TU WIEN Informatics

A Study of Multi-Document Active Reading in Analog and Digital Environments

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Software Engineering & Internet Computing

eingereicht von

Jasmin Mahler, BSc.
Matrikelnummer 01325188

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.Ass. Dr.techn. Manuela Waldner, MSc.

Wien, 8. April 2021

Jasmin Mahler

Manuela Waldner

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

A Study of Multi-Document Active Reading in Analog and Digital Environments

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Software Engineering & Internet Computing

by

Jasmin Mahler, BSc.
Registration Number 01325188

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Ass. Dr.techn. Manuela Waldner, MSc.

Vienna, 8th April, 2021

Jasmin Mahler | Manuela Waldner

Erklärung zur Verfassung der Arbeit

Jasmin Mahler, BSc.

Hiermit erkläre ich, dass ich diese Arbeit selbstständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 8. April 2021

Jasmin Mahler

Acknowledgements

Throughout the writing of this diploma thesis, I have received a great deal of support and assistance.

First of all, my sincere thanks go to Manuela Waldner, who supported me throughout my thesis and always stood by my side. Thank you for your supervision and assistance, for your invested time and for always believing in me. You kept finding new ways to motivate me, in good times as well as in hard times. Thank you for your engagement, as well as your expertise and ideas. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I am so grateful to have you as my advisor. Without you, I would definitely not have reached the end of my studies by now.

Special thanks go to my sister and all my friends who took the time and participated in my study with great enthusiasm. Without them, this work would not have been possible. My thanks go to their willingness and their interesting contributions and answers to my questions.

In addition, I would like to thank my parents for their wise counsel and sympathetic ear. You are always there for me, not only as support during the thesis and my studies but throughout my entire life. Finally, I could not have completed this thesis without the support of Simone, Mathias, Melanie and Lukas, who provided stimulating discussions and opinions and reminded me never to quit and always keep going as well as provided happy distractions to rest my mind outside of my research.

Kurzfassung

Trotz der vielen Verbesserungen im digitalen Bereich wechseln Wissensarbeiter immer noch häufig zwischen digitalen und analogen Materialien und Werkzeugen während ihrer Arbeit. Dabei nehmen sie unter anderem Zeit- als auch Ressourcenkosten in Kauf, um in ihrer bevorzugten (analogen) Umgebung aktives Lesen und damit verbundene Tätigkeiten durchführen zu können. Frühere Studien belegen, dass aktives Lesen mit analogen Materialien und Werkzeugen effizienter ist als unter der Verwendung von digitalen Materialien und Werkzeugen. Jedoch ist bis heute nicht vollständig geklärt, was genau zur Überlegenheit von analogem aktivem Lesen gegenüber dem digitalen führt.

Ziel dieser Diplomarbeit ist ein direkter Vergleich der Verhaltensweisen und eingesetzten Strategien von Benutzern während des aktiven Lesens mehrerer Dokumente in analogen und digitalen Umgebungen. Dieser Vergleich dient dazu, genauere Einblicke zu bekommen, in welchen (Teil-)Bereichen des aktiven Lesens (annotieren, markieren, Notizen schreiben und räumliche Organisation) in den beiden Umgebungen Unterschiede vorhanden sind, was etwaige Gründe für diese Unterschiede sein könnten und vor allem auch, wie die Erfahrung des digitalen aktiven Lesens in Zukunft verbessert werden könnte. Im Zuge des Vergleichs lässt sich auch feststellen, ob analoges aktives Lesen immer noch besser als digitales aktives Lesen ist, wenn im digitalen Bereich ein großer Bildschirm verwendet wird, der ähnlich viel Platz wie im analogen bietet.

In einer qualitativen, kontrollierten, teils konfirmatorischen, teils explorativen Benutzerstudie werden somit die Verhaltensweisen und Strategien von Benutzern während dem aktiven Lesen mehrerer Dokumente in analoger und digitaler Umgebung verglichen, um die zuvor genannten Aspekte zu untersuchen. Die Ergebnisse zeigen, dass trotz der Verwendung eines großen Bildschirms analoges aktives Lesen immer noch effizienter ist als digitales aktives Lesen. Zusätzlich konnten in der Evaluierung Unterschiede in den Verhaltensweisen und Anpassungen der eingesetzten Strategien aufgrund der Zugänglichkeit und Verfügbarkeit von Materialien und Werkzeugen festgestellt werden. Insbesondere im Bereich der räumlichen Organisation während des digitalen aktiven Lesens gibt es noch erhebliches Verbesserungspotential.

Abstract

Despite the many improvements in the digital domain, knowledge workers still frequently switch between digital and analog materials and tools during their work. In doing so, they accept "switching costs" (such as time and resources) to perform active reading and related activities in their preferred (analog) environment. Previous studies show that active reading is more efficient using analog materials and tools than using digital ones. However, up to now, it is not fully understood what exactly leads to the superiority of analog active reading over digital active reading.

The goal of this thesis is to directly compare the behaviors and strategies employed by users during active reading of multiple documents in analog and digital environments. This comparison serves to gain more detailed insights into which (sub-)areas of active reading (annotating, highlighting, note-taking, and spatial organization) are different in the two environments, what might be possible reasons for these differences, and most importantly, how to improve the experience of digital active reading in the future. As part of the comparison, it is also possible to determine whether analog active reading is still more efficient than digital active reading when using a large screen that provides a similar amount of space as an analog workstation.

Thus, in a qualitative, controlled, partly confirmatory, partly exploratory, user study, users' behaviors and strategies during active reading of multiple documents in analog and digital environments are compared to investigate the previously mentioned aspects. The results show that analog active reading is still more efficient than digital active reading despite the use of a large screen. Additionally, the evaluation was able to identify differences in behaviors and adaptations of strategies used due to the accessibility and availability of tools. In particular, there is still considerable potential for improvement in the area of spatial organization during digital active reading.

Contents

CHAPTER 1

Introduction

Reading is an important part of our daily lives. A distinction is made between active and passive reading. With passive reading, it can often happen that even though words are being (technically) read, their context and meaning are not absorbed by the reader, because the reading itself is done less carefully and effortfully. *Active Reading*, on the other hand, combines reading with critical thinking, learning and decision making [SPG+99]. It describes a form of knowledge work that requires a high level of interconnectivity among reading-related tasks, which supports the primary task of deep engagement with the text in the form of immersive reading [HPW+12]. Those reading-related tasks frequently involve acts like annotating, note-taking, searching, highlighting, comparing, and non-sequential navigation [TE11a].

In order to convert information to knowledge, knowledge workers process a large amount of data that they have collected from many different sources on a daily basis [KS11]. During this process, knowledge workers consciously or unconsciously apply different *Active Reading* strategies, such as annotating source documents or taking notes on an extra sheet of paper. Since the processed data mainly consists of multiple documents, that must be cross-referenced [AGH+98], spatial organization as part of knowledge workers' active reading strategy might play a crucial role. For example, documents might be sorted and grouped by their contents' topic, and piles might be placed all across the table depending on their importance.

Even though personal computers play an essential role in our daily life, they have not managed to eliminate the use of paper for office and education work for a long time [MBM07], and paper still remains an essential part of offices and schools up to today. As more and more information is available only in digital form, it is therefore often necessary for knowledge workers to switch between analog and digital tools while performing their tasks. For instance, researchers search for appropriate papers on a specific topic online. For reading, they might print them out and mark and annotate them on the physical paper. However, they draft their final report using text processing software on their computers, and, finally, publish it online. Such an approach is not particularly surprising, considering that there is already some evidence showing that analog active reading is more effective than digital active reading [GCR+19, MWB13, KSZ18, TSO15]. Up to now, it is not fully understood what exactly makes active reading using analog tools so much more effective compared to active reading using digital tools.

Many studies have already been conducted in the field of active reading, though, there are still many open questions, mainly since previous studies were either limited to specific parts of active reading processes (such as annotating and highlighting [GCR+19], note-taking [NA06], or spatial organization [TSO15]), did not make a direct comparison between analog and digital work environments [AEN10, AN13, HPW+12, PML+18, TE11a], or have not dealt with active reading of multiple documents and spatial organization on large displays [OS97, MBM07]. But especially when working with multiple documents, spatial organization plays an essential role, as humans are cognitively well adapted to making use of space to express and perceive relationships between objects [Kir95]. Moreover, active reading is not limited to reading (single) individual documents, but rather focuses on internal mental activities and the creation of new analyses that might not be present in an individual book [AD72].

Therefore, this thesis' goal is a preferably direct comparison of users' active reading behavior in analog and digital environments. This comparison aims not only to gain insights on reading supportive activities, such as highlighting, annotating, and note-taking, but also on better understanding the spatial organization approaches knowledge workers employ to their workspace and reading materials when they are confronted with reading multiple documents just like during a typical day at work. A better understanding of possible differences should then help design better user interfaces of future applications to support and improve active reading processes even more in the digital context.

In the context of this work, therefore, the following research questions will be explored:

Q1. *Is analog multi-document active reading also more effective than digital multi-document active reading when using a large display, which facilitates spatial organization?*

Q2. *Which active reading behaviors do users employ in the analog and digital world?*

Q3. *How do these active reading behaviors differ between the analog and digital worlds?*

Q4. *What adjustments are users doing to their active reading strategies, based on the availability and accessibility of operations and tools?*

Q5. *What might be an appropriate way to transform (missing) active reading elements from the analog to the digital world?*

In the course of this thesis, users' active reading behavior in analog and digital environments will be compared. At the beginning, the necessary theoretical foundation and background information (Chapter 2) as part of a comprehensive literature research are provided. This includes not just definitions (Section 2.1), but also an analysis of the current state of active reading (activities) in analog and digital environments (Section 2.2), as well as studies related to this thesis (Section 2.3). The empirical part of the work (Chapter 3), in the form of a controlled user study to answer the above research questions, is then conducted. It is shown which behaviours users employ during analog and digital active reading, where there are differences, and which active reading environment is, in the end, more efficient. The observations and findings of the study are then summarized (Chapter 4) and discussed in more detail (Chapter 5) to determine what possible reasons there might be for any (behavioral) differences, also in light of existing findings from previous studies. Finally, the findings are used to present design guidelines (Section 6.2) for future user interfaces that tackle identified (multi-document) active reading issues on computers.

The contributions of this thesis can be summarized as follows:

1. The results from a qualitative user study, which is the first to provide a direct comparison of users' typical active reading behaviours (annotating, highlighting, and note-taking) and spatial organization approaches when confronted with multiple documents in analog and digital environments using a large display (Chapter 3 – 5).

2. The definition of design guidelines for user interfaces, in order to tackle identified active reading issues on computers (Section 6.2).

CHAPTER 2

Background

The focus of this thesis is on gaining insights on active reading strategies knowledge workers employ during their sensemaking process in analog and digital settings. This chapter provides the necessary background information. It starts with definitions of *Visual Thinking* and *Active Reading* and then links to knowledge workers' sensemaking process. Next, typical active reading activities, such as reading, annotating, note-taking and (spatial) organization and navigation, during the sensemaking process of knowledge workers in analog and digital environments are discussed in detail. At first, existing work about reading in terms of reading comprehension and behavior differences between paper-based and screen-based reading is discussed. The focus then switches to annotating and note-taking, which are highly connected to reading and supported differently by paper and digital environments. Afterwards, existing work on analog and digital spatial organization and navigation in relation to sensemaking are given. Lastly, related studies and their open questions are listed.

2.1 Definitions

2.1.1 Visual Thinking

Visual Thinking is one of the most frequent cognitive activities people perform when receiving and understanding information. Although the tasks of *Visual Thinking* are common, basic, and primary, *Visual Thinking* plays a fundamental part in knowledge accumulation and management [SF11]. It comes directly from the emotional experience perceived from the physical world [SF11]. It is the ability to conceptualise and present thoughts and ideas, and thus, enhances the learning process of people [Cyr97].

Visual Thinking can be defined as the composition of three overlapping strategies of thought: seeing, imaging, and designing (see Figure 2.1) [Cyr97]. While the seeing part is quite self-explanatory, imaging and designing need a little bit more explaining. *Imaging*

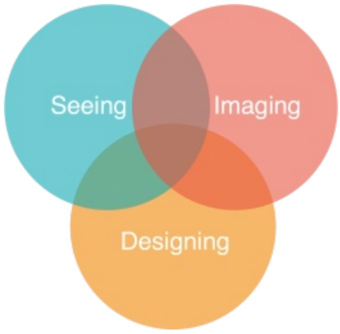


Figure 2.1: Components of Visual Thinking (adapted from E. Cyrs [Cyr97]).

can be put on a level with understanding the perceived information and considering different meanings of it as some kind of internal thinking. *Designing*, on the other hand, involves expressing an idea in some type of visual form such as verbal symbols represented as words [Cyr97]. Such external representations serve as memory aids and allow memory to be shared between people [Zha97]. So in order to extract meaning, draw conclusions, and deepen the understanding of information, people often mark, annotate, write notes and rearrange things [Kir10] as part of their *Visual Thinking* strategies.

2.1.2 Active Reading

In order to convert information to knowledge, knowledge workers process a lot of data that they have collected from many different sources on a daily basis [KS11]. Therefore, active reading is a critical task of knowledge workers. The term *Active Reading* describes a form of knowledge work that requires a high level of interconnectivity among reading-related tasks. It is a broad category description for behaviors that aid in understanding and retaining written content and encompasses both writing and reading activities. Furthermore, it can be characterized by fluid transitions between the primary task of deep engagement with the text in the form of immersive reading and a set of subtasks that support the active reading experience [HPW+12]. Those subtasks frequently involve annotating, note-taking, searching, highlighting, comparison, and non-sequential navigation [TE11a].

Active Reading combines reading with critical thinking, learning, and decision making, whereas passive reading is less careful and less effortful [SPG+99]. *Active Reading* is therefore supported by various strategies that can take place internally in mind or can be further aided by externalization, i.e., the act of making one's thoughts visible to support cognition [Kir10]. Deciding to focus on specific parts of the text is an example of internal active reading, whereas highlighting parts of a text, making annotations within a text or taking notes about a text are examples of externalization-based active reading [WHP+18].

The concept of *Active Reading* is quite similar to *Visual Thinking*, and both concepts have many overlaps. The main difference between these two concepts is that the main focus of *Active Reading* lies, as the name already suggests, on reading and retaining written content, whereas for *Visual Thinking* there is no such restriction. In *Visual Thinking*, the knowledge gain can come from any source, may it be in written form, via listening or even feeling. As *Visual Thinking* is defined more generally, one might argue that *Active Reading* is a special form or subcategory of *Visual Thinking*. Since most of the work of knowledge workers lies in gathering information and actively reading it, *Active Reading* is a major component of this thesis.

Adler [AD72] defines active reading of a text with a focus on internal mental activities, describing it as "the asking of questions" about a text. Depending on the reader's goals, efforts and skills, questions are asked and answered differently. Adler [AD72] identifies four levels at which a reader's goals can differ and which capture the wide applicability of active reading from basic understanding of text to developing new ideas within entire subject areas. Those levels are the *elemantary*, *inspectional*, *analytical* and the *syntopical* level. In this thesis, we investigate active reading strategies on a syntopical level, which involves reading multiple sources and creating new analyses that might not be present in any individual book [AD72], such as in knowledge work when multiple documents must be cross-referenced [AGH+98]. We further are interested in the externalization strategies that result from those internal mental activities.

2.1.3 Sensemaking

The process of building understanding out of a collection of data is called sensemaking. This process is often complex and ill-defined, involving data that is incomplete, dynamic, and, in some cases, even wrong or deceptive [AEN10]. It ranges from fitting information directly to its need to very complex sensemaking activities that require synthesis and assimilation of information into the users' existing knowledge structure to establish an understanding [ZS16]. Knowledge workers encounter sensemaking tasks every day. Sensemaking is a fundamental human activity, as technology can provide support for activities like searching, filtering and visualizing, but it cannot provide understanding [AEN10].

In general, knowledge workers' sensemaking is completed in two major loops: the information foraging loop and the sensemaking loop [PC05]. In the foraging loop, knowledge workers discover new information and sort and filter this information as necessary. An example of a task that can be situated in this phase is the screening and rough categorization of the raw material and, in the course of this, a first rough brainstorming about relevant contents, categories, and topics. Institutional teaching and personal knowledge developed by the knowledge worker are combined in the sensemaking loop [BPP+19]. Figure 2.2 shows the notational sensemaking model, which describes the cognitive processes knowledge workers undergo when completing tasks in intelligence analysis.

Even though this model is widely used in the visual analytics community and has guided the design of many visual analytics systems, it does not provide rich details

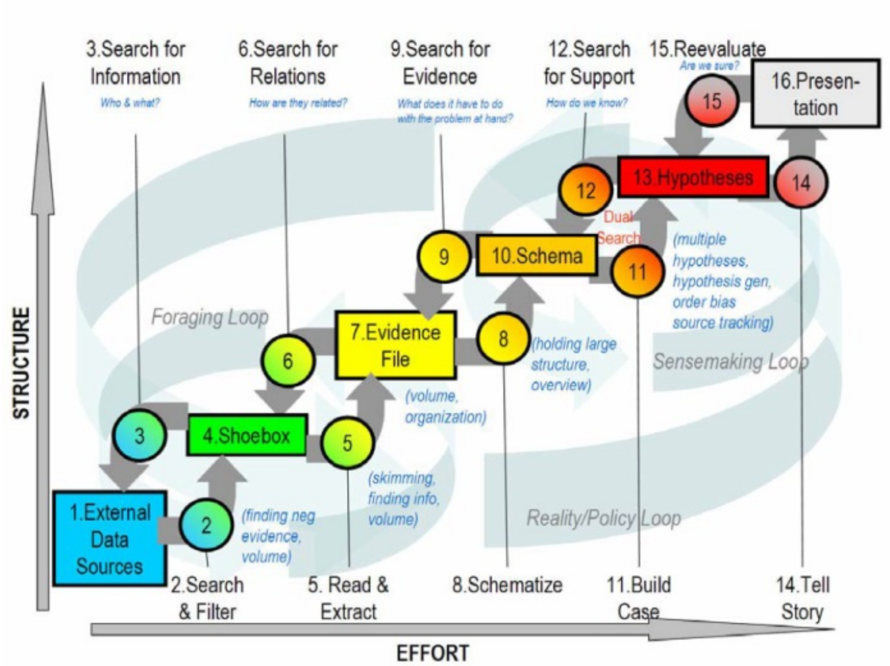


Figure 2.2: Pirolli and Card's sense-making model [PC05].

of how intelligence analysts work in the real world. Rather than working linearly, knowledge workers operate on everything in parallel during almost the entire project [KS11]. Although the model acknowledges either top-down or bottom-up movements or jumps to different stages, it still implies that the stages of the intelligence process are sequential and discrete. For example, an explanation for the reasons for frequent jumps of knowledge workers' from one state to another that are not adjacent in the model is not given by the model. Rather than describing how analysts work and how they transition, the model describes how information transforms and how data flows [KS11].

