDAs (Personal Digital Assistant). Text is presented to us on mobile devices in various situations such as e-mails, long text messages or web pages if you use mobile internet. The present survey aims at learning what makes reading such text a fluent and comfortable experience. If you volunteer to participate you will be asked to complete an anonymous questionnaire.

The questions include a number of participant background questions for the purpose of categorising different mobile device user profiles, but at no time will it ask for any personal details. This research project has received ethical clearance by the Psychology and Sports Sciences Ethics Committee at Northumbria University. If you require further information please contact the Chair of the said committee, Dr Nick Neave by email at nick.neave@northumbria.ac.uk Please tick to confirm you have read and understood the information and agree to take part.

□ Yes, I have read and understood the information and agree to take part in the survey.

Question 2:

Age:
□ - 17
□ 18 - 24
□ 25 – 30
□ 31 - 45
□ 46 –

Question 3:

Gender:
□ Female
□ Male

Question 4:

First language:
□ English
□ Other

Question 5:

Which of the following features or applications do you use on your mobile telephone?

□SMS (text messages)
□ Multimedia messages
□ Camera / Video camera
□ Games
□ Internet

□ E-mail
□ Social networking applications (Facebook, Twitter, etc.)
□ E-book readers
□ Other (please specify)

Question 6:
How long have you used your current mobile phone?
□Less than 3 months
□3 - 6 months
□Over 6 months

Question 7:
How frequently do you use a computer (desktop or laptop)?
□Daily
□A few times a week
□Once a week or less

Question 8:
In general, how interested are you are in new technologies and electronic devices? Answer on a scale from 1 to 7 (below), where 1 indicates very low interest and 7 extremely high interest.

1 (low interest)	2	3	4	5	6	7 (high interest)

Question 9:
In your opinion, how fluent and competent do you estimate you are in using computers and other digital devices such as mobile phones, games consoles, music players, etc? Answer on a scale from 1 to 7 (below), where 1 indicates extremely low level of fluency and 7 indicates an extremely high level of fluency.

1 (not fluent)	2	3	4	5	6	7 (extremely fluent)

Question 10:
How fluent do you find reading on the following formats? Answer scale: 1 = extremely poor fluency, 7 = extremely high fluency.

		Poor fluency		Extremely high fluency

10a	Reading text on paper (for example books, newspapers, magazines, etc.)	1	2	3	4	5	6	7
10b	Reading text on computer screens (desktop or laptop)	1	2	3	4	5	6	7
10c	Reading text on mobile phones	1	2	3	4	5	6	7

Question 11:
How comfortable do you find reading when the text is presented in the following formats? Answer scale: 1 = extremely uncomfortable, 7 = an extremely comfortable.

		Very low importance					Extremely high importance	
11a	Reading text on paper (for example books, newspapers, magazines, etc.)	1	2	3	4	5	6	7
11b	Reading text on computer screens (desktop or laptop)	1	2	3	4	5	6	7
11c	Reading text on mobile phones	1	2	3	4	5	6	7

Question 12:
How important do you find the following items in being able to read text fluently and comfortably? Answer scale: 1 = very low importance, 7 = extremely high importance.

		Very low importance					Extremely high importance	
12a	Width of columns / row of text on a page	1	2	3	4	5	6	7
12b	Orientation of the screen or a page (landscape or portrait)	1	2	3	4	5	6	7
12c	Use of headings, paragraphs and other items that indicate structure of content	1	2	3	4	5	6	7

Question 13:
How important or not important is the design and appearance of the following elements to your reading experience? Answer scale: 1 = the design or appearance does not matter at all, 7 = the design or appearance is highly important to the reading experience.