Nevertheless, the model still provides valuable insights as the information transformations are part of the knowledge workers' externalization of their thoughts. As analysts alter their way of thinking and refine their visual model as they learn more about a specific domain, support for externalization should occur throughout the sensemaking process [KS11]. Moreover, externalizing thoughts has many cognitive and social benefits, such as lowering the working memory load, supporting idea reformulation, which can lead to substantial improvements in understanding and retention, and providing common ground for sharing insights with others to support communication and decision-making [RHRH+19].

2.2 Active Reading in the Analog and Digital World

Depending on the tasks to be performed, people prefer to use either analog or digital tools. The decision which context is the right one for the activity in question depends strongly on individual preference, but above all on the benefits and advantages that the analog or digital environment brings. In the following, the characteristics of typical active reading activities (such as reading, annotating, note-taking and (spatial) organization and navigation) as part of knowledge workers' visual thinking strategies using either analog or digital tools of a computer setup are discussed, as well as their differences and expected effects on the actors' approach. This analysis provides an overview of both worlds' characteristics (which is given in Table 2.1) and serves as a basis for a better understanding of the active reading strategies users employ during the study. Notice that the digital world in the context of this thesis refers only to a computer setting, leaving out other digital solutions such as e-readers and tablets.

Area/Activity	**Analog**	**Digital**
Documents	Tangible Visually distinguishable from each other (when not in focus, e.g. size, thickness, look of first [visible] page)	No tangibility provided Look almost the same (filename; icon depending on document type); do not have visually distinguishable attributes (except for the name and position in the file system)
Reading	Reading angle can be adapted arbitrarily View is restricted to field of vision; content is fixed and restricted to paper size Text has a fixed position on paper Manual and time-consuming search External lighting No power source needed (except for external lighting in dark environments)	Reading angle is limited to the monitors rotating and tilting possibilities View and content is restricted to screen size and resolution; content in most cases adaptable (e.g. font size and layout on the web) Scrolling imposes a spatial instability Automatic keyword search (often) supported Screens emit light and reflect Power source is needed (battery or electricity)
Annotating/ Highlighting	Relatively effortless, quick and smoothly integrated Flexibility in precision (can be quite accurate or rough depending on needs)	More complicated and cumbersome due to inflexibility of interaction techniques Partially inaccurate or even wrongly placed
Writing/Note-taking	Pen allows free-form notes and sketches Exhausting in the long run Strenuous transcription word by word (or use of additional devices like copiers) Manual spell-checking Handwriting sometimes challenging to decipher for others Physical act of writing provides an additional layer of memory that assists performance (better recall and recognition; motor memory effect) [SMR09]	Keyboard good for writing; creation of diagrams requires computer skills and more time Quick and less exhausting (for text input only!) Ease of copy-paste Automatic spell-checking Readable by everyone Additional layer of memory is not found while typing
Spatial Organization/ Navigation	Quick and easy Easily interweaved with reading using two-handed movements Filing and piling supported Arbitrary positioning	Complicated and sometimes time-consuming window management One handed input via computer mouse (or keyboard) Filing usually better supported than piling Positioning possible on different levels: positioning of screens, positioning of windows on the screen and virtual desktops, window management in general (minimization, maximization, window scaling, window grouping e.g., as tabs) and desktop icon positioning

Table 2.1: Characteristics of the analog and digital world in regards to typical active reading activities.

2.2.1 Reading

Reading is a highly practiced activity, not just by knowledge workers but by all kinds of people, as it is an important part of peoples' daily (work) lives. It forms a component of a wide range of different activities and serves many different purposes. The reading processes are shaped by the reasons of why a text is read and the broader context within which the reading activity is embedded [OS97]. People either read analogously on paper such as books, magazines or printed articles, or digitally on their (computer) screens. Paper is a material that is tangible and enables haptic perception. The haptic interactions with paper text afford readers richer sensorimotor engagement with the text compared to screen text, which enhances information encoding and comprehension [HRL17]. Digital documents do not provide such tangibility and materiality. They look almost the same and do not have visually distinguishable attributes such as the document's size (= number of pages and therefore its thickness), the location where it is placed on the desk, or the look of the topmost document on a pile [BJ05]. People cannot hold digital documents in their hands or recognize the document's length by its weight or thickness. They cannot change the reading angle as arbitrarily as possible with the paper since monitors only provide limited rotating and tilting possibilities.

Moreover, computer screens glow, which is not the case with paper. Computer screens also reflect, especially in the sun, making it harder to read outdoor or inside near windows. Although the quality of computer screens has immensely improved in the last decades, reading on paper is still less eye-straining than reading digitally. LCD computer screens are, for example, known to cause visual fatigue due to their emitting light [MWB13]. In contrast, e-book technologies based on electronic ink, such as the Kindle, are merely reflecting light and are hence more reader friendly with respect to the visual ergonomics [MWB13]. When reading paper text, the haptic modality might offload some cognitive demands onto the visual modality, thereby alleviating such visual fatigue. As a result, screen-based reading is more physically tiring and mentally taxing than reading paper [HRL17]. Another interesting difference is that paper documents are mainly in portrait format, while computer screens are, by design, of landscape format. Most digital documents are then again designed in portrait mode (even most websites), resulting in only having small parts of the document visible in full-screen mode (or without any window size and zoom level adaptations). This restricted view leads to much scrolling, primarily since scrolling is also used to assess a document's length.

Compared to turning pages, scrolling increases cognitive and visual demands [GCR+19]. Even though scrolling itself can quickly be executed, it is known to hamper the reading process [MWB13]. This hindrance of the reading process might be due to the lack of hard cuts between pages and the fact, that the pages are usually displayed one below the other, imposing a spatial instability making it much more difficult for the digital reader to create a mental representation of the page and the location of its contents [MWB13]. By contrast, the fixity of a text on paper supports the construction of spatial representation [MWB13], making it easier for analog readers to remember and find read information again. With books, for example, readers often know the exact position of an image: They can tell precisely whether the image was on the left or right page, and whether it appeared at the beginning or end of the page.

For a while now, paper has ceased to be the only medium used for reading. Especially digital natives, those born after 2000, do much of their reading in digital environments [GCR+19], such as on their smartphones, tablets, or computers. With the technological improvements in screen design and reader software enhancements, reading on screens presents a less aggravating and more flexible reading experience to an increasing number of people [J.01], not just the digital natives. For this reason, countless studies deal with reading in digital environments compared to typical paper-based settings and numerous projects investigated the creation of digital (active) reading environments [WHP+18] including XLibris [PSG98], PapierCraft [LGHH08], LiquidText [TE11b], GatherReader [HBPB12] and Matulic and Norrie's pen-and-touch active reading environment [MN12].

The majority of studies focus on performance measures of reading such as speed, proofreading accuracy and comprehension [SA17, KSZ18, Cli19, DVAS18, PAM+16, MWB13]. In general, many of those came to the conclusion that there is a benefit for reading comprehension when reading from paper compared to screens. Reading times do not reliably differ by medium, indicating that reading from paper is more efficient, considering that there is better performance with similar time investment [Cli19]. One study [PAM+16] could observe equivalent performance outcomes for computers and paper. Reasons for different findings in the literature might be because of different study designs and, more importantly, the different texts being read in various studies [GCR+19]. For example, the text was confined to a single page in the study with no reading comprehension differences [PAM+16].

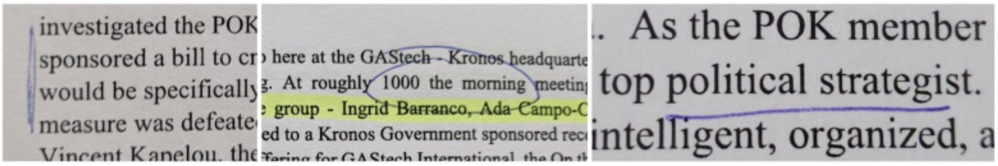
One might think that potential comprehension difficulties in digital reading might disappear once people have enough experience with digital technologies [DVAS18]. However, there are no intergenerational differences anymore [MWB13]. Effect sizes favoring paper-based reading have increased in recent years and the advantage of print reading has significantly increased since 2000, suggesting the digital disadvantage may be growing instead of dropping [GCR+19, KSZ18, Cli19]. These surprising findings suggest that we cannot idly wait for screen inferiority to disappear as children are exposed to digital devices earlier in their lives, as adults gain more experience with technology or as technology improves. Digital reading is a major challenge across age groups that becomes even more severe as the presence of technology increases [DVAS18].

It seems unlikely that the computer will replace the printed book as a reading medium soon in the same way it replaced the typewriter as a writing tool [Zim05], as there is a legitimate concern that reading on paper may be better in terms of performance and efficiency [Cli19]. Therefore, it is essential not just to focus on performance outcomes but also to understand and observe reading behaviors, strategies, and (sensemaking) processes people use in the two different mediums, which is exactly what the study of this thesis does. Screen-based reading behavior is characterized by more time on browsing and scanning, keyword spotting, one-time reading, non-linear reading, and more reading selectively, while less time is spent on in-depth reading and concentrated reading and sustained attention is decreasing [Zim05]. Evidence is provided that people adopt a more shallow processing style in digital environments, which may be related to more mediocre quality of attention. Reported negative correlations between the frequency of digital media use and text comprehension further support the fact that the more people use digital media for shallow interactions, the less they will be able to use them for challenging tasks [DVAS18].

Many studies [Zim05, TSO15, TE11a, HPW+12, OS97, MBM07, GCR+19] observed their subjects while performing a more challenging task and encountered differences in essential reading supporting activities such as annotating, note-taking, intra-document navigation, and spatial organization. While some of these studies only deal with reading per se (e.g., proofreading), the focus of this thesis is on reading for the purpose of writing, as this is one of the most common tasks knowledge workers perform. Due to the different reading purposes, however, the observations might drastically differ as subjects may behave differently. Although some studies focus on reading for the purpose of writing, they often do not take into account the importance of spatial organization throughout the whole sensemaking process. In this thesis' study, a large high-resolution screen is provided to enable almost similar spatial organization as on a desk with paper documents. Besides, the materials used by previous studies are mostly single-page documents [GCR+19, TSO15, PAM+16, MWB13], or, in some cases, there is even only one document to be read [OS97, TSO15, PAM+16, MBM07]. Especially in the digital world, the number of pages (and documents) and the associated movements might influence the way people interact with the documents and work and might unravel further challenges for the user. By using a complex task involving the reading of multiple documents of different lengths, a realistic scenario can be created, and the observations of behaviors and problems correspond more closely to those that would occur in reality.

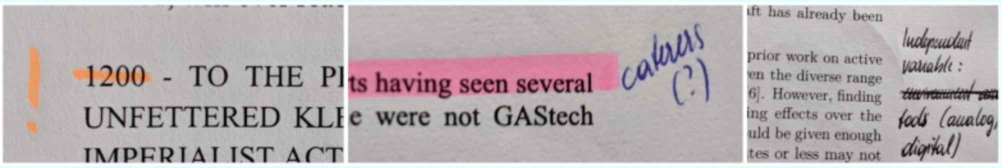
2.2.2 Annotating & Highlighting

People like to annotate the original content when they read, especially for in-depth reading, as it is part of their sensemaking process [Zim05]. Annotations on the content enable people to limit their working memory load and articulate and reformulate thoughts [RHRH+19]. Major functional roles of annotations include procedural signalling for future attention, to place marking, and to aid memory [HPW+12]. They further aid understanding and facilitate the building of an internal representation of the contents [OS97]. Annotations can be divided into three major types: anchor-only, content-only, and compound annotations [MBM07]. *Anchor-only* annotations can be equated with highlighting portions of a text and are the most common use of annotations. Highlighting helps to direct one's attention to important information [PML+18]. Highlighted passages in a document are a form of identification and extraction that isolates information without removing it from the context. For the document as a whole, highlights provide additional visual structure [AEN10]. People can use this additional visual structure of highlighted text portions to associate with text content for reading comprehension as it helps the reader to quickly retrieve critical points during the reviewing process without having to highlights

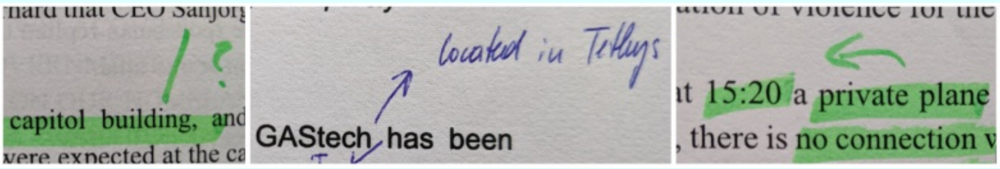


(a) Anchor-only annotations occur in the form of margin bars, circles, highlights or underlines.

annotations



(b) Content-only annotations are notes or marks without any explicit reference to the documents content.



(c) Compound annotations can be complex anchor-only, complex content-only, or a combination of both annotation types.

Figure 2.3: The three types of annotations: anchor-only, content-only, and compound.

skim through the content again [LTC16]. Words or marks added to the page without explicit connection to any portion of the text are denoted as *content-only* annotations. If a line or other connection indicates a relationship to portions of the text, these words or marks are labeled as *compound* annotations [MBM07]. Figure 2.3 shows examples of all three types. In the following, anchor-only annotations are regarded as an independent group, such that to avoid confusion, the term “highlights” is used when talking about anchor-only annotations, while the term “annotations” only refers to content-only and compound annotations.

Findings of prior studies suggest that people find annotation and highlighting more complicated and do it less when reading digitally as compared to paper [MBM07, GCR+19]. On paper, markings can be made quickly and interwoven with the ongoing reading with relatively less effort and smooth integration. In digital environments, the same process is complicated, inflexible, cumbersome, and detracts from the reading task due to the inflexibility of interaction techniques via mouse and keyboard, as the functionality must be activated first before being used [OS97]. Findings show that using annotation tools while reading printed text leads to more accurate and faster responses than using the same tools when reading digital text, and that the use of annotation tools in general significantly increased text reading time [BYEA14]. Furthermore, studies observed subjects struggling to digitally highlight the exact text desired [GCR+19], as typically, the process of digital highlighting requires finding and pressing a button to enter the highlighting mode before selecting text portions vaguely with the mouse cursor [WHP+18]. Making annotations with a pen or highlighting text portions with a marker is physically easier (and more precise) than clicking and highlighting on a computer [GCR+19]. Interestingly, a positive correlation between digital highlighting and comprehension could be found, while paper highlighting correlated negatively to comprehension [GCR+19]. This correlation might be due to the observable tactile challenge of physically highlighting digital text. Subjects used fewer digital annotations and highlights, but they were more strategic and placed in more critical areas of interest. In comparison, subjects used much more paper highlights, of which many occurred outside of relevant information in the text [GCR+19]. Annotating and highlighting electronic documents is certainly possible, but it requires much more resources and additional skills than a simple pencil or highlighter [Zim05]. Suggestions for improving the digital experience from annotations and highlights include the ability of annotations to span page and document boundaries, having properties like priority or category, and being able to aggregate annotations and highlights together and organize them with a level of structure chosen by the user [TE11a].

2.2.3 Note-Taking

Another task people often perform while reading is taking notes. Taking notes is a commonly practiced strategy while performing knowledge-intensive tasks. It is the activity to write or draw short sentences or rough figures to retain information, usually on a blank page [SO18]. It is intended to direct one's attention to relevant information, resulting in encoding it in working memory and storing it in long term memory [PML+18]. From a purely technological perspective, note-taking seems archaic as there exists a wide range of inexpensive and convenient duplicating procedures, making notes quite redundant. There is, however, a guiding hypothesis that note-takers would learn more because they are more actively engaged in manipulating incoming information [Car83]. Note-taking may increase general attentional processes, leading to a higher concentration on the material. It furthermore encourages people to process materials at a more meaningful and more profound level. Thus, the degree of energy or effort devoted to the new material may lead to greater learning [CT79].

Many studies compared note-takers with non-note-takers in lecture settings (listening to a lecture and having tests or quizzes directly afterward [Car83, SO18, MO14]). In Ladas' review [Lad80], he reports that of the eleven studies of this nature he located, eight found that note-taking led to better performance and three did not find significant differences [Lad80, Car83]. The utility of note-taking seems to depend on the presence of several factors, such as the allowance for a review period, the quantity and quality of notes, the type or timing of test situations, and the characteristics of the note-takers themselves

[CT79]. Interestingly, studies found that subjects who expected an immediate test took only half as many notes and performed better on an essay-like test [Wee71]. This finding results in speculations that immediate-expectation subjects actively manipulated the materials during learning while the other group adopted a passive, verbatim recording set, which did not include performing transformations on the materials [CT79].

Research also found that subjects who were allowed to both generate and review their own notes scored higher on immediate free-recall and short term objective tests than those who were not allowed to takes notes and those who were allowed to review notes taken from someone else [FH73]. These findings suggest that the practice of both taking notes and being allowed to review one's notes are optimal study strategies. In this thesis' study, subjects are allowed to review their notes and use them to give an oral summary as well as answer the questions of a post test, as reviewing notes is a typical activity done before writing a summary or report. However, taking notes is not an efficient process, especially while listening, as people can only record modest amounts of information, even at comfortable presentation rates. In the cases of people with poor short term memory abilities, note-taking even interfered with the ability to recall information [Car83]. In an eye-tracking study [PML+18], researchers have shown the utility of examining integrative saccades (= rapid movements of the eye from one fixation point to another) between areas of interest in addition to fixations (= the eye coming to rest on one specific part of the text) on key areas of interest as an index of depth of cognitive processing during learning. Their results show that note-taking requires only lower level-processing and did not encourage the subject to reorganize material mentally; thus, it did not improve memory for the material [PML+18].

Handwriting versus Typing

People are used to taking notes with pen and paper. It is a popular activity that is learned early in school. Analog tools allow free-form notes and enable users to sketch and draw wherever and whatever they want. However, taking lots of notes per hand is exhausting and increases the feeling of hand fatigue and cramps. Typing on a keyboard can be much quicker and less exhausting [MBM07]. Computers are widely used as a writing tool, especially if documents are formal or lengthy and are read by other people [SO18]. Studies have therefore investigated differences in learning effects between writing with conventional pen and typing on a keyboard [OIN+19]. Due to their findings, many researchers have suggested that laptop note-taking is less effective than longhand note-taking for learning [MO14]. One possible explanation for performance differences is that laptops offer access to a massive range of other programs and applications, leading to distractions that can decrease performance [MDR19]. But even when laptops were used solely to take notes, they may still be impairing learning because their use results in shallower processing [MO14].

Most writing tools on computers support the ability to utilize special functionalities such as searching for keywords inside a document, having anchors to different parts of the document, or even referencing completely different documents by hyperlinks. Automatic spell-checking [MBM07], for example, makes it easier to write error-free texts. The ability to utilize special functionalities is especially advantageous for writing formal and long documents [SO18], which then seem much cleaner and can be read by everyone without problems, while handwriting can be, in some cases, quite challenging to decipher. Another functionality is the ease of copy and paste. Most of the time, copied information is kept in its original format. The context around that piece of data is, however, lost, resulting sometimes in the user needing to go back to the source to remind themselves about the context within which the extracted fragment appeared [ZS16]. Although writing on the computer is comfortable and in most cases faster than writing by hand, most people start taking notes on the computer only after reading the document and generating the notes entirely from their memory, while analogous note-taking is more integrated into the flow of reading [OS97]. People can take notes or annotations with one hand while skipping and skimming through documents with the other. The tangibility and materiality of paper often outweighs the benefits afforded by computers [HPW+12].

Laptop note takers are more likely to take lengthier transcription-likes notes with more considerable verbatim overlap rather than processing and reframing information in their own words because most people can type significantly faster than they can write [MO14]. However, verbatim note-taking seems to be detrimental to learning. In general, taking more notes, thereby having more information, is beneficial; mindless transcription seems to offset the benefit of increased content, at least when there is no opportunity to review [MO14, Kob05]. Although differences in word count and verbatim overlap between note-taking with traditional pen or keyboard were large, differences in performance were small and not statistically significant in some cases and do not appear sufficient to produce performance differences between longhand and laptop note-takers. Mixed results run counter to making any general claims about which note-taking method is superior [MDR19]. However, the ability to generate text quickly now appears to be at the risk of learning what we have written down [SMR09]. Even people themselves think that when they are using handwriting, they can write something easily, start writing quickly, take notes while listening, and, most importantly, memorize words they have written down [SO18].

Experiments showed that cognitive load (= load for working memory when performing a task) of typing was larger than of handwriting regardless of the typing skills. In other words, handwriting did not interfere with thinking or memorizing, which is one of the most important reasons why people prefer handwriting [SO18]. They think, that they can remember handwritten passages easier than typed ones. Handwriting seems to be more effective for conceptual comprehension than typing [OIN+19]. A comparison of recall and recognition for common words demonstrates that memory is better for words when they have been written down rather than when they are typed [SMR09]. The physical act of writing provides an additional layer of memory that assists performance that is not found in typing. Recent imaging and memory performance studies support the hypothesis that due to the additional context provided by handwriting, people remember target words more accurately when they take the time and effort to write them out than to type them [SMR09]. An advantage of handwriting over typing has also been indicated in neuroscientific approaches using electroencephalography and magnetic resonance imaging [LBG+08, OIN+19]. Interestingly, recent studies have shown that the movements of handwriting with a digital pen on a tablet are not the same as with a conventional pen on paper. For example, the movements differ by pen pressure, speed but also the length of the pauses during writing, suggesting that segment trajectory calculation and control of muscular adjustment may be disturbed. From the points of view of movements and brain activities, handwriting with a digital pen on a tablet might, therefore, disturb cognitive activities, such as learning [OIN+19].

Mind Maps & Concept Maps

Millions of people are using mind maps for brainstorming, note-taking, document drafting, and many other tasks that require hierarchical structuring of information [BL11]. They are great personal learning tools that result in attractive, colorful, and memorable solutions [Epp06]. Several studies have highlighted the beneficial use of mind mapping for note-taking, as they are easy to learn and apply, provide a concise hierarchical overview, and are easy to extend and add further content [JPM99, SBW02, FHH02]. However, they are also idiosyncratic, hard to read for others, represent mostly hierarchic relationships, and can become overly complex such that the bigger picture is lost [Epp06]. As they are not the best way of sharing information and ideas, mind maps would be best for personal note-taking [Epp06], and it seems that they are used for rather short term activities.

Besides mind mapping, concept mapping is another available qualitative visualization technique that fosters learning and knowledge sharing in a constructive and systematic manner [Epp06]. A concept map is a top-down diagram showing the relationships between concepts, including cross-connections among concepts, and their manifestations [Epp06]. Concepts are usually organized hierarchically, from most general, most inclusive to most specific [NC07]. Relationships between concepts are indicated by a connecting line linking the two concepts [NC07]. Those links may be labelled or unlabelled, directional or non-directional [NA06]. An example of a concept map can be seen in Figure 2.4.

Concept mapping is often used as media for constructive learning activities and as communication aids in lectures, study materials, and collaborative learning [NA06]. It is particularly suitable for acquiring main ideas, but inadequate for acquiring detailed, nuance-laden knowledge [NA06]. When the process is done well, concept mapping is an easy way to encourage very high cognitive performance levels. This is also one of the reasons why concept mapping can be a very powerful evaluation tool [NC07]. However, it is not a simple, seamless, or very rapid visual externalization technique, as there are relatively strict formal rules that need to be adhered to when drawing a concept map [Epp06]. Moreover, the top-down structure may not be adequate to represent or structure sequential content such as processes, timelines, or developments. Concept maps also tend to be less memorable, because most of them look very much alike [Epp06]. However, there is evidence that concept mapping is slightly more effective than other constructive activities such as writing summaries and outlines [NA06].

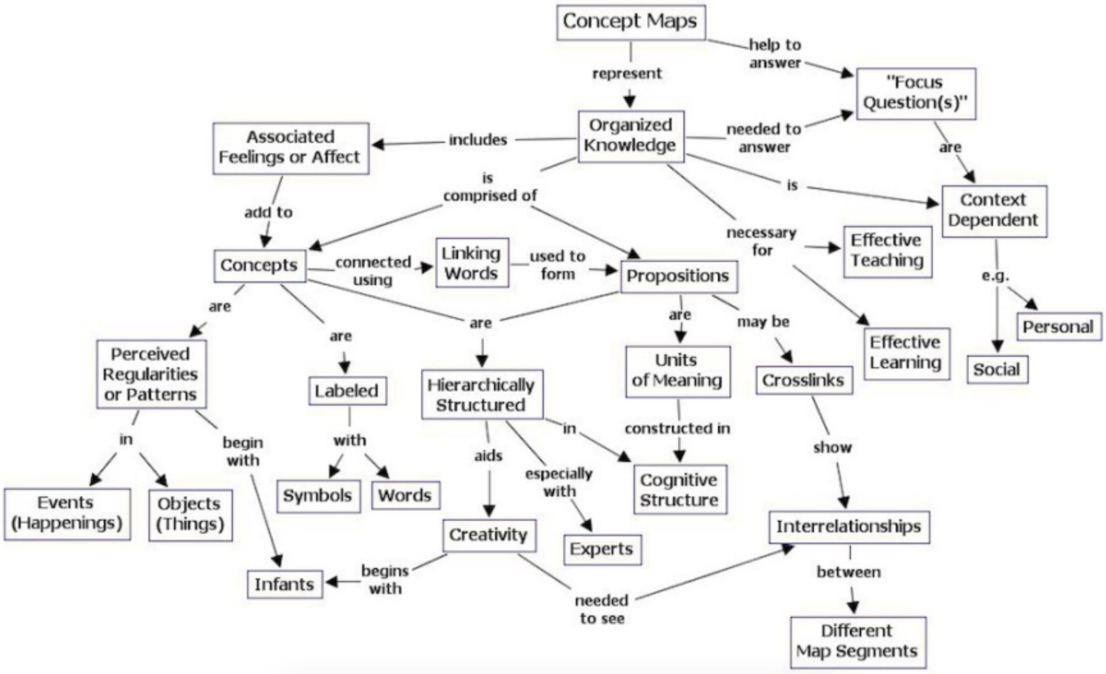


Figure 2.4: A concept map showing key features of concept maps [NC07].

In general, the construction of a concept map is intended to reveal the map's author's perceptions, rather than reproducing memorized facts. Therefore, concept maps can be seen as a portrayal of a mental model, which is helping in organizing knowledge and, as such, in understanding [KHA00]. The best way to construct concept maps is by having a reference to some particular question one seeks to answer [NC07]. Often, concept maps are created when a general understanding is established, such that the facts are already stored in notes during the information gathering stage [ZS16]. The created concept maps can have different levels of detail: Some people only create a few nodes (and even without connections), others create detailed graphs with cycles (and not necessarily a graph tree) [WGSS21]. When transforming from notes to concept maps, sensemakers sometimes make implicit concepts and relationships explicit by creating a node or link in the concept map, which are, in many cases, hidden in the text format [ZS16]. Good concept maps usually result from three to many revisions [NC07] and would be best for self-study and review purposes, because they take longer to develop [Epp06].

The construction of knowledge representations during sensemaking resembles meaningful learning. Concept- and mind-mapping tools can be used to build structures or make implicit structures explicit [ZS16]. There are many software solutions for creating mind or concept maps on computers, like Docear [BGLG11], MindMeister [Mei20b] and CmapTools [CHC+04] as a few examples of such applications. Such software tools enable users to create such maps easily in a digital environment. Figure 2.5 shows a digital mind map, which was created with the help of MindMeister. Mind- and concept-mapping

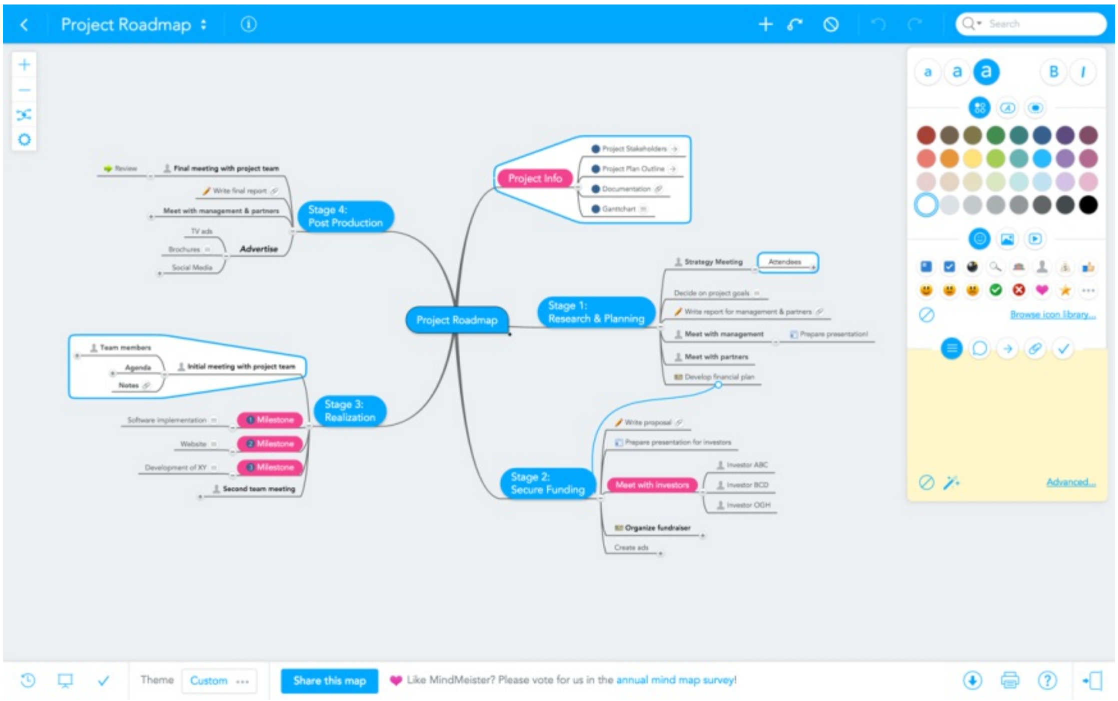


Figure 2.5: A mind map created with MindMeister [Mei20b].

applications further support additional features not possible on paper, such as adding hyperlinks and references as well as additional notes to the mind maps' nodes, though not many users seem to use this kind of feature [BL11]. Differences between software-based use and analog application using pen and paper have not yet been analyzed [Epp06].

2.2.4 Spatial Organization and Navigation

Humans are cognitively well adapted to making use of space to express and perceive relationships between objects [Kir95]. On a physical desk, people typically organize their documents and pieces of information into files and piles. Files are units, where elements are explicitly titled and arranged in some systematic order. In piles, on the other hand, individual elements are not necessarily titled and, in general, are not arranged in any particular order. Because piles do not have a systematic order, their spatial location is often especially important in finding them [Mal83]. Furthermore, information that is visible on the top of the desks is in many cases a reminder, that there are still things to do without having the people intentionally to look for what needs to be done [Mal83].

Digital systems usually support filing better than piling: People have to categorize a document to store it, and by being placed in the system, a document already belongs to a certain category. Knowledge workers, in particular, experience difficulties in categorizing documents [BJ05]. The difficulty of deciding how to classify something can be an important barrier to filing information. Deciding what categories there are and what category something is in is organizationally the hardest problem [Mal83]. It was noticed that the contemporary hierarchical file folder structure does not support natural ways of organizing, as it does not allow one to put a document aside to work on it later [BJ05]. People spend many of their precious daily working hours looking for lost information, that they know exists somewhere but are failing to find [Jon07]. Digital documents do not contain meaningful context information and are not well distinguishable from each other. They look very similar to each other, and the content of the documents is entirely hidden from the user's view. Furthermore, file folder structures can go almost endlessly deep in the hierarchy, while the structure is not transparent except for a tree-like hierarchic overview. Even though the virtual desktop allows easy spatial regrouping, it is not as extensively used for managing important documents than traditional desks with paper [BJ05]. Paper documents give almost infinite flexibility in (re-)structuring documents without an (explicit) categorization effort, and the unique affordances of paper emphasize the convenience of it for collaborative working and sharing. The tangibility of paper makes regrouping of documents according to changes in task planning almost effortlessly. Regrouping is, in general, very easy with paper documents, as paper piles can easily be removed or extended by inserting new paper documents [BJ05]. Even though digital documents can also be easily duplicated, shared, extended and removed, they need some categorization when being put in the file system, since piles without any structure are usually not supported on virtual desktops.

In most people's lives, the use of spatial organization to make sense of and recall information, as well as reveal relationships, is practically a daily occurrence [AN12, AEN10]. Examples of such usage might be notes attached to the side of a monitor or papers spread out on the desktop [AEN10]. Figure 2.6 shows a desk with an exemplary organization of documents. Multiple studies explicitly compared virtual and physical workspaces for sensemaking. The results show that the physical space provided by a large display biases the user towards working spatially, leading to increased externalization of the user's synthesis [AEN10]. In contrast, most multi-monitor configurations encourage the user to think in terms of separate workspaces, usually associated with a distinct application or task [AN12]. Displays that are not just physically large, but high-resolution as well, allow users to place detailed views of documents into spatially meaningful representations and, therefore, enable them to increase their simultaneous access to information [BEK+13, AEN10], as can be seen in Figure 2.7. Furthermore, these meaningful spatial representations enabled by large high-resolution displays can then be used to recall information through physical navigation easily as well as to organize the display space semantically. These properties have been shown to improve user performance on many tasks ranging in difficulty from simple pattern finding to cognitively demanding sensemaking [BEK+13]. This performance improvement might be because physical navigation is fundamentally more embodied than virtual navigation, which requires technological mediation and internal mappings to maintain spatial understanding, whereas physical navigation leads to the development of more effective externalizations [AN13].