		Very low importance					Extremely high importance	
13a	Size of text	1	2	3	4	5	6	7
13b	Typeface used	1	2	3	4	5	6	7
13c	Use of colour	1	2	3	4	5	6	7
13d	Contrast	1	2	3	4	5	6	7
13e	Resolution (sharpness) of screen	1	2	3	4	5	6	7
13f	Ability to se an overview of the whole content	1	2	3	4	5	6	7
13g	Ability to estimate how much content is available in addition to what is visible at one time	1	2	3	4	5	6	7
13h	Clarity of links and directions to related or further sections to the content	1	2	3	4	5	6	7
13i	Ability to filter visible content (for example, expanding lists)	1	2	3	4	5	6	7
13j	Ability to adjust size of text	1	2	3	4	5	6	7
13k	Ability to adjust the overall zoom level of content	1	2	3	4	5	6	7
13l	Ability to adjust the orientation of a page (landscape or portrait)	1	2	3	4	5	6	7
13m	Ability to move content on screen	1	2	3	4	5	6	7

13n	Ease of selecting links, menus, expanding lists, dropdown menus, etc.	1	2	3	4	5	6	7

Question 14:
How important are the following elements to you in creating a fluent reading experience? Answer scale: 1 = very low importance, 7 = very high importance.

		Very low importance						Extremely high importance
14a	Visibility of text	1	2	3	4	5	6	7
14b	Amount of content clearly visible on the screen	1	2	3	4	5	6	7
14c	An overview of content (even if details are not fully visible)	1	2	3	4	5	6	7
14d	Ease of adjusting the visibility of content on screen	1	2	3	4	5	6	7
14e	Ease of moving the contents on the screen	1	2	3	4	5	6	7
14f	Personal interest in the topic / subject matter	1	2	3	4	5	6	7
14g	Existing knowledge about the topic / subject matter	1	2	3	4	5	6	7
14h	Style of writing (how complex or simple the language is)	1	2	3	4	5	6	7
14i	Length of sentences and phrases	1	2	3	4	5	6	7
14j	Seeing more than one sentence/phrase at a time	1	2	3	4	5	6	7

14k	Electronic content looking the same as if it was in print format (for example, newspaper style on a website is similar to what the actual paper newspaper looks)	1	2	3	4	5	6	7

Question 15:
How much do you agree or disagree with the following statements on reading? Answer scale: 1 = completely disagree, 7 = completely agree.

		Completely disagree						Completely agree
15a	Personal interest in subject matter affects reading speed.	1	2	3	4	5	6	7
15b	Personal interest in subject matter affects comprehension of text.	1	2	3	4	5	6	7
15c	Generally, I am a fast reader.	1	2	3	4	5	6	7
15d	Generally, I am an attentive reader.	1	2	3	4	5	6	7
15e	I read frequently and regularly.	1	2	3	4	5	6	7
15f	Different writing styles for different publication types support reading comprehension.	1	2	3	4	5	6	7
15g	I have a broad vocabulary.	1	2	3	4	5	6	7
15h	I read professional specialist texts for my work or studies.	1	2	3	4	5	6	7

Question 16:
How important are the following elements to you in deciding to read or not to read a piece of text on a mobile phone screen? Answer scale: 1 = very low importance, 7 = very high importance.

		Very low importance		Extremely high importance

16a	The information is needed immediately.	1	2	3	4	5	6	7
16b	I am very interested in the topic/subject matter of the text.	1	2	3	4	5	6	7
16c	The feeling that reading is fluent.	1	2	3	4	5	6	7
16d	The feeling that reading is a comfortable experience.	1	2	3	4	5	6	7
16e	The ease of interacting with the device I'm using to read the text.	1	2	3	4	5	6	7
16f	Familiarity with the topic/subject matter.	1	2	3	4	5	6	7

Question 17:
How likely would you estimate it would be for you to read the following types of text on a mobile device screen. Answer scale: 1 = not at all likely, 7 = very likely.