The impoverished environment of single monitors forces users to make explicit context

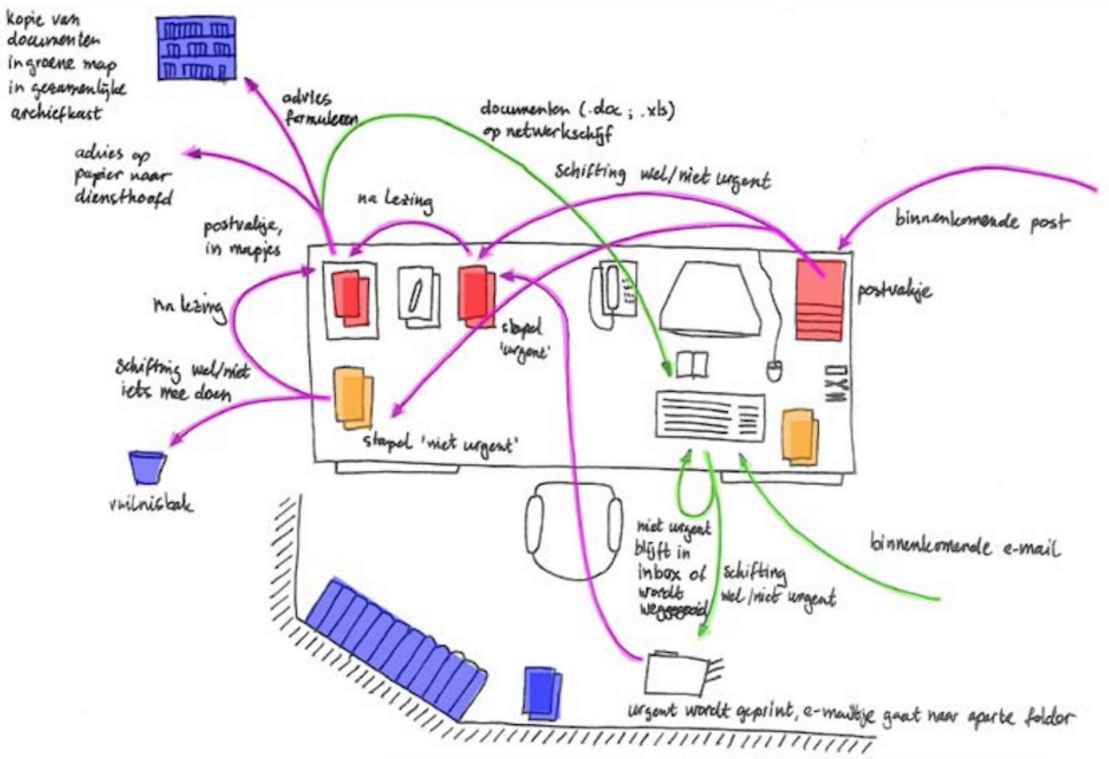


Figure 2.6: Example of the organization of documents on a desk [BJ05].

switches on the introduction of new information, frequently in the form of a new window overlaying the previous one. Such context switches severely affect the user's ability to make comparisons and require the user to expend valuable mental resources on the minutiae of managing views rather than on the problem at hand [AEN10]. Furthermore, on conventional displays, documents can either be arranged spatially or viewed at a detailed level, but not both. On large high-resolution displays, even the detailed view of a document is relatively small in relation to the available space, thus allowing documents to be placed in spatially meaningful ways while retaining their detailed representations [AEN10]. Such a low interaction overhead for expressing relationships encourages externalization [AN12]. The primary difference between these two types of displays lies in the approaches to document management: On conventional displays (with relatively small resolution), users mainly maximize documents and use the taskbar to switch rapidly through open documents. Even without any special tools, such as Jigsaw [SGL08], AbiWord [Com], Analyst's notebook [IBM], Sentinel Visualizer [FMS], Palantir Government [Tec], Entity Workspace [BIC06], Sandbox [WSP+06], or Cambiera [IF09], the inherently spatial environment of large (high-resolution) displays already provides support for activities typically done with physical artefacts [AEN10]. The available space is used primarily when there are otherwise only rudimentary options for active reading and visual thinking [WGSS21].

Figure 2.7: A 32 megapixel large, high-resolution display running Analyst's Workspace [AN12].

In general, paper allows for quick navigation, both within a document and between several documents. People can easily interweave navigation with reading using two-handed movements [MBM07], allowing them to interleave and overlap activities effectively [OS97]. On the contrary, navigation on computers can be irritating, slow, and distracting. The one-handed input via the computer mouse only allows for the serial performance of activities and often draws the user's attention away from its original task, namely the reading [OS97]. On the computer, it is much easier to lose the point at which one stopped reading because it is not possible to mark the spot with the finger of the second hand like on paper while doing another activity (e.g., writing) in the meantime.

Not only the possibility to work with both hands in parallel is an advantage of paper, but also the flexibility of spatial layout. Documents can be dynamically arranged throughout the whole workspace, while there is a restricted field of view on computers [OS97]. The window management is quite complicated in most cases, consuming much time for managing the position and size of windows [MBM07].

2.3 Related Studies

This section summarizes the most similar prior studies of this work in the area of active reading and its related subdomains (highlighting, annotation, note-taking, and spatial organization). Comparisons are made between these studies and ours to highlight and address identified research gaps of previous studies and, in the process, making our research contributions more clearly.

In 1997, O'Hara and Sellen [OS97] conducted a laboratory study that compared reading from paper to reading on computers. Ten years later, Morris et al. [MBM07] revisited the issues of reading by comparing paper use to using a dual-monitor desktop system, a pen-enabled horizontal display surface, and multiple tablet computers. Like in this thesis, both studies analyzed active reading behaviors in detail. However, both studies were only concerned with the summary of a single document. This thesis, on the other hand, deals with the active reading behaviors of people and their interactions with several documents, which makes spatial organization more difficult but also more useful. Besides, both studies were conducted many years ago. In the meantime, a lot has happened and changed in computerized reading technology, especially regarding computer screens. The underlying technology fundamentally changed: A move from cathode ray tube screens to retina displays allows the manufacturing of way flatter screens. These are now available in many different sizes and, most importantly, have a much higher resolution.

Tashman and Edwards [TE11a] used diaries, interviews, and workshops in their study to let the participants tell them about their real-world active reading frustrations and which kinds of new functionality they would want in an ideal active reading system. Participants did not report anything about their spatial organization of documents. Furthermore, the authors did not perform a controlled user study in which differences between the two modalities (analog and digital) could have been seen. Hong et al. [HPW+12] studied students at their desks during standard paper reading and reviewing activities to understand the context of active reading better. However, they had not make a direct comparison to active reading behaviors in digital environments.

Goodwin et al. [GCR+19] also analyzed the differences between paper and digital reading processes and the links to comprehension. However, they only focused on reading, highlighting, and annotating without considering note-taking, intra- and inter-document navigation and spatial organization. This is due to the fact, that their task consisted of reading a single paragraph only. Takano et al. [TSO15] compared the reading performances and processes on cross-reference reading for multiple documents between paper and computer displays. Their reading goal was to proofread, which is why their main focus was on the movements of documents and pointing gestures of participants. In contrast to proofreading, reading with the goal to write or summarize results in different active reading behaviors. Mangen et al. [MWB13] and Sun et al. [SSH13] studied the comprehension differences between print and screen reading without further analyzing the reading processes and behaviors.

Carrier and Titus [CT79], as well as Jansen et al. [JLI17] analyzed in their review the cognitive costs and benefits of note-taking in lecture settings. Ponce et al. [PML+18] examined in their study the effects of using one or two computer-supported learning strategies on learning processes and learning outcomes. Mueller and Oppenheimer [MO14], Shibata and Omura [SO18], as well as Morehead et al. [MDR19] compared the effects and the cognitive load of handwriting and typing. Our research compliments their results by analyzing and comparing note-taking in analog and digital active reading scenarios.

Nesbit and Adesope [NA06] conducted a meta-analysis that reviews learning with knowledge and concept maps. The meta-analysis found that in comparison with activities such as reading text passages and attending lectures, concept mapping activities are more effective for attaining knowledge retention and transfer. This effect might be due to the greater learner engagement occasioned by concept mapping compared to reading and listening. There is also evidence that concept mapping is slightly more effective than other constructive activities such as writing summaries and outlines [NA06]. In our study, users could create concept maps, but they were not required or encouraged to do so.

Mander et al. [MSW92] investigated in the early 90ies how people dealt with the flow of information in their physical workspaces. They found out that subjects created piles of documents in an attempt to quickly and informally manage their information. As a result, they developed and prototyped a new desktop interface element: “The pile” [MSW92]. In 2006, Agarawala and Balakrishnan [AB06b] evaluated interaction and visualization techniques that explore the use of piles as the primary organizational entity for desktop objects as part of a detailed exploration of the piling metaphor. One year later, in 2007, Apple first released in macOS X the now renamed “Stacks” [Appa]. Since then, Apple has adapted “Stacks” further. Apple users can now set up sophisticated smart folders on the desktop, that look like a folder but do not exactly behave like one [Appb]. We will discuss these approaches with respect to our results and suggest future ideas and improvements.

Hutchings et al. [HSM+04] conducted a study of the window management practices of both single and multiple monitor users, in which the computing event activities of participants were logged over a time period of three weeks. Their results show that more monitors, and therefore more display space, results in more visible windows. Interestingly, users tend to have some part of the screen empty with more screen space available [HSM+04]. Bi and Balakrishnan [BB09] conducted a diary study that investigates users’ behaviors when switching from standard computing environments (i.e., single- or dual-monitor) to using a large high-resolution display for five days. They found out that more display space resulted in a higher effort in managing windows and optimizing the window layout to improve participants’ workflow. The results also reveal that users on a large display perform more window moving and resizing, but less minimizing and maximizing operations than a single- or dual-monitor, and that a large display could benefit multi-window tasks [BB09].

Waldner et al. [WGSS11] investigated in their study how users place windows in large irregular environments. Their results show that users established unconventional window management strategies to cope with the display’s size and irregularities. They observed a clear tendency of the participants to divide the display into a close focus area and a distant context area [WGSS11]. Andrews et al. [AEN10] studied the spatial organization and navigation of participants during sensemaking on large, high-resolution displays. A few years later, Andrews and North [AN13] compared the sensemaking processes of participants using small and large, high-resolution displays, focusing again mainly on spatial organization and navigation. Our study compares digital spatial organization of participants on a large display to analog spatial organization of participants working with paper on a desk.

Summarized, many of the previously conducted studies were either limited to specific parts of active reading and sensemaking processes or did not make a direct comparison between analog and digital work environments. We investigate the interaction of all aspects (reading, highlighting, annotating, writing, spatially organizing and navigating) and provide a direct comparison between the active reading behaviors in analog and digital environments. In particular, the (spatial) organization and navigation for more complex tasks involving multiple documents have not yet been well researched between analog and digital setups. Table 2.2 gives an overview of the studies as mentioned earlier, and their research focus in relation to active reading behaviors. As can be seen, only three studies [OS97, MBM07, TSO15] cover all aspects and provide a comparison between analog and digital environments. However, the first two have not dealt with the active reading of multiple documents and spatial organization on large displays, while the last one focused on a different reading goal and was mainly interested in actions concerning document moving and placement, pointing to text and moving between pages.

Author/s (year)	Reading	Annotating/Highlighting	Writing/Note-taking	Spatial Organization	Navigation
Agarawala & Balakrishnan (2006) [AB06b]				D	
Andrews et al. (2010) [AEN10]	D	D	D	D	D
Andrews & North (2013) [AN13]	D	D	D	D	D
Bi & Balakrishnan (2009) [BB09]				D	D
Carrier and Titus (1997) [CT79]		A	A		
Goodwin et al. (2019) [GCR+19]	**AD**	**AD**			
Hong et al.(2012) [HPW+12]	A	A	A	A	A
Hutchings et al. (2004) [HSM+04]				D	D
Jansen et al. (2017) [JLI17]			**AD**		
Mander et al. (1992) [MSW92]				A	
Mangen et al. (2012) [MWB13]	**AD**				
Morehead et al. (2019) [MDR19]			**AD**		
Morris et al. (2007) [MBM07]	**AD**	**AD**	**AD**	**(AD)**	**AD**
Mueller & Oppenheimer (2014) [MO14]			**AD**		
Nesbit & Adesope (2006) [NA06]	**AD**		**AD**		
O'Hara & Sellen (1997) [OS97]	**AD**	**AD**	**AD**	**(AD)**	**AD**
Ponce et al. (2018) [PML+18]	D	D	D		
Shibata & Omura (2018) [SO18]			**AD**		
Sun et al. (2013) [SSH13]	**AD**				
Takano et al. (2015) [TSO15]	**AD**	**AD**	**(AD)**	**AD**	**AD**
Tashman & Edwards (2011) [TE11a]	A	A	A		A
Waldner et al. (2011) [WGSS11]				D	

Table 2.2: Summary of studies (sorted by author) and their research area focus in relation to active reading behaviours in analog conditions (A), digital environments (D) or a combination/comparison of both (AD).

CHAPTER 3

User Study

To improve the digital sensemaking process for knowledge workers, a better understanding of how they employ different active reading behaviors depending on the available tools and materials is necessary. In particular, it is necessary to understand which activities users employ during their active reading processes and how they differ between the analog and the digital world. Ultimately, the study focuses on finding out what makes active reading in the analog world more effective than in the digital world.

3.1 Hypotheses

This thesis' most central research question is whether analog multi-document active reading is also more efficient than digital multi-document active reading using a large display (Chapter 1 - Q1.). Prior work found much evidence that this is the case for "simple" active reading. However, they did not compare analog active reading to digital active reading using a large display and while interacting with multiple documents.

Based on prior work (see Section 2.2), a list of hypotheses related to corresponding expectations about analog and digital multi-document active reading has been derived. Those hypotheses focus on (partial) aspects of active reading, such as annotating, highlighting, note-taking, and spatial organization, but also on their synergy and their impacts on the users' active reading processes and results.

H1. *Analog multi-document active reading leads to better performance with similar time investment than digital multi-document active reading using a large display.*

Although it would be desirable that digital multi-document active reading using a large display is more efficient than analog multi-document active reading (or at least equally efficient), it is rather unlikely, that this is actually the case. The haptic interactions with paper text affords readers richer sensorimotor engagement with the text, which enhances information encoding and comprehension [HRL17]. In addition to that, screen-based reading is found to be more physically tiring and mentally taxing than reading paper because of the emitted light of computer screens which causes visual fatigue [HRL17]. In general, there is a lot of evidence that there is a benefit for reading comprehension when reading from paper compared to screens, showing that paper reading is more effective, considering that there is better performance with similar time investment [Cli19]. Not only is reading itself more efficient in the analog world, but also taking notes by hand compared to typing [OIN+19]. The tangibility and materiality of paper still often outweighs the benefits afforded by computers [HPW+12].

Since the employed (multi-document) active reading behaviors (Chapter 1 - Q2.) as well as their differences in the analog and digital worlds (Chapter 1 - Q3.) are also of interest, results about specific active reading subtasks are transferred to corresponding expectations in multi-document active reading environments, too:

H2. *Annotations and highlighting are used less during digital multi-document active reading using a large display than while analog multi-document active reading.*

Making annotations with a pen or highlighting text portions with a marker is physically easier and more precise than clicking and highlighting on a computer [GCR+19]. Studies observed subjects struggling to digitally highlight the exact text desired [GCR+19]. As a result, people find annotating and highlighting more complicated and do it less when reading digitally as compared to paper [MBM07, GCR+19]. It is unlikely that this behaviour changes in a multi-document active reading setting.

H3. *People take more notes while digital multi-document active reading than while analog multi-document active reading.*

Even though people prefer handwriting, since it does not interfere with thinking or memorizing [SO18], it is likely that they generate more notes in a digital environment. Taking a lot of notes per hand is exhausting and increases the feeling of hand fatigue and cramps. Typing on a keyboard can be much quicker and less exhausting [MBM07].

H4. *During digital multi-document active reading using a large display, the available workspace is used and organized similarly to the one during analog multi-document active reading.*

It is known that humans are cognitively well adapted to making use of space to express and perceive relationships between objects [Kir95]. Recalling information through physical navigation and organizing (display) space semantically have been shown to improve user performance on many tasks ranging in difficulty from simple pattern finding to cognitively demanding sensemaking [BEK+13]. Furthermore, the inherently spatial environment of large (high-resolution) displays already provides support for activities typically done with physical artefacts [AEN10], leading to the expectation of a similar organization and behaviour in both environments.

H5. *Spatial organization consumes more time during digital multi-document active reading than during analog multi-document active reading.*

On the one hand, the window management on computers is quite complicated in most cases, consuming much time for managing the position and size of windows [MBM07], while in the analog world, people can easily interweave navigation (and organization) with reading using two-handed movements [MBM07]. On the other hand, more display space has also been found to result in a higher effort in managing windows and optimizing the window layout to improve participant's workflow [BB09].

Moreover, we are interested in peoples' adjustments on (multi-document) active reading strategies based on the availability and accessibility of operations and tools (Chapter 1 - Q4.) and expect the following hypotheses due to prior research on active reading:

H6. *Peoples' approaches during analog multi-document active reading differ from the ones used during digital multi-document active reading.*

Since active reading subtasks, such as highlighting and taking notes, are supported and used differently in analog and digital environments, it can be assumed that peoples' approaches will differ accordingly. For example, it is expected that in analog settings, text portions will be highlighted while reading and (few) notes will be taken at the same time, since both activities are more integrated into the flow of reading in the analog world than in the digital one [OS97]. In digital settings, on the other hand, it can be expected that mainly notes are made, although not necessarily during reading but rather afterwards. The one-handed input via the computer mouse only allows for serial performance of activities and often draws the user's attention away from its original task, namely the reading [OS97].

3.2 Study Design

The study can be characterized as a controlled, qualitative study, that combines a confirmatory approach for answering the defined hypotheses with an exploratory analysis to also inspect areas not covered by the hypotheses, and to find possible explanations and reasons for specific outcomes. Based on the qualitative observations, it is possible to characterize differences between (active reading) behaviors of users in the two environments (analog and digital) and to generate further hypotheses for more detailed future investigations.

The study consists of two different conditions: the analog and the digital condition. As for the digital condition, a standard operating system with a large display to facilitate spatial organization has been chosen.

The study uses a between-subjects design. Prior work on active reading indicated that individual variability would be high, given the diverse range of strategies for sensemaking and behaviors for externalization [RHRH+19]. Individual variability can be problematic for a between-subjects design because it might lead to Type II errors, resulting in the acceptance (non-rejection) of a false null hypothesis, even though this null hypothesis is not applicable to the entire population. By increasing the sample size, the power of the test could be increased. As a result, the risk of committing a Type II error would decrease. However, since this is a qualitative study with limited resources, it is not possible to choose the sample size so that a high enough test power is achieved to avoid or reduce the occurrences of such errors. Therefore, not only p-values but also standardized effect sizes and graphical representations of the observed effects are used. When considering a within-subjects design, finding two similar tasks in complexity as well as preventing learning effects over the different conditions is hard. Furthermore, participants should be given enough time to employ their active reading behaviors. Twenty minutes or less may not be sufficient and may put unnecessary pressure on the participants. In a within-subjects design, however, each condition's duration has to be much more limited to enable completion of all conditions within a reasonable time frame. By choosing a between-subjects design, participants were given more time to unfold their active reading strategies to a higher degree.

3.2.1 Independent Variable

The independent variable of this study is the environment (or condition), in which users participate and work. The environment was either analog or digital. In both cases, users were provided with tools of this kind of environment only. In the analog condition, for example, typical tools such as pen and paper were provided and all materials were printed on paper. For the digital condition, a large display was connected to a computer with standard software tools and applications. Section 3.5 describes both environments in more detail.

3.2.2 Dependent Variables

Since we want to study a variety of effects depending on the used environment, we need to measure multiple dependent variables. As a basis for the definition of our dependent variables as well as the coding scheme (see Subsection 3.8.4), we used typical active reading activities (annotating, highlighting, note-taking) and aspects of spatial organization, on which the focus of this entire thesis already lies. Table 3.1 lists all dependent variables with short explanations.