		Not likely					Very likely	
17a	Latest news headlines.	1	2	3	4	5	6	7
17b	Newspaper website.	1	2	3	4	5	6	7
17c	Novels.	1	2	3	4	5	6	7
17d	Newsgroups or blogs.	1	2	3	4	5	6	7
17e	Sports results.	1	2	3	4	5	6	7
17f	Social networking sites.	1	2	3	4	5	6	7
17g	User guides.	1	2	3	4	5	6	7
17h	Reports or other professional communications.	1	2	3	4	5	6	7
17i	Learning materials.	1	2	3	4	5	6	7

Question 18:

Thank you for taking part in this survey. If you wish to give any feedback on this questionnaire, please write your comments below:

[open text field provided]

Question 19:
If you would like to receive general information regarding the results of this study, please leave your e-mail address in the space below. This information will be available in approximately six months from now.

[open text field provided]

References

90percentofeverything.com (2009). Just add an egg – Usability, User Experience and Dramaturgy. Http://www.90percentofeverything.com/2009/10/20/just-add-an-egg-usability-user-experience-and-dramaturgy/ (date of reference: 7.2.2013.)

About.com's Web Trends (2013). Top Mobile Web Sites You Have Bookmarked. Http://webtrends.about.com/od/mobileweb20/tp/Top-Mobile-Websites-.htm (date of ref. 1.1.2013.)

Afflerbach, P., Pearson, P. D. and Paris, S. G. (2008), Clarifying Differences Between Reading Skills and Reading Strategies. *The Reading Teacher*, 61: 364–373. doi: 10.1598/RT.61.5.1

Ahn, L. von, Maurer, B., McMillen, C., Abraham, D., Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science* 12 September 2008 321: 1465-1468. Reproduced - http://recaptcha.net/reCAPTCHA_Science.pdf

Anderson, R. & Davison, A. (1988). Conceptual and Empirical Bases of Readability Formulas. In Davison, A. & Green. G. M. (eds).

Bailey, R. W., 1993, Performance vs preference. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 282–286.

Balota, D.A and Rayner, K. (1991). Word Recognition Processes in Foveal and Parafoveal Vision: The Range of Influence of Lexical Variables. In Besner, D. and Humpreys, G.W. (eds). (1991). *Basic Processes in Reading: Visual Word Recognition*. Lawrance Erlbaum Associates, Publishers, New York.

Baker, E., Atwood, N. & Duffy, T. (1988). Cognitive Approaches to Assessing the Readability of Text. In Davison, A. & Green, G. M. (eds), 55-83.

Baker, J. Ryan. (2005). Is Multiple-Column Online Text Better? It Depends! *Usability News, Usability News*, 7(2),

Bernard, M.L., Chaparro, B.S., Mills, M.M., Halcomb, C.G. (2003). Comparing the Effects of text Size and Format on the Readability of Computer-Displayed Times New Roman and Arial Text. *International Journal of Human-Computer Studies*, 59 : 823-835.

Bernard, M., Mills, M., Peterson, M., and Storrer, K. (2001). A Comparison of Popular Online Fonts: Which is Best and When? *Usability News*,

3(2).

Besner, D. and Humpreys, G.W. (eds). (1991). *Basic Processes in Reading: Visual Word Recognition*. Lawrance Erlbaum Associates, Publishers, New York.

Boyarski, D., Neuwirth, C., Forlizzi, J., and Regli, S. H. (1998). A study of fonts designed for screen display. *Proceedings of CHI'* 98, 87-94.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. *Usability Evaluation in Industry*. London: Taylor and Francis.

Bruijn, David de, Mul, Sjaak de and Oostendorp, Herre van (1992): The Influence of Screen Size and Text Layout on the Study of Text. *Behaviour and Information Technology*, 11 (2) pp. 71-78.

Bruce, B. C. & Rubin, A. D. (1988). Readability Formulas: Matching Tool and Task. In Davison, A. & Green, G. M. (eds) 5-22.