3.2.3 Confounding & Random Factors

As already discussed in Subsection 2.2, there are many differences between active reading in the analog and the digital world. Table 2.1 gives a detailed comparison of both worlds. Therefore, it is hardly surprising that the differences result in a number of confounding

Variable	Explanation
Task Performance	
Summary	Duration of the oral report given by the participants about their task results
(Total) Response time	Response time of participants to each of the six questions about the task materials and in total
(Total) Points	Points for each question and in total
Points per Second	Score, that measures the achieved points in relation to the total response time
Annotating & Highlighting	
Highlights	Number of anchor-only annotations
Content-only	Number of content-only annotations
Compound	Number of compound annotations
Colors	Number of different colors used for highlighting and annotations
Mark thickness	Number of different thickness styles of highlights
Note-Taking	
Type of Notes	Categorizes the notes (e.g., summary, timeline, concept map, ...)
Formats	Number of text formatting used (such as bold or bigger font size)
References	Number of references made to the original data/documents
Connection Lines	Number of lines and arrows (as annotations or in notes)
List Items	Number of list items in the notes
Timeline Items	Number of timeline items in the notes
Paragraphs	Number of paragraphs in which notes are structured
Words	Number of written words
Tilted elements	Number of elements, that are written down italic and/or rotated
Shapes	Number of drawn graphical elements other than connection lines
Copy/Paste*	yes, in case copy-paste operations have been used; otherwise no
Spatial Organization & Navigation	
Average visible documents	Number of visible documents on average
Average open documents	Number of open documents on average
Visible/Open Ratio	Ratio of the average visible and open documents
Groups	Number of groups/piles at the end of the task
Total distinct open documents	Number of distinct documents opened (to verify whether all documents have been read)
Search*	yes, in case the search functionality has been (intentionally) used; otherwise no
Average Display Space Usage*	Percentage, of how much display space participants used on average
Behaviour Analysis (see Table 3.7 for details about the specified behaviours)	
Occurrence	Number of how often the corresponding behaviour occured
Duration	Number of how long a specific behaviour lasted in total
Mean Duration	Mean duration of a specific behaviour
Approach	Categorizes the approach/strategy employed by participants (e.g., reading only, highlighting only, a combination of highlighting and writing, ...)

Table 3.1: Dependent variables of this study (* = Measurement for the digital environment only).

and random factors for this study. These factors must be identified and, if possible, avoided.

The most significant confounding factor for this study is the possibility of a keyword search in documents in the digital condition. The search is a massive advantage over the analog condition and makes the digital one way stronger. As this study is only interested in information management strategies rather than focusing on users' performance, the search functionality could be easily disabled. However, in the digital world, users usually organize their information with the knowledge in the back of their minds that they can search in documents. By disabling this feature, people would have to work with a very unnatural version of information management. In order to ensure the same conditions for both environments, users of the digital condition were not allowed to use any search functionalities during their oral summary or when answering questions about the task. While solving the task and working through the provided materials, they were allowed to use all functionalities and advantages of the digital world as they are used to in their daily lives. It was assumed that the search functionality would hardly be used, as the participants were not asked any concrete questions initially, which could have been used as a starting point for a search. Only towards the end, when names were read more often, or connections were unclear, a targeted search would be beneficial. It turned out that the assumption of low usage of participants' search in the digital condition was right: Only one of a total of nine in the digital condition actually used this functionality.

Another significant difference is that typing on a computer is much more efficient than taking notes per hand, at least as long as the input consists of simple notes only and without diagrams or the like. Taking notes per hand usually takes longer and is, in the long run, more exhausting. That is also why participants had to give an oral report at the end of the study's task instead of handing in a detailed written summary. In a realistic setting, one would assume that the final summary would have been written on a computer, based on the notes created earlier. However, since giving an oral summary is faster than writing one, and neither of both conditions is disadvantaged, the study used this type of summary.

Furthermore, the operating system itself as well as the interaction techniques (including the window management) provided by the system may be factors over which we have no influence. They, for example, influence the way of people highlighting digitally or positioning/structuring documents. In our case, Mac users may have had a small advantage, as they are used to the operating system (macOS X) used in this study. However, this advantage is offset by the use of a Windows keyboard, as this way keyboard shortcuts are more unfamiliar to execute for them.

3.3 Participants

In the end, a total of 17 participants (eight females, nine males) volunteered and completed the study. Participants were recruited by word of mouth and did not need to fit any special requirements. However, they should be knowledge workers of some kind, such as (former) students, researchers, engineers, or journalists.

Eight of the volunteers were part of the analog condition, while the remaining nine belonged to the digital one. The reason for the supernumerary participant in the digital condition is that one of the former participants of this group seemed to be a possible outlier in the beginning. His behavior and procedure did not correspond to what we initially expected in the study, as he made almost no notes, marked nothing, and also scored rather poorly on the subsequent questions. In order to ensure that there would be enough participants per group even after a possible exclusion, another person was recruited for the digital condition. In the course of the study, however, it turned out that the mentioned participant's behavior does not appear to be an unusual one in the digital environment and that he is by no means an outlier. Therefore, none of the participants were excluded, resulting in an odd number of participants for the analysis.

Participants ranged in age in 15 out of 17 cases from 25 to 34 years. In the remaining two cases, they belonged to the age group of 18 to 24-year-olds and 35 to 44. Fourteen participants have at least a Bachelor's degree (ten with a Bachelor's degree, four with a Master's degree). The remaining three have not yet completed any studies and have, therefore, a regular High School degree or equivalent. More than half of the participants (11 out of 17) are currently mainly working, while the rest is still focusing on their studies. Users had different educational and professional backgrounds, such as Computer Science (eight out of 17), Marketing & Communication (two out of 17), Finance, Geoscience, Sustainability Science, Law, Biology, Biomedical Engineering and Research & Technology (one out of 17 each).

Table 3.2 lists the exact demographic statistics of the participants, divided into the two conditions. It was desired to have an almost even distribution in most of the areas mentioned above (such as age, gender, highest degree, occupation, and type of area). By having an almost even distribution across the conditions, effects due to specialized skills, increased previous knowledge of tools or the impact of dominant groups, such as having only participants from the IT sector allocated to the digital condition, should be mitigated.

Participants did not receive compensation for taking part in this study. However, a voluntary contest to encourage active participation enabled the user with the highest score to receive a small reward. Other than that, participants were provided with drinks and snacks.

3.4 Task

This study's task is based on the first part of VAST's 2014 challenge scenario, "The Kronos Incident" [Com15]. The scenario is about a celebration of the company GAStech, which has been operating a natural gas production site in the island country of Kronos. During the celebration, several employees of GAStech go missing. An organization known as the Protectors of Kronos (POK) is suspected in the disappearance. The celebration and related events are documented by different newspapers, which form the basis of the dataset. The original dataset of the first mini-challenge of 2014 consists of approximately 900 documents, mostly news articles, a few historical documents, a map and supporting pieces of information, and is available to the public to download. Usually, the datasets of VAST Challenges are intended for automated analysis. In our case, however, the dataset should be worked through manually by the participants of the study, which is not possible with 900 documents. Since the whole dataset would have been overwhelming for the

Demographic	A	D
Age		
< 18	0	0
18 - 24	0	1
25 - 34	8	7
35 - 44	0	1
> 45	0	0
Gender		
Male	4	5
Female	4	4
Native language		
German	8	7
English	0	1
Serbian	0	1
Dominant Hand		
Left	1	0
Right	7	9
Visual Acuity		
Normal	4	0
Uncorrected	0	2
Corrected	4	7

Demographic	A	D
Highest Degree		
Less than a high school diploma	0	0
High school degree or equivalent	1	2
Bachelor's degree	5	5
Master's degree	2	2
Doctorate	0	0
Occupation		
Studying	3	3
Working	5	6
Domain		
Computer Science	3	5
Marketing & Communication	1	1
Biomedical Engineering	1	0
Business & Accounting	1	0
Sustainability Science	1	0
Geoscience	1	0
Research & Technology	0	1
Biology	0	1
Law	0	1
Total	8	9

Table 3.2: Demographic participant statistics of the analog (A) and digital (D) condition.

users, the original dataset was filtered and reduced to a smaller, more reasonable amount of documents to fit the session's anticipated length and complexity. Mainly the blog posts of the fictional newspapers "Homeland Illumination," "Kronos Star" and "The Abila Post" were re-used and combined to three documents of two to three pages each, as well as the map, the three pages long five-year report about the history of the Protectors of Kronos and ten smaller articles of about a quarter of a page each from the newspapers "Kronos Star," "The Abila Post" and "The World". Articles and blogposts of the four mentioned fictional newspapers were chosen since they were published by the major newspapers covering the key elements of this mini-challenge. Additionally, the publication date was of importance: The blogposts were all published during the key kidnapping period of January 20-21. Seven of the smaller ten articles were also published on these days. The remaining three include information about the day before the kidnapping took place as well as information about an IPO, in which GASTech International sold shares a month earlier.

The task comes with a vague goal of figuring out a kidnapping that took place during a celebration and characterizing the events surrounding the disappearance. Figure 3.1 shows the task description, with which participants of this study were provided. From there on, participants need to work through the data set. Solving the task requires reading articles in detail, connecting the data, and making several intuitive leaps since the newspapers' reports are partially inconsistent or incomplete. Although all articles together provide all the relevant information about the kidnapping, there is still enough space for speculations, hypotheses, and creative ideas. Participants had a maximum of 45 minutes to work on the task. Afterward, they gave an oral summary of around five minutes about what they have found out.

VAST Challenges are widely known and recognized as standard visual analytics tasks, as they provide realistic tasks and data sets and are designed so that analysts could be expected to make reasonable progress, if not solve the scenario. Different kinds of studies already used several VAST challenges. Andrews et al. [AEN10] used in their study the dataset of the VAST 2006 contest. Isenberg et al. [IFP+12], Jakobsen M. and HornbÆk K.[JH14], and Mahyar N. and Tory M. [MT14] respectively did the same in their studies. The task description and data from the 2011 VAST MiniChallenge 3 was used by Geymayer et al. [GWLS17]. Those are just a few examples of studies that used VAST Challenges and their datasets as basis for their research.

The VAST 2014 challenge is one of the challenges, which includes thorough text analysis. Text analysis is precisely what this study was looking for since it is a typical task of knowledge workers. The dataset is usable (and understandable) by non-experts without specialized training and does not require any previous knowledge. Nonetheless, the task provides enough complexity in order for participants to act out their active reading strategies and behaviours. Participants may find it useful to take notes, highlight text passages, and arrange or group documents depending on their contents or relevancy. Such activities are especially relevant for this study since the main interest is in describing how users worked rather than in the outcome and performance of their work. Nevertheless, a description of how participants progressed on the task is given, too, to be able to assess whether analog or digital multi-document active reading is more efficient.

3.5 Apparatus

The study took place in the library (room number HG0502) of the Research Unit of Computer Graphics at the TU Wien Institute of Visual Computing & Human-Centered Technology (Faculty of Informatics), which provided a quiet space for the study and enough room for creative thinking and undisturbed work. A glass of water was available for drinking, as well as a small bowl of snacks to increase participants' concentration.

Participants of the analog condition worked on a large table (1.6 x 1.2 m). They had access to empty sheets of paper (DIN A3 and A4), blue and pink sticky notes, a green, yellow and pink highlighter and pens in three different colors (blue, black and red) as well as scissors and adhesive tape. Print outs of the task materials were placed next to each other on the top end of the table. The smaller articles as well as the map were

TASK DESCRIPTION

Note: This scenario and all the people, places, groups, technologies, contained therein are fictitious. Any resemblance to real people, places, groups, or technologies is purely coincidental. Document titles and colors as well as their arrangement aren't of any importance for this task and only important for a better analysis of the resulting video materials.

Background: In the roughly twenty years that GAStech, a company based from Tethys, has been operating a natural gas production site in the island country of Kronos, it has produced remarkable profits and developed strong relationships with the government of Kronos. However, GAStech has not been as successful in demonstrating environmental stewardship.

In January, 2014, the leaders of GAStech are celebrating their new-found fortune as a result of the initial public offering of their very successful company. But the celebration doesn't go as planned.

As an expert, you are called in to help law enforcement from Kronos and Tethys assess the situation. Time is of the essence.

Task: Your task is to find out what happened on the day of the celebration in only 45 minutes. In order to better understand the occurrences, you should also get familiar with the organization known as the Protectors of Kronos (POK).

The result should be a 5 minute oral report of your most important findings about what has happened from January 20, 2014 until January 21,2014. Details, such as names and timestamps, play an important role.

You can use any provided material for solving this task and giving your oral report. Feel free to do whatever helps you for this task, might it be notes taking, drawing, annotating or something completely different. Modifying provided material as well as creating new ones is completely fine.

Figure 3.1: The task description that users have received (background taken over of VAST's 2014 challenge [Com15]; task slightly adapted).

Document	Position	Size	Orientation	Sides	Pages	Words	Variant	Color
A Map of Kronos	1	A5	Landscape	1	1	4	Loose	None
5 year report	2	A4	Portrait	2	3	1431	Loose	
Kronos	3	A4	Portrait	1	2	933	Loose	
Abila	4	A4	Portrait	1	3	1058	Bound	
Illumination	5	A4	Portrait	2	3	1194	Bound	
50	6.1	A5	Landscape	1	1	178	Loose	
174	6.2	A5	Landscape	1	1	138	Loose	
461	6.3	A5	Landscape	1	1	131	Loose	
481	6.4	A5	Landscape	1	1	149	Loose	
559	6.5	A5	Landscape	1	1	166	Loose	
693	6.6	A5	Landscape	1	1	112	Loose	
713	6.7	A5	Landscape	1	1	209	Loose	
718	6.8	A5	Landscape	1	1	187	Loose	
764	6.9	A5	Landscape	1	1	99	Loose	
824	6.10	A5	Landscape	1	1	129	Loose	

Table 3.3: Characteristics of the print outs of the task materials.

each printed on DIN A5 in landscape mode, while the longer documents including the blogposts of the fictional newspapers as well as the five-year report about the history of the Protectors of Kronos were each printed in portrait mode on DIN A4, resulting in 11 A5 papers and four A4 documents. Every article was printed out one-sided, except for the blogpost of "Homeland Illumination" and the history of the POK, which were printed out two-sided to make reading a little bit more difficult for the analog condition and to create a more similar setup to scrolling on computers. Both documents were the longest with three pages of content, which is why they were chosen for two-sided printing. Some of the longer documents were then stapled together while others were kept loose.

In the end there were two loose documents and two stapled documents consisting of one one-sided document and one two-sided document each to represent all possible combinations. The small articles were further sorted by their document name (located in the upper right corner) in ascending order and placed in a corresponding stack (largest document name at the very bottom) at the right end of the printouts. All print outs, except for the map, had a color tag in the upper right (A4 documents) or the bottom right corner (A5 documents) to be able to better distinguish the used materials in the recordings. Table 3.3 summarizes the properties of the print outs of the task materials. A tripod was located on the table across the participants' seat with a Logitech webcam (1.3 megapixels with integrated microphone) mounted on it for capturing the user's activities during the study.

The digital setup consisted of a MacBook Pro (Retina, 15-inch, Mid 2015) with macOS X, which was connected to a 54" LG monitor with a resolution of 1920 x 1080 pixels. Such a big monitor might seem unusual. However, the analog workspace in this study is also uncommonly big (and comparatively empty and tidy). With the large screen we wanted to offer the user the possibility to perform spatial organization in a digital environment, because it has been shown that the physical space provided by a large display biases the user towards working spatially, as described by Andrews et al. [AEN10]. The user was sitting around one meter away from the monitor. An external Windows keyboard, and a Logitech USB cable mouse, including a mouse pad, were provided so that participants would not need to work and interact with the laptop directly. Different kinds of text-processing and auxiliary software, such as Word, Pages (Apple's counterpart to Microsoft's Word), Acrobat Reader, Notes and TextEdit (simple text editors), and Chrome, were already pre-installed on the computer.

One approach would have been to provide the user with as many options as possible by installing a set of specialized tools for e.g., creating mind maps in addition to these conventional tools. However, since this would be more of a second-guessing of what special tools knowledge workers would usually end up using, we decided to keep the setup simple and minimalistic (similarly to O'Hara et al. [OS97]). In addition, we can investigate if users employ active reading behaviours and adapt them according to available tools even without any special tools. In order to enable the users to work with already known software tools in a preferably familiar environment, they were given the opportunity to install missing tools at the beginning of the study. However, none of the participants took advantage of this offer. The best-case scenario would have been if participants were able (and willing) to use their computers. By doing so, however, too many (confounding and random) factors and variables would have been added. Furthermore, problems of recording and activity tracking would have arisen. However, considering the operating system used, it would probably have been an advantage for the users to be able to use their own computers. Only four of the nine participants in the digital group are used to macOS X, while the remaining five are ingrained windows users.

No webcam was necessary for capturing the activities of users of the digital condition since the screen was recorded during the whole session. Further field logging was added by tracking the window management as well as keyboard strokes (see Subsection 3.8.1 for more details). The task materials were provided digitally (as .pdf, .doc or .txt format) and arranged on the desktop the same way as for the analog condition on the table. The ten small articles, which were stacked on the table for the analog condition, were grouped in a folder called "articles" in the digital condition. All files were initially closed. Participants of the digital setup were not allowed to use any analog material at all.

Figure 3.2 shows both setups. As can be seen in the pictures, both conditions provide more free space than most users' usual work area. The initial arrangement of the task materials in both settings can be seen in detail in Figure 3.3.

3.6 Procedure

The procedure for both conditions was almost the same. Participants were greeted and then seated themselves around the table. They received an information sheet about the study and were asked to read it thoroughly. The information sheet included general information about the study, such as the expected duration, as well as information

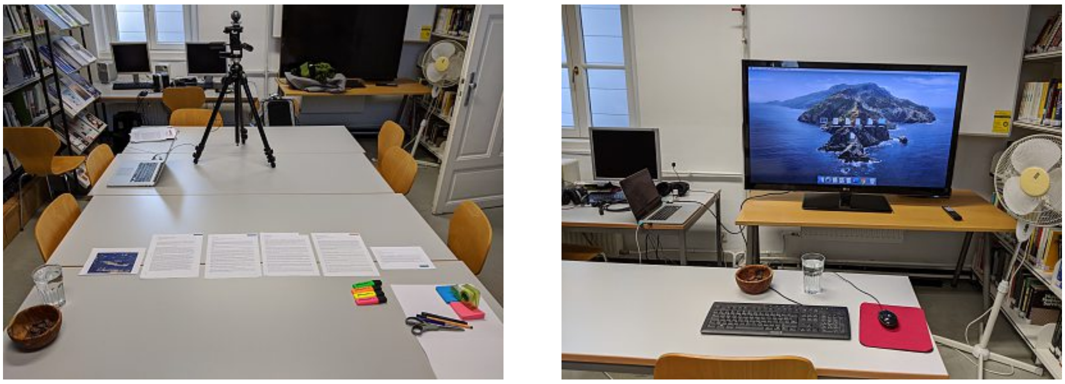


Figure 3.2: The setup of the analog (left) and digital (right) condition.

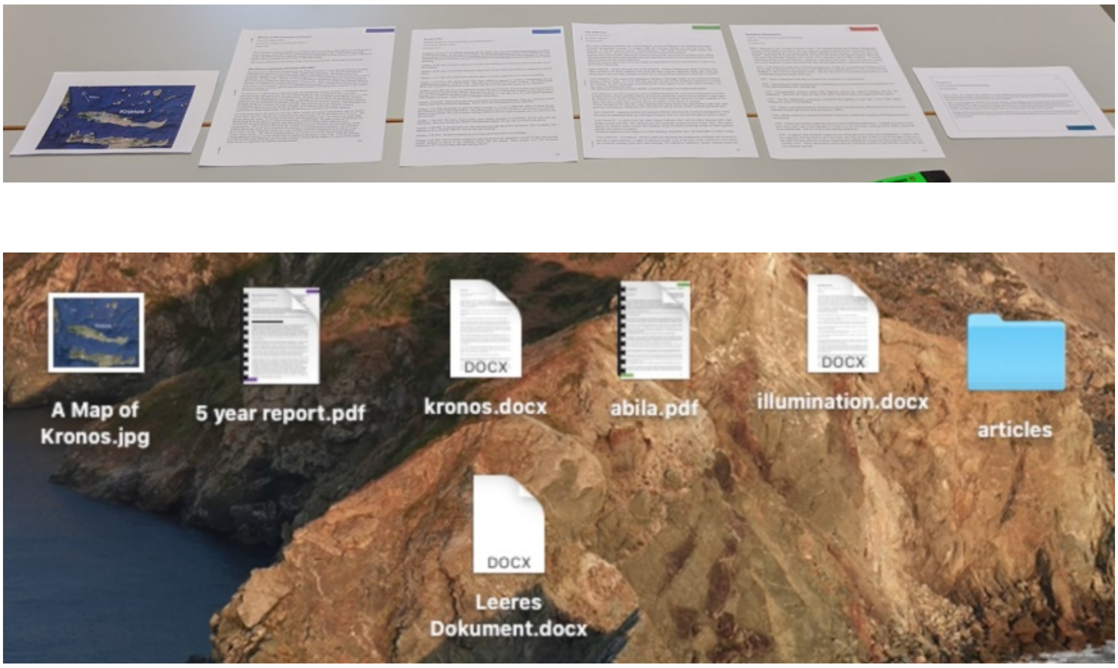


Figure 3.3: The initial arrangement of the task materials in the analog (top) and digital (bottom) condition.

about data collection (such as video- and audio-recordings) for later analysis. The study administrator then answered open questions and handed out the consent form and a demographic questionnaire. After signing the consent form and filling out the demographic questionnaire, participants could sign up for a voluntary contest that should engage users in active participation. Furthermore, users of the digital condition received a cheat sheet with practical keyboard shortcuts and were able to get to know the setup by trying those shortcuts in a couple of minutes. No further tutorials about the usage of provided software (features) and tools were held, as this would bias the participants and might have changed the active reading strategies used by them.

Having assured that the participants were comfortable with the setup and provided materials, the study administrator handed out the task description and gave participants a short briefing, including the clarification of further questions. Participants were asked to set a timer for 45 minutes on their mobile phones for keeping track of time. That was also the maximum time they were given for solving the task. Besides, the study administrator set the alarm, too, and notified participants, if they wanted, ten minutes before the end of their given time. The video- and audio-recording started simultaneously with the timer. Participants could speak freely throughout the whole session. Furthermore, the study administrator has taken a few field notes while participants were working on the task.

After having worked on the task for a maximum of 45 minutes, participants verbally reported their results in around five minutes. Upon completion, participants answered six questions concerning the contents of the "Kronos Incident":

1. How many GasTech employees are missing?
2. Are there concrete indications (regarding the responsibility for the kidnapping)? If so, which ones?
3. What was celebrated at the GasTech meeting on 14 January?
4. Why does POK have a problem with GasTech or what does POK criticize about GasTech?
5. List all names of missing GasTech employees you discovered / found out.
6. List all members of the Vann family and their roles in / for PoK.

For the oral report as well as answering the questions about the task, participants could use all provided materials, including their notes. Following these questions, the study ended with a debriefing and a semi-structured, open-ended interview, where participants were asked to comment on their strategy of solving the task, the reasons for doing things the way they did, and what they might have handled differently in their usual (work) environment. The sessions were recorded until the end of the interview. Sessions lasted one hour on average and were held individually.

3.7 Pilot Study

A preliminary study has been conducted with two participants, one for each condition. These participants were peer researchers and students who performed the task voluntarily. The expected outcome of the pilot study was the finalized study design.

The pilot study showed that the setup of both conditions was not equivalent enough and needed a lot of adaptations and improvements. One significant difference was the initial arrangement and order of the provided articles, which led to confusion and extremely different outcomes. Since the participant of the digital condition did not take any notes at all, an empty document was added to engage notes taking. Furthermore, the task description was adapted for resulting in a more detailed oral report by bringing the focus on details such as names and timestamps. That way, participants should further be engaged in highlighting and taking notes. The task description clearly states that participants are allowed to use any provided material for solving the task and giving the oral report. It further states that participants could feel free to do whatever helps them for the task, whether notes-taking, drawing, annotating, or something completely different. Modifying provided material as well as creating new ones is fine.

Another addition, due to the preliminary study, was the introduction of the contest. The contest should be an incentive for active participation by rewarding the best-ranked participant with a small price at the end of the study. In order to be able to rank users and rate their performance, concrete questions about the contents of the task were defined. These questions are not announced to the participants in advance, which means that they do not know in advance what they should focus on in particular for answering the questions. Since these questions did not yet exist in the pilot study, it is difficult to assess which of the two participants ultimately performed better. In their oral report, however, both participants mentioned roughly the same events, albeit with different details. Furthermore, field- and key-logging was added to the digital condition to allow easier and more detailed analysis.

3.8 Data Collection and Analysis

Several types of data were collected and analyzed. First, each session was recorded with video and audio. The recordings comprised 17 hours and 17 minutes in total, excluding time spent on introductions and filling in the demographic questionnaire. On average, sessions lasted around one hour ($\sigma = 7.28$ minutes), including the time provided for solving the task ($\mu = 42.4$ minutes of the provided 45 minutes, $\sigma = 5.36$ minutes), the oral summary ($\mu = 4.44$ minutes, $\sigma = 1.52$), the post-questionnaire ($\mu = 4.24$ minutes, $\sigma = 1.64$ minutes) as well as the interview ($\mu = 9.89$ minutes, $\sigma = 3.24$). Because of a short interruption, one participant of the digital condition, D3, was given one more minute for solving the task. Figure 3.4 shows the distribution of the durations (in minutes) for the task, the oral summary, the post-questionnaire, the interview as well as the total duration of the sessions divided into the two conditions.

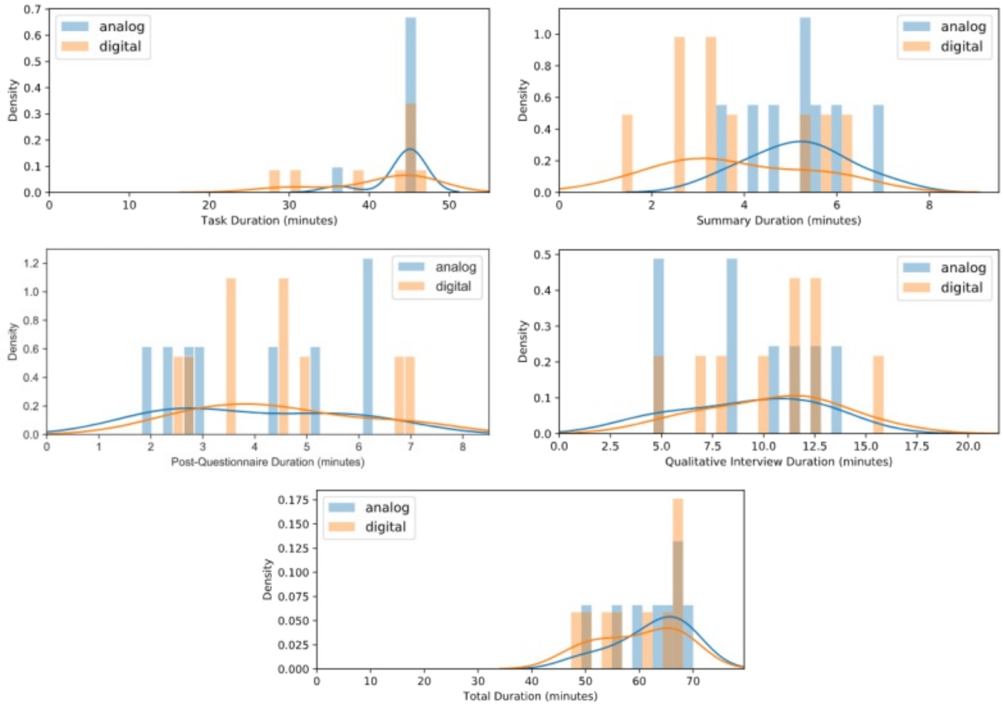


Figure 3.4: The density distribution of durations (in minutes) of different parts of the study, divided by condition.

Second, we instrumented the digital condition system to automatically collect time-stamped data describing participants' interaction with the system, including the tracking of window events so that we could determine how the display space was used throughout the task. Additionally, a picture of the workspace was taken at the end of each analog session to capture the desk organization after finishing the task. At the end of each digital session, a screenshot was also taken to capture the arrangement of the windows at that time. Finally, we collected all artifacts used or created during the task by participants, including their notes and annotations.

For the statistical analysis we used the free and open-source statistical software jamovi [Mei20a].

3.8.1 Window and Keyboard Logging

In total, 1 373 window events (e.g., dragging a window, minimizing or maximizing a window or closing a document, but excluding keyboard taps) and 10 311 keyboard strokes were logged. The logs of the window events include timestamps, the application which fired the event, as well as the title of the window/document and the type of the event. A list of all possible event types including short descriptions is given in Table 3.4.

Furthermore, each time a window event was triggered, we logged the current state,

Event Type	Description
CREATE	A new window was created
DESTROY	A window was destroyed
FOCUS	A window received focus
HIDDEN	A window was hidden (its application was hidden, e.g., via cmd−h or via right click on the application and then “Hide”; only possible for all windows of an application and not for specific ones)
MINIMIZED	A window was minimized
MOVE	A window was moved or resized, including toggling fullscreen/maximize
UNHIDDEN	A window was unhidden (its application was unhidden, e.g., via cmd−h)
UNMINIMIZED	A window was unminimized

Table 3.4: Descriptions of logged window event types.

including the count of distinct open applications, open windows (in total as well as divided into visible and overlapped/hidden ones), the title and application of the window that had focus at that moment, and the percentage of used display space, again provided with a timestamp. The keyboard logs look quite similar: They include a timestamp, the application name and title of the window/document with the focus, the character of the keyboard stroke and a count, indicating how often the search functionality has been used at that time.

We added the window and keyboard logging mainly to determine how the display space was used, but also to be able to track events even if they were not “visibly” triggered by mouse events but rather via keyboard shortcuts. The keyboard logging was especially crucial for tracking whether (and how often) participants used the search functionality, as this would have been a significant advantage over the analog condition (as already discussed in Subsection 3.2.3).

For the implementation of the logging, we used Hammerspoon. Hammerspoon [Ham20] is a desktop automation tool for OS X, which bridges various system-level APIs into a Lua scripting engine, allowing to have powerful effects on the system and manipulating it by writing Lua scripts [oRdJ20]. Since OS X does not provide an integrated window manager to move windows into corners or resize them by halves, thirds or quarters, we used the pre-made Hammerspoon plugin “MiroWindowsManager [Man20]” to overcome this problem and to be able to use keyboard shortcuts for managing windows. Unfortunately, this window manager does not support to snapping of windows to a side via drag-and-drop as people might be used to on Windows. Participants of the digital condition received a cheat sheet with the keyboard shortcuts for managing windows. They were also given a few minutes to get to know the setup and try those keyboard shortcuts by themselves. Hammerspoon enabled us to easily add a window manager as well as to be able to log window events and keyboard strokes, all at one place.

Code	Type	Description
Annotations	Binary	Indicates whether annotations have been made to the source documents or not.
Marks	Binary	Indicates whether text portions have been highlighted (anchor-only annotations) or not.
Notes	Binary	Indicates whether notes have been written down on an extra sheet of paper/document or not.
Colors	Numeric	Number of different colors used for highlighting/writing/grouping.
Format	Numeric	Number of different font styles used for formatting notes, such as bold and underlined text. A different font color does not count to this code, as it is part of the code above.
References	Numeric	Number of document references written down as notes.
Connections	Numeric	Number of connections made as notes, such as connection lines or arrows.
List Items	Numeric	Number of list items in the notes.
Timeline Items	Numeric	Number of timeline items in the notes.
Paragraphs	Numeric	Number of paragraphs/ text groups in the notes.
Words	Numeric	Word count of the notes (as calculated by Microsoft Word).
Writing Style	Nominal	Indicates the writing style, which can either be none (no notes), full sentences, shorthand, or a combination of both.
Content Annotation	Numeric	Number of words or marks added to the page of a document, but not explicitly connected to any portion of the article's text.
Compound Annotation	Numeric	Number of words or marks added to the page of a document with a line or other connection indicating a relationship to a portion of the article's text.
Shapes	Numeric	Number of shapes or symbols (other than lines and arrows, including mind-maps, drawing, sketching) used for note-taking and annotations.
Tilted Elements	Numeric	Number of text parts of notes that have not been written horizontally but tilted.
Mark Thickness	Numeric	Number of different thickness types of marks.
Type of Notes	Nominal	Categorizes the type of notes (e.g., summary, timeline, concept map, ...)

Table 3.5: Coding scheme of artifacts (one value per participant for all of its documents).

3.8.2 Artifacts

At the end of the study, all artifacts used or created by participants were collected. Those artifacts include the provided task materials such as printouts and digital documents, and notes created by participants. The artifacts were then used for analysis: Highlights, annotations, words of notes and the different colors used for highlighting and writing were counted. Additionally, the structure of the created notes was analyzed in more detail, such as the number of list items, timeline items, paragraphs and references made to other documents. Table 3.5 gives a full list of the established codes for the analysis of the artifacts that resulted during the study.

We have collected the artifacts, as it is easier to directly analyse them instead of trying to do so via the recordings. In fact, it would have been impossible to recognize in the recordings, for example, what participants of the analog condition have written down. Furthermore, by collecting all materials we ensured, that participants could not share (detailed) information about the study and its task with other potential candidates.

3.8.3 Post-Questionnaire

The purpose of the post-test is mainly to be able to measure and evaluate the performance of participants and their understanding of the contents. The post-questionnaire consisted of six questions concerning the 'Kronos Incident'. Participants commented in free text on the questions and were allowed to use their notes and all provided materials for answering them. Their answers were audio recorded, while their skimming and skipping through

Question	Answer	Points
How many GasTech employees are missing?	14	10
	14; 4 reappeared	100
	10	100
Are there concrete indications (regarding the responsibility for the kidnapping)? If so, which ones?	POK claims responsibility	80
	Ransomnote	100
What was celebrated at the GasTech meeting on 14 January?	Anniversary (without any more details)	40
	Kronos-GASTech partnership/20 year anniversary (reception afterwards)	70
	Successful IPO (meeting)	100
Why does POK have a problem with GasTech or what does POK criticize about GasTech?	IPO (payout)	10
	Environmental issues (Pollution)	45
	Corruption of the government	45
	Environmental issues & IPO	55
	IPO & Corruption	55
	Environmental Issues & Corruption	90
	Environmental Issues, Corruption & IPO	100
List all names of missing GasTech employees you discovered / found out.	Not Stan Sanjorge, Jr. (CEO)	10
	Ingrid Barranco (CFO)	20
	Ada Campo-Corrente (CIO)	20
	Orhan Strum (COO)	20
	Wilhelm Vasco-Pais (ESA)	20
	GASTech executives (without names)	40
	GASTech executives & not Sanjorge, Jr.	50
	GASTech executives (including their names)	95
	GASTech executives (including their names) & not Sanjorge, Jr.	100
List all members of the Vann family and their roles in / for PoK.	Mandor Vann (+ POK's top political strategist + uncle to Isia and Julianna Vann)	10 (+10 +5)
	Juliana Vann (+ POK martyr, died by drinking contaminated water + sister of Isia Vann, niece of Mandor Vann)	10 (+10 +5)
	Isia Vann (+ current POK member, who advocates a more forceful approach + brother of Juliana Vann, nephew of Mandor Vann)	10 (+10 +5)
	Edvard Vann (+ GASTech Security Guard + not related to POK)	10 (+10 +5)
	Combination of everything	100

Table 3.6: Point scale of the questions of the post-test.

the documents was video recorded. Since participants were allowed to "look up" the answers, we measured their response time to weight their (correct) answers accordingly. In case participants did not fully understand a question, the question was repeated and paraphased, trying to give not too many clues in the right direction. We further created a point scale for each question to be able to grade each participant the same way. Table 3.6 gives a summary of all six questions and the points for (partly) correct answers.