Colley, A., Huhtala, J., Tikka, P. Hakkila, J. (2012). Mobile GUI Legibility – Comparing Automated vs. User Study Based Evaluation. Unpublished short paper (work in progress).

Cattell, J. (1886). The time taken up by cerebral operations. *Mind*, 11, 277-282, 524-538.

Collins English Dictionary. (1999). HarperCollins Publishers. First edition 1979.

Costello, A.B., Osborne, J.W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assesment, Research & Evaluation*. Volume 10 Number 7. pareonline.net/pdf/v10n7.pdf (date of reference August 13, 2012).

Daneman, M., & Carpenter, P.A.,(1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19, 450 – 460*.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3): 319-340.

Davison, A. & Green, G.M. (eds), (1988). *Linguistic Complexity and Text Comprehension: Readability Issues Reconsidered*. Hillsdale, New Jersey: Lawrance Erlbaum Associates Publishers.

DeMaio, T.J. & Landreth, A. (2004). Do Different Cognitive Interview Techniques Produce Different Results? In Presser, S. et al. (eds) (2004). *Methods for Testing and Evaluating Survey Questionnaires*. John Wiley & Sons, Inc., Hoboken New Jersey.

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

Dillon, A., Richardson, J. and McKnight, C. (1990). The effect of display size and paragraph splitting on reading lengthy text from screen. *Behaviour and Information Technology*, 9(3), 215-227.

DuBay, W.H. (2004). The Principles of Readability. *Impact Information*. Costa Mesa, CA. Http://www.nald.ca/library/research/readab/readab.pdf (date of reference: 7.2.2013).

Dyson, M.C. (2004). How physical text layout affects reading from screen. *Behaviour & Information Technology*, Volume 23, Number 6, November-December 2004. Taylor & Francis, 377-393. DOI=http://dx.doi.org/10.1080/01449290410001715714.

Dyson, M.C., and Haselgrove, M. (2001). The influence of reading speed and line length on the effectiveness of reading from screen. *Int. J. Hum.-Comput. Stud*. 54, 4 (April 2001), 585-612.

Dyson, M.C., and Kipping, G.J. (1997). The legibility of screen formats: Are three columns better than one? *Computers and Graphics*, 21, 6 (Dec. 1997), 703 -712.

Ellis, A. (1983) *Reading, Writing and Dyslexia*. London: Lawrance Erlbaum Associates.

Enkvist, N. E., & Kohonen, V. (1976). Cloze-testien teoriaa. *AFinLa:n julkaisut 1971-1979*, (13), 31–54.

Erdfelder, E., Faul, F. and Buchner, A. (1996). About G*Power. Http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/about/about_gpower_01.html (date of reference 8.1.2013).

Eysenck, M. and Keane, M.T. (2006). *Cognitive Psychology: A Student's Handbook.* Psychology Press, Hove and New York. First pub. 2005.

Farstrup, A.E. & Samuels, J.S. (eds) (2002). *What Research Has to Say About Reading Instruction*. Delaware: International Reading Association.

Flanagan, J.C., (1954). The critical incident technique. *Psychological Bulletin* 5, 327–358.

Flesch R,(1948). A new readability yardstick. *Journal of Applied Psychology* 32: 221–233.

Florida Laws: FL Statutes - Title XXXVII Insurance Section 627.011 Short title. http://law.onecle.com/florida/insurance/627.4145.html (date of reference 02.01.2013).

Foddy, W. (1994). *Constructing Questions for Interviews and Questionnaires*. Cambridge University Press.

Frase, L.T. and Schwartz, B.J. (1979). Typographical Cues that Facilitate Comprehension, *Journal of Educational Psychology*, 71, 2 (Apr. 1979), 197 – 206.

Gibson, E., (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 1998, Vol. 68, Issue 1, pp. 1 – 76.

Gigaom.com, (2010). Mary Meeker: Mobile Internet Will Soon Take Over Fixed Internet. Http://gigaom.com/2010/04/12/mary-meeker-mobile-internet-will-soon-overtake-fixed-internet/ (date of reference 24.01.2013).