3.8.4 Coding of Activities

For describing participants' (active reading) behaviours, we coded the activities employed by participants and their handling of provided task materials during task solving. To form initial coding categories, we used 30 percent of the video materials (three random videos of the analog condition, and three random videos of the digital condition). As a basis for the identification of relevant activities, we followed the coding scheme by Isenberg et al. [ITC08], extended it and tried to map those activities to stages of Pirolli & Card's sensemaking loop [PC05] (see Figure 2.2). After the first pass, we refined our coding set, as some codes could not successfully be mapped onto the activities employed

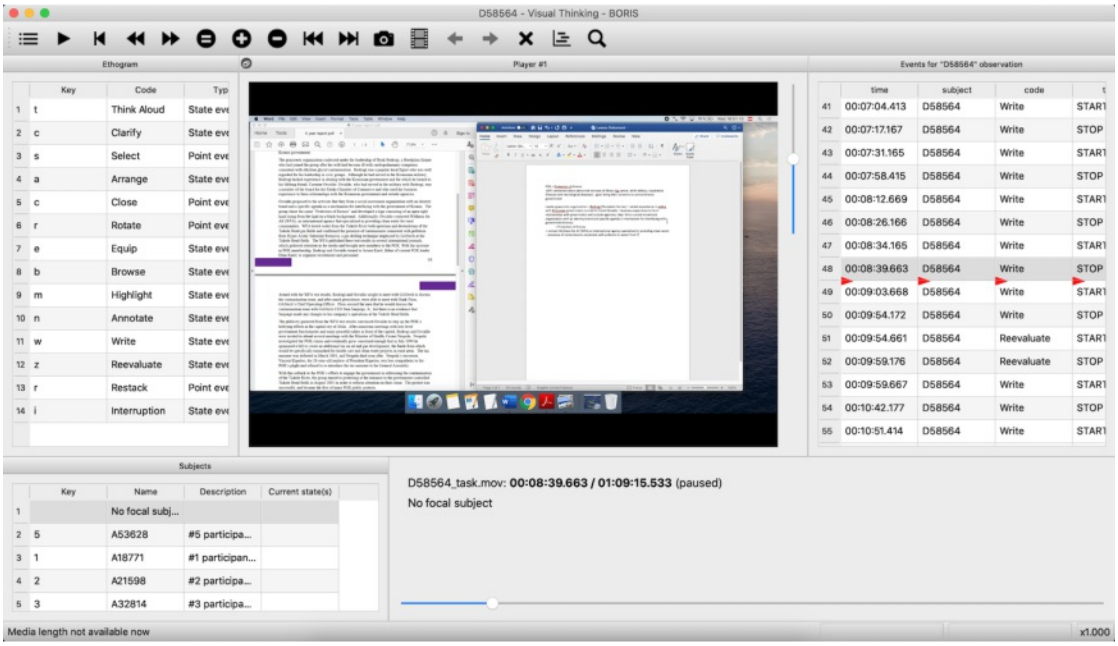


Figure 3.5: A running observation in BORIS [FG16].

in both conditions. This was partly due to the fact that the video recordings of the digital condition only captured the screen and not the participants themselves. Therefore, we were not able to distinguish whether participants were reading a document, thinking about the contents or doing something completely different (e.g., looking at the timer or adapting their position). As we did not track the eye movements, we could not track at which documents participants were looking, especially when they had multiple documents open for comparison. On the other hand, we could not distinguish in the analog setting whether participants were searching for a specific information or if they were just skimming through the documents. Table 3.7 describes the final coding set used for the coding of the activities. Video parts without a code include untraceable actions, such as reading, thinking or looking at a timer, that would need precise eye-tracking.

For the coding of the videos, we used the free and open-source software BORIS (Behavioral Observation Research Interactive Software, [FG16]). It is an easy-to-use event logging software for video/audio coding and live observations. One major advantage of using this software is that it includes a number of possibilities to extract miscellaneous visualizations, such as a plot of the recorded events or a time budget analysis. Figure 3.5 shows a running observation in BORIS.

Code	Type	Description	Analog example	Digital example
Talk	State event	Any vocal activity that does not need a dialogue partner.	Soliloquies, murmuring, expression of annoyance, ...	
Clarify	State event	Any vocal activity that involves another person.	Talking with the study administrator, asking questions, ...	
Select	Point event with modifier	Choosing/picking/selecting a document (includes initial position changes and opening a document). Includes the name of the selected document, to recognize re-opening of documents.	Choosing one of the documents and placing it in front of oneself	Opening a document/folder, unminimizing windows, ...
Arrange	State event	Any kind of arranging and moving documents/windows in x-/y-direction.	Moving documents, putting documents aside, arranging pens/tools, ...	Moving windows, resizing windows, minimizing windows, ...
Restack	Point event	Any kind of movements of documents/windows in z-direction.	Bringing documents from the back to the front or vice versa	Bringing windows from the back to the front, switching between tabs of one program (e.g., in Acrobat Reader), ...
Close	Point event with modifier	Moving documents/windows out of the current workspace. Includes the name of the closed document.	Putting documents back to their original positions	Closing windows
Rotate	Point event	Rotation of documents/windows.	Rotating documents for easier writing/reading	Rotation of windows or their contents
Equip	State event	Being equipped with tools and ready to use them.	Holding a pen/marker	Enabling the quick marking functionality, keeping the focus on the notes document while reading/scrolling through one of the other documents, ...
Browse	State event	Scanning through documents to see, e.g., the length of the document or the document structure.	Scanning through documents, searching for information	Scrolling through documents, searching for information
Highlight	State event	Highlighting text portions.	Using a marker or pen, highlighting text portions in source documents	Using the highlight functionality
Annotate	State event	Adding notes to source documents.	Adding text pieces to source documents, adding symbols/shapes to source documents	Adding sticky notes/comments in Adobe Acrobat, adding text pieces/symbols/shapes to source documents
Write	State event	Taking notes on anything other than the source documents.	Writing, note-taking	Typing, copy & paste, formatting of notes, ...
Reevaluate	State event	Reevaluation of notes/extracted information.	Reading notes, checking whether notes include read information, ...	Scrolling through the notes document
Interruption	State event	Indicates an interruption during the study.	Someone entering the room, someone other than the participant or study administrator talking, ...	

Table 3.7: Coding scheme of the participants' behaviour during task solving.

3.8.5 Transcription of Interviews

All 17 interviews – a total of two hours and 48 minutes of audio data – have been transcribed manually. Using automatic speech recognition tools, such as AmberScript [Amb20] and f4x speech recognition [ddpG20], the quality was below an acceptable standard. The poor outcome might be the result of the relatively low audio quality of the interviews and the not fully developed speech recognition of the German language, which is currently not as advanced as for the English language. We then used Express Scribe [Sof20] for the transcription process, as it allowed us to insert timestamps easily and control the playback of the audio data via keyboard shortcuts during transcribing. As the type of transcription, we decided to use intelligent verbatim transcription, which omits laughter, pauses, and fillers throughout the conversation. We performed some light editing to correct sentences and grammar, and irrelevant words were eliminated.

In a next step, we assigned the codes used for the coding of the activities to interview answers, to be able to get more insights in reasons for participants behaviour. We also had a look at how they felt about the setup, how they think they would have worked differently if they could have used tools of both environments and how their typical workplace would have looked like.

CHAPTER 4

Results

In this chapter, we present both quantitative and qualitative findings of the study. We start with an evaluation of the participants' task performance in order to be able to make statements about the efficiency of both conditions. The focus then switches to identified behaviour differences between the analog and the digital world as well as (possible) links to performance. Typical active reading activities, such as annotating, highlighting, taking notes, (spatially) organizing documents, and navigating between them are analysed and compared in more detail. Participants' strategies and approaches for solving the task of the study are evaluated, too. Finally, we present possible confounding and random effects that emerged during the analysis.

The results are a composition of concrete statistical tests to be able to confirm or reject the hypotheses defined in Section 3.1, as well as an exploratory analysis. The exploratory analysis serves in particular to find possible reasons and explanations for potential differences (and their links to performance) between the analog and digital worlds despite the use of a large screen. Therefore, the reported p-values of tests which are not directly related to answer one of our hypotheses can be seen as "exploratory p-values". An exploratory interpretation of a significant p-value typically establishes a new hypothesis [Gau15], which will then be discussed in Chapter 5 and summarized in Chapter 6.

Table 4.1 summarizes the notation of this works' results. For all our statistical tests, we used an α-level of .05. To test whether the underlying population of the selected sample is normally distributed, the Shapiro-Wilk test is used. Depending on its result, we use, in general, Student's t-test for all tests with one independent variable of two independent levels whose population is normally distributed. For those samples whose population is not normally distributed, a Mann-Whitney U test is used. The Kruskal Wallis test is used for tests with one independent variable of two or more independent levels, with ε^2 as effect size. The X^2-test is used for tests with one independent variable of two or more independent levels and with inherently categorical dependent variables. In those cases,

Symbol	Meaning
N	Population size
n	Sample size
μ	Mean
$\tilde{\mu}$	Median
σ	Standard deviation
z	Standardized test statistic
p	Probability
d	Cohen's d (effect size)
V	Cramer's V (effect size)
ε^2	Epsilon square (effect size)
r	Pearson's r (correlation)

Table 4.1: Notation of the results.

the effect sizes are reported via Crammer's V. The effect sizes of Mann-Whitney-U tests have been calculated by dividing the absolute (positive) standardized test statistic z by the square root of the population size: $\frac{z}{\sqrt{N}}$. For the interpretation of all effect sizes, we use Cohen's classification of effect sizes [Coh88]:

Small effect: $.2 \leq |d| < .5$
Moderate effect: $.5 \leq |d| < .8$
Large effect: $|d| \geq .8$

4.1 Task Performance

To test the hypothesis that analog multi-document (active) reading is more efficient than digital multi-document (active) reading, considering better performance with similar time investment (**H1**), a test for each of our four performance measures was conducted. Those performance measures are the achieved total points, the total response time, the achieved points per second and the summary duration. Two out of the four tests showed a statistically significant difference, while all tests revealed an effect.

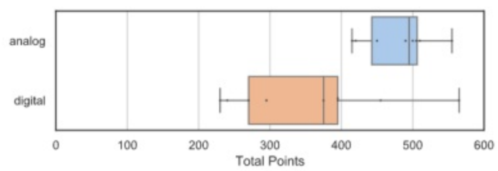


Figure 4.1: Box-plot diagram showing the achieved points of participants using only analog materials and tools ($n = 8$; $\tilde{\mu} = 495$) and those working with a digital setup ($n = 9$; $\tilde{\mu} = 375$).

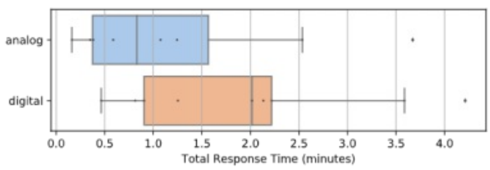


Figure 4.2: Box-plot diagram showing the total response times (in minutes) of participants' answers of the analog condition ($n = 8$; $\tilde{\mu} = .832$) and the digital one ($n = 9$; $\tilde{\mu} = 2.017$).

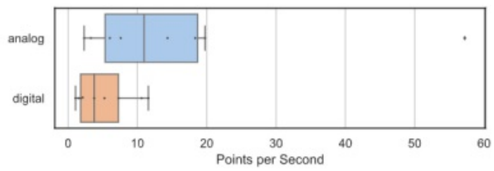


Figure 4.3: Box-plot diagram showing the achieved points of participants in relation to their response times (resulting in the achieved points per second) using only analog materials and tools ($n = 8$; $\tilde{\mu} = 11$) and those working with a digital setup ($n = 9$; $\tilde{\mu} = 3.76$).

Participants of the analog condition achieved considerably higher scores on the immediate post-test than those of the digital condition. As illustrated in Figure 4.1, scores of the analog condition ranged from 415 to 555 ($\mu = 481$, $\sigma = 48.4$), whereas digital conditions' scores were from 230 to 565 ($\mu = 358$, $\sigma = 110$). The maximum possible score was 600. A two-tailed t-test showed a statistically significant difference between the performance of analog and digital groups ($t(15) = 2.9$, $p = .011$, $d = 1.411$). The effect size for this analysis ($d = 1.411$) was found to exceed Cohen's convention for a large effect.

Even though users of the analog condition performed considerably better on the post-test, there seems to be no indication that people of the analog condition are, in general, providing correct answers faster than those of the digital one. Figure 4.2 illustrates the distribution of the users' total response times (in seconds) in each condition. A Mann-Whitney U test revealed no statistically significant difference in the total response times between both conditions ($U = 22$, $z = -1.347$, $p = .2$), however, there seems to be a small effect (.327) favouring the analog condition.

Even when taking the participants' response times into account, participants of the analog condition still achieved more points per second than those of the digital condition. A Mann-Whitney U test showed that there was a small to moderate effect (.49) and a marginally statistically significant difference ($U = 15$, $z = -2.02$, $p = .046$) between the achieved points in relation to the response times for participants of the analog condition compared to those working on the computer. The median achieved points per second were 11 points for the analog group compared to 3.76 points for those working with digital materials and tools only. The distribution of the users' achieved points per second in each condition is shown in Figure 4.3.

Furthermore, the oral reports of users of the analog condition lasted longer than those of the digital condition (Figure 4.4, $\mu_{analog} = 5.17$, $\sigma_{analog} = 1.09$, $\mu_{digital} = 3.78$, $\sigma_{digital} = 1.6$, $t(15) = 2.06$, $p = .057$, $d = 1$), with a nearly statistically significant difference and a large effect. There is also a positive correlation between the oral summary duration and the total points, $r = .637$, $p = .006$, Figure 4.5

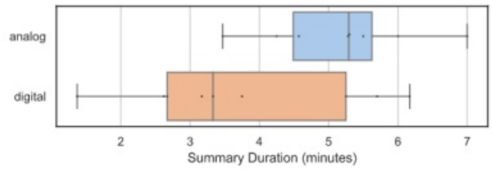


Figure 4.4: Box-plot diagram showing the durations of the participants' oral reports (in minutes) of the analog condition ($n = 8$; $\tilde{\mu} = 5.29$) and the digital one ($n = 9$; $\tilde{\mu} = 3.33$).

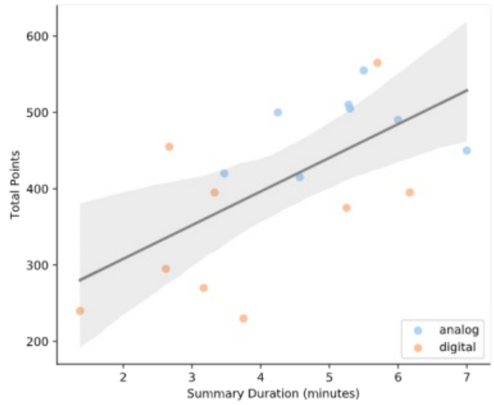


Figure 4.5: Scatterplot showing a strong, positive, linear correlation between the duration of summaries and total achieved points.

4.2 Behaviour Differences & Links to Performance

To understand how active reading behaviors and activities relate to one another in terms of temporal relationship and comparison to the different environments (analog and digital), we analyzed the video data from our study, coding each individual's activities using the behavior labels as described in Table 3.7 (see Subsection 3.8.4 for more details about the coding of the activities). This analysis revealed two aspects of participants' activity: first, while certain processes frequently occurred before others (e.g., *Select* before *Arrange* and *Equip* before *Highlight/ Write*), no common overall pattern appeared. Second, individuals varied in how they approached the task, not just between the two conditions but also within.

Figure 4.6 and Figure 4.7, respectively, show the coded temporal sequences of active reading behaviors during the task of all eight participants of the analog condition and all nine of the digital one. In the following, the number of event occurrences as well as the time spent on the respective events are used for further statistical analysis. Measures used for the evaluations are referenced back to those events for better understanding.

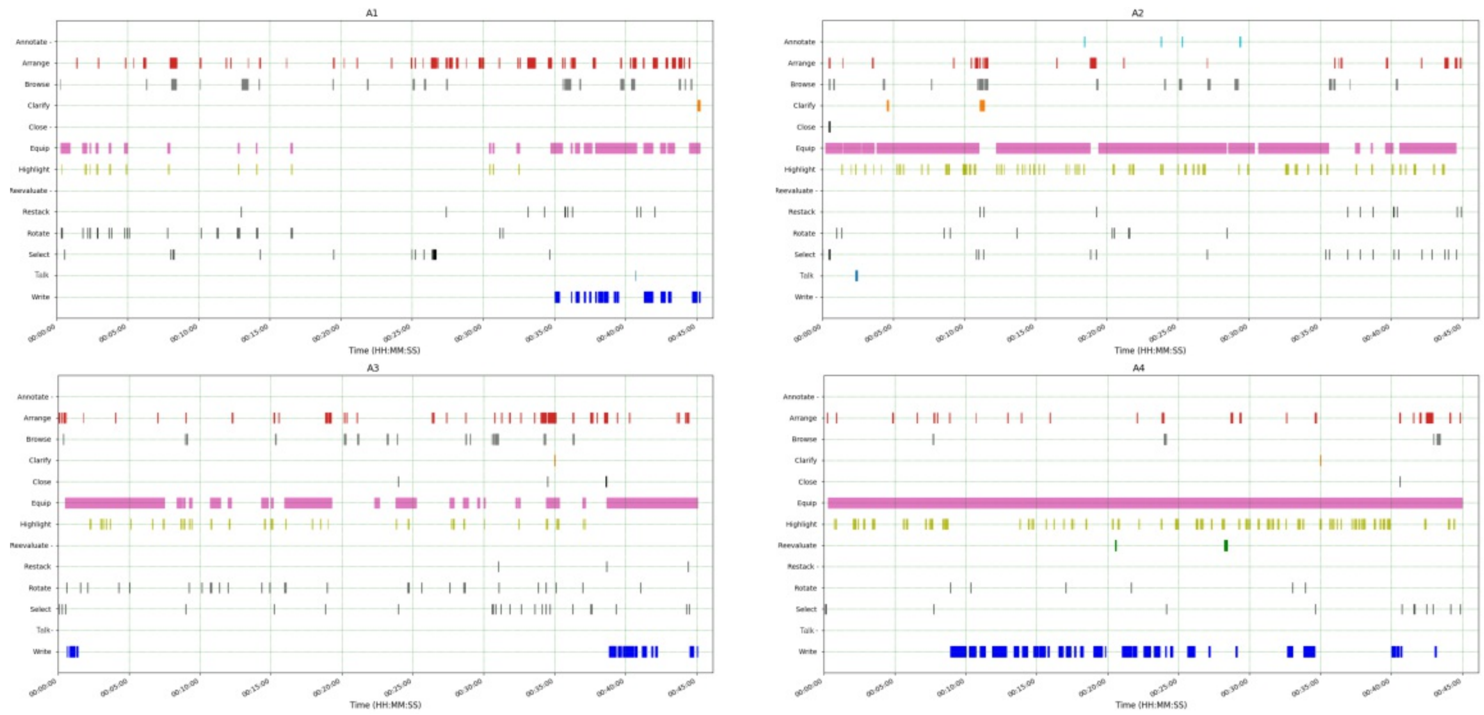


Figure 4.6: Temporal sequences of processes and activities of all eight participants of the analog condition during the complete task.