Gilliland, J. (1972). *Readability*. University of London Press Ltd.

Gould, J.D., Alfaro, L., Finn, R., Haupt, B., Minuto, A. and Salaun, J. (1987). Why reading was slower from CRT displays than from paper, *ACM SIGCHI Bulletin*, Vol. 18, No. 4, pp. 7 -11.

Graeser, A.C., Singer, M. and Trabasso,T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 3, 371-395.

Grzeschik, K., Kruppa, Y., Marti, D., and Donner, P. (2011). Reading in 2110 – reading behaviour and reading devices: a case study. *The Electronic Library*, Vol.29 No. 3, 2011, pp. 288 – 302.

Haber, R.N. & Schindler, R.M. (1981). Errors in proofreading: Evidence of syntactic control of letter processing? *Journal of Experimental Psychology: Human Perception and Performance*, 7, 573-579.

Henson, R.K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counselling and Development*, 34, 177-189.

Hudson, R.A., (1995). Measuring syntactic difficulty. Manuscript, University College, London. Unpublished paper.

Huenerfauth, M., Feng, L., Elhadad, N. (2009). Comparing Evaluation Techniques for Text Readability Software for Adults with Intellectual Disabilities. *ASSETS'09*, October 25–28, 2009, Pittsburgh, Pennsylvania, USA

Hyvärinen, T., Kaikkonen, A. and Hiltunen, M. (2005). Placing Links in Mobile Banking Application. *MobileHCI'05*, September 19-22, Salzburg, Austria.

Hyönä, J. (1993). *Eye Movements During Reading and Discourse Processing*. University of Turku Psychological Research Reports 65. Academic dissertation. University of Turku.

ISO 9241-11, Ergonomic requirement for office work with visual display terminals (VDTs) – Part 11 Guidance on usability. (1998). Http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=16883 (date of ref. 29.11.2012).

Just, M.A., Carpenter, P.A., (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review* 99, 122–149.

Järvinen, P. and Järvinen, A. (2000). *Tutkimustyön metodeista*. Opinpajan kirja, Tampere.

Kaasinen, E., Roto, V., Roloff, K., Väänänen-Vainio-Mattila, K., Vainio, T., Maehr, W., Joshi, D., and Shrestha, S., (2009). User Experience of Mobile Internet: Analysis and Recommendations. *International Journal of Mobile Human Computer Interaction*, Vol. 1, Issue 4. IGI Publishing.

Kaikkonen, A. (2009). *Internet on Mobiles: Evolution of Usability and User Experience.* Doctoral dissertation. Helsinki University of Technology. http://lib.tkk.fi/Diss/2009/isbn9789522481900/ (date of reference 22.01.2013).

Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.

Klare, G.R. (1963). *The measurement of readability.* Ames, IA: Iowa State University Press.

Krug, S. (2006) *Don't Make Me Think! A Common Sense Approach to Web Usability. Second edition*. New Riders Publishing, Berkeley, California.

Laarni, J. (2002). Searching for Optimal Methods of Presenting Dynamic Text on Different Types of Screens. *NordicCHI 10/02*, Århus, Denmark. ACM.

Lamberg, E.M., Muratori, L.M., (2011). Cell phones change the way we walk. *Gait & Posture* 35 (2012) 688–690.

Legge, G., & Rubin, G. (1986). Psychophysics of reading. IV. Wavelength effects in normal and low vision. *Journal of the Optical Society of America A, 3*, 40-51.

Legge, G., Rubin, G., and Luebker, A. (1987). Psychophysics of reading V. The role of contrast in normal vision. *Vision Res*. 27, 1165-1177 (1987).

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology,* 140: 1 – 55.

Lund, A.M. (2000). Measuring Usability with the USE Questionnaire. STC Usability SIG Newsletter, 8:2. Http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html Originally: *Usability & User Experience*, October 2001 issue (Vol. 8, No. 2) (date of reference 10.2.2013).

MacLean, A., Barnard, P. J., & Wilson, M. D. (1985). Evaluating the human interface of a data entry system: user choice and performance measures yield different tradeoff functions. In P. Johnson and S. Cook (eds.) (1985). *People and Computers: Designing the User Interface*. Cambridge University Press, Cambridge, pp.172-185.

Mills, C.B. and Weldon, L.J. (1987). Reading text from computer screens. *ACM Computer Surveys*, 19(4).

Monk, A.F. & Hulme, C. (1983). Errors in proofreading: Evidence for the use of word shape in word recognition. *Memory and Cognition*, 11, 16-23.

Nielsen, J. (2000). Hard to use sites will fail. *The Irish Times*. Http://www.ireland.com/newspaper/computimes/2000/0110/compu1.htm (date of reference 10.1.2000).

Nielsen, J. (1993). *Usability Engineering*. San Diego: Academic Press.

Nielsen, J. and Levy, J. (1994). Measuring usability: Preference vs. performance. *Communications of the ACM*, Vol. 37, no. 4, pp. 66-75.

Nokia (2011). Http://conversations.nokia.com/2011/02/11/mobile-phones-the-next-billion/ (date of reference 6 August 2012).

Norman, D.A. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books. New York.

Oostendorp, H. van (ed.), (2003). *Cognition in a Digital World*. Mahwah, New Jersey: Lawrence Elrbaum Associates, Inc., Publishers, 2003.

Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. (2005). Interaction in 4-Second Bursts: The Fragmented Nature of Attentional Resources. *Mobile HCI*. CHI'05, Portland, Oregon.

Paas, F. and van Merrienboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors*, 35, 737-743.

Paas, F. and Merrienboer, J. van (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psycholog*y, 86, 122-133.

Pastoor, S. (1990). Legibility and subjective preference for color combinations in text. *Human Factors, 32*, 157-171.

Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods*. SAGE Publications. Newbury Park, London, New Delhi. Second edition. First pub. 1980.

Pirozzolo, F.J. and Wittrock, M.C. (eds). (1981). *Neuropsychological and Cognitive Processes in Reading.* Academic Press, New York.

Presser, S. et al. (eds) (2004). *Methods for Testing and Evaluating Survey Questionnaires*. John Wiley & Sons, Inc., Hoboken New Jersey.

Pressley, M. (2002). Metacognition and Self-Regulated Comprehension. In Farstrup, A. & Samuels, J. S. (eds), 291-309.

Qiu, M.K., Zhang, K., and Huang, M., (2004). An Empirical Study of Web Interface Design on Small Display Devices. *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence* (WI'04).

QUIS questionnaire, University of Maryland. Http://www.lap.umd.edu/quis/ (date of reference: 7.2.2013).

Rayner, K., Slattery, T.J., Drieghe, D. and Liversedge, S.P. (2011). Eye Movements and Word Skipping During Reading: Effects of Word Length and Predictability. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 37, No. 2, 514 – 528.

Redish, J. (2000). Readability formulas have even more limitations than Klare discusses. *ACM Journal of Computer Documentation*, *24*(3), 132–137. doi:10.1145/344599.344637.

Reicher, G.M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275-280.

Samuels, J.S. and Eisenberg, P.A. A Framework for Understanding the Reading Process. In Pirozzolo, F.J. and Wittrock, M.C. (eds). (1981). *Neuropsychological and Cognitive Processes in Reading.* Academic Press, New York.

Scharff, L. F. V., Hill, A. L. and Ahumada, A.J. Jr. (1999). Discriminability measures for predicting readability of text on textured backgrounds. *Optics Express 81, Vol.6, No. 4.* Revised February 2000.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, *14*(4), 433–440. doi:10.1080/026999300402745.

Shrestha, S. (2007). Mobile Web Browsing: Usability Study. *Mobility'07 Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology.* Pages 187 – 195. ACM New York, NY, USA, 2007. DOI: 10.1145/1378063.1378094

Snitker, T. V. (2004). *Breaking Through to the Other Side: Using the User Experience in Web, Interactive TV and Mobile Services*. Ny Teknisk Forlag.

Spence, R. (2007). *Information Visualization: Design for Interaction*. Pearson Education Limited, Harlow. First pub. 2001.

Stanza e-reader application for iPhones. Http://www.mystanza.com/ (date of ref. 14 July 2012).

Stein, G. (1960). *Autobiography of Alice B. Toklas*. New York: Random House, Inc. First pub. 1933.

Stevens, J.P. (2002). *Applied multivariate statistics for the social sciences*. 4th edition. Hillsdale, NJ: Erlbaum.

Strauss, A. and Corbin, J. (1990). *Basics of qualitative research - Grounded Theory Procedures and Techniques.* Sage Publications, Newbury Park, Ca.

Sweller, J. (1994). Cognitive Load Theory, learning difficulty, and instructional design. *Learning and Instruction* **4** (4): 295–312. doi:10.1016/0959-4752(94)90003-5

Sweller, J., and Chandler, P. (1994). Why some material is difficult to learn. Cognit. Instr. 12: 185-233. In Oostendorp, Herre Van (ed.),

(2003). *Cognition in a Digital World*. Mahwah, New Jersey: Lawrence Elrbaum Associates, Inc., Publishers, 2003.

Sweller, J., Merrienboer, J. J. G. V., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review,* 10(3), 251–296.

Thaler, R.H. and Sunstein, C.R. (2008). *Nudge: Improving Decisions, About Health, Wealth and Happiness*. Yale University Press. New Haven & London.

Tinker, M.A., (1963). *Legibility of Print*. Iowa State University Press, Ames, IA.

Usabilitynet.org. Http://www.usabilitynet.org/tools/criticalincidents.htm (date of reference August 16, 2012).

Useit.com. Http://www.useit.com/alertbox/20000319.html) (date of reference Aug 16, 2012).

Waycott, J. and Kukulska-Hulme, A. (2003). Students' experiences with PDAs for reading course materials. *Pers Ubiquit Comput* 7: 30–43 DOI 10.1007/s00779-002-0211-x.

Webitect (2013). Webitect: A resource for webmasters. The Top 10 Typography Rules All In One Place. Http://webitect.net/design/the-top-10-typography-rules-all-in-one-place/ (date of reference 11.2.2013).

Williams, S.W., Ari, O., and Santamaria, C.N., (2011). Measuring college students' reading comprehension ability using Cloze tests. *Journal of Research in Reading, Vol. 34, Issue 2*, pp. 215 – 231.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology* **1** (1): 59–85. DOI:10.1016/0010-0285(70)90005-8. Retrieved 2008-06-30.

Wilson, R., Landoni, M., and Gibb, F. (2002). A user-centered approach to e-book design. *The Electronic Library*, 20, 4, 322-330.

Yi, W., Park, E. Cho, K., (2011), E-Book Readability, Comprehensibility and Satisfaction. *ICUIMC' '11*, February 21–23, 2011, Seoul, Korea.

Zibell, K. (2000). Most readability principles apply to Web-site design: Klare's "Useful Information" is Useful for Web Designers. *ACM Journal of Computer Documentation*, 24:141 – 147. Vol. 24, No. 3.

Öquist, G. and Lundin, K. (2007). Eye Movement Study of Reading Text on a Mobile Phone using Paging, Scrolling, Leading, and RSVP.

Mobile and Ubiquitous Multimedia; Vol 284: Proceedings of the 6th international conference on Mobile and ubiquitous multimedia. ACM Digital Library, New York.