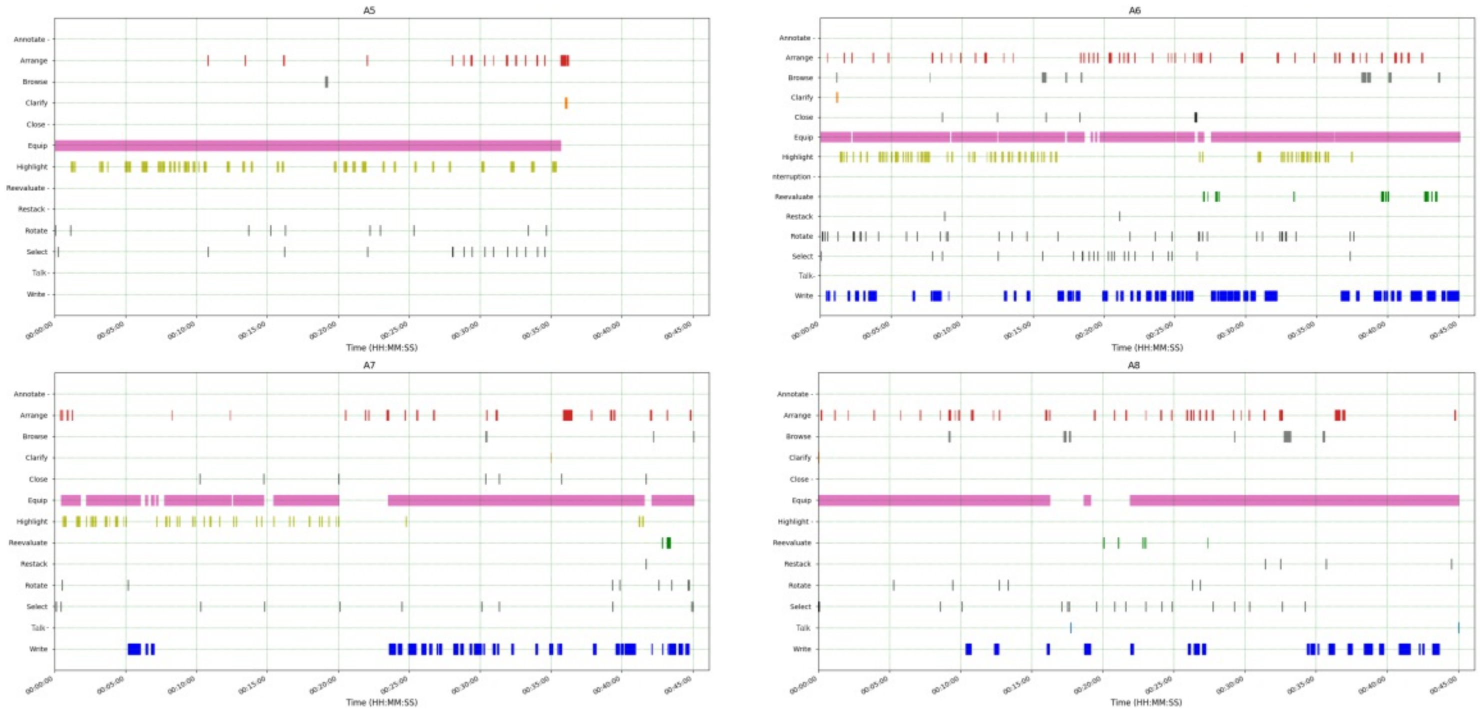


Figure 4.6: Temporal sequences of processes and activities of all eight participants of the analog condition during the complete task (cont.).

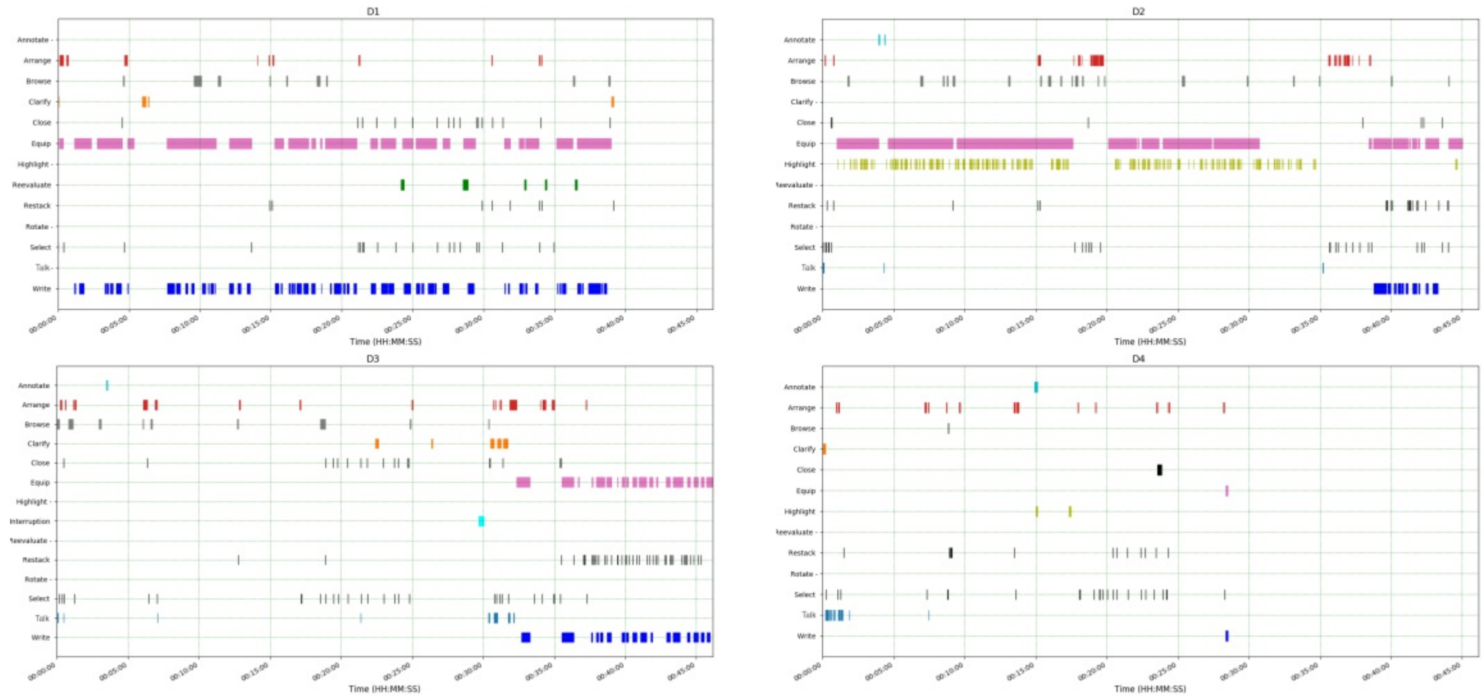


Figure 4.7: Temporal sequences of processes and activities of all nine participants of the digital condition during the complete task.

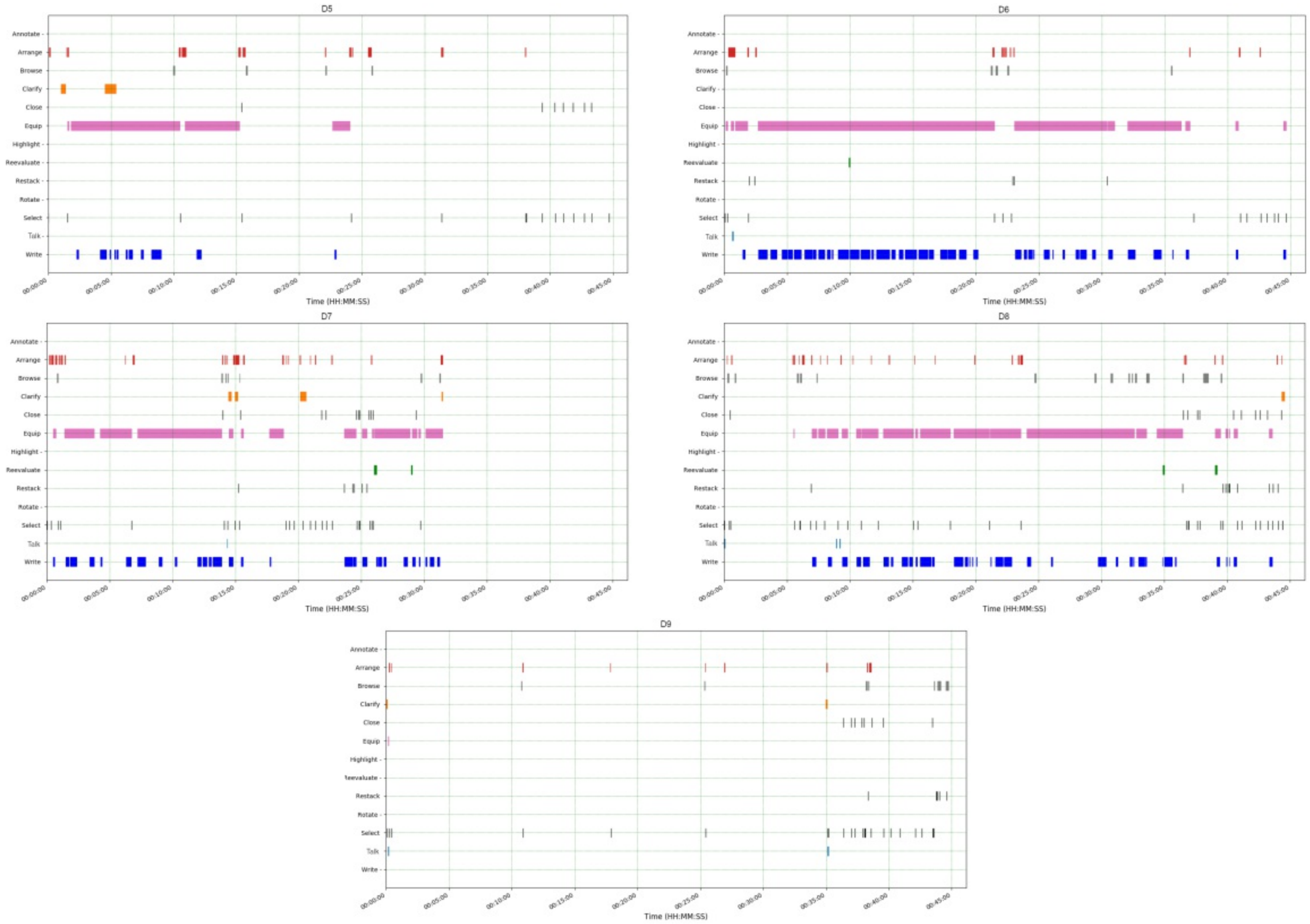


Figure 4.7: Temporal sequences of processes and activities of all nine participants of the digital condition during the complete task (cont.).

4.2.1 Annotating & Highlighting

Participants of the analog condition actively engaged in task solving and reading and made use of a lot of highlights. Figure 4.8 shows that almost no highlighting was used by participants of the digital condition, with two exception: D4 only highlighted an unclear statement and a spelling error, while D2 frequently highlighted text portions and also used the quick highlighting function provided by Microsoft Word and Adobe Acrobat Reader, mimicking the holding of a pen/marker. Figure 4.9 shows that annotating is less frequently used than highlighting, especially in the analog condition. To test the hypothesis that annotations and highlighting are less used during digital multi-document active reading using a large display compared to analog multi-document active reading, two Mann-Whitney U tests were conducted (**H2**). The first test showed a moderate effect (.548) and a statistically significant difference ($U = 12.5$, $z = -2.261$, $p = .019$) between the occurrences of participants' highlighting in the analog condition ($\mu = 40.4$, $\sigma = 24.9$) and the digital one ($\mu = 12.9$, $\sigma = 37.9$). The second test did not reveal any statistically significant effect or difference ($U = 34$, $z = -.193$, $p = .857$) between the occurrences of participants' annotations in the analog condition ($\mu = .628$, $\sigma = 1.41$) and the digital one ($\mu = .444$, $\sigma = .726$).

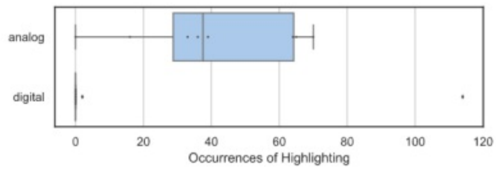


Figure 4.8: Box-plot diagram showing the highlighting occurrences of participants in the analog condition ($n = 8$; $\tilde{\mu} = 37.5$) and the digital one ($n = 9$; $\tilde{\mu} = 0$).

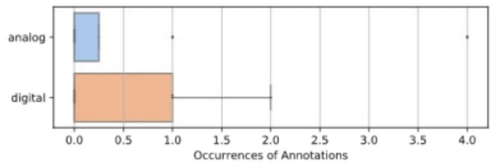
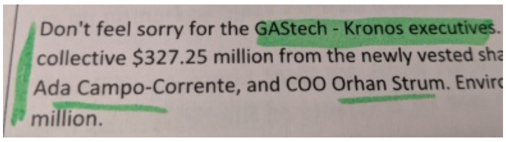
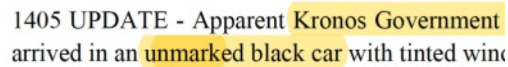


Figure 4.9: Box-plot diagram showing the annotation occurrences of participants in the analog condition ($n = 8$; $\tilde{\mu} = 0$) and the digital one ($n = 9$; $\tilde{\mu} = 0$).

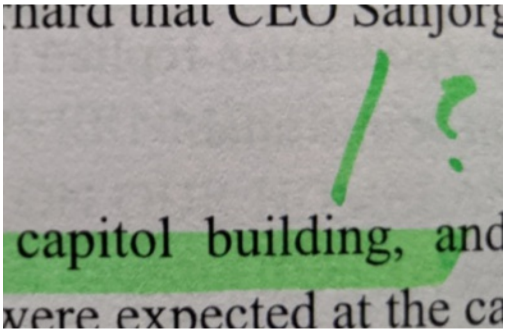
Participant A7 explained why highlighting text portions is so essential for him during active reading. He said that he uses markers *"so that I can see at a glance what was important."* Although he is the only participant who has explicitly mentioned this reason, the frequency of use of markers by the other participants of this condition shows that they feel similarly. Participant D1 mentioned that he is not used to highlighting on computers. All others of the digital condition claimed that the functionality is *"unappealing"*, *"impractical"* and *"cumbersome"*, as a lot of clicks are involved. D2, who was the only one in the digital condition who highlighted a massive amount of text portions, reported that he had difficulties with digitally highlighting since the wrong word (or row) is often marked, and corrections are tedious. According to him, however, it works better in a PDF document (with Adobe Acrobat Reader) than in a Word document, because in the PDF, the markings are more accurate. D6 also argued that she *"has written down the information anyway."* Figure 4.10 shows examples of highlightings and



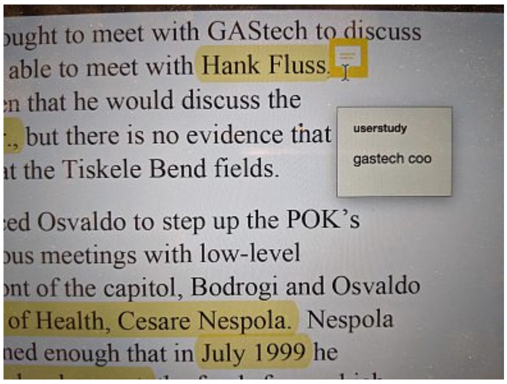
(a) Highlights of different stroke sizes made by participant A2 and A6 of the analog condition.



(b) Digital highlightings made by participant D2 in Adobe Acrobat Reader, showing "*nmark*" has been highlighted twice (stronger color saturation).



(c) Small annotations in form of symbols or characters such as a question mark, an exclamation mark or arrows made by participant A2 and A6 of the analog condition.



(d) An annotation made in Adobe Acrobat Reader by participant D2.

Figure 4.10: Examples of highlightings and annotations made in the analog and the digital condition.

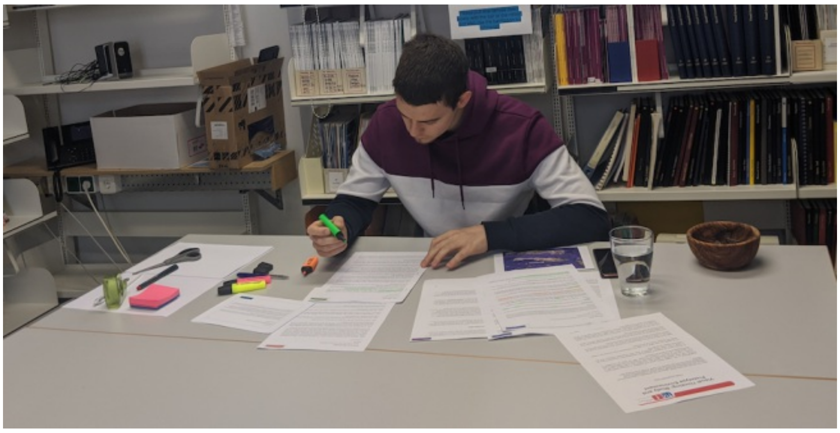


Figure 4.11: Participant A2 while working on the task: He can be seen while highlighting text portions during reading.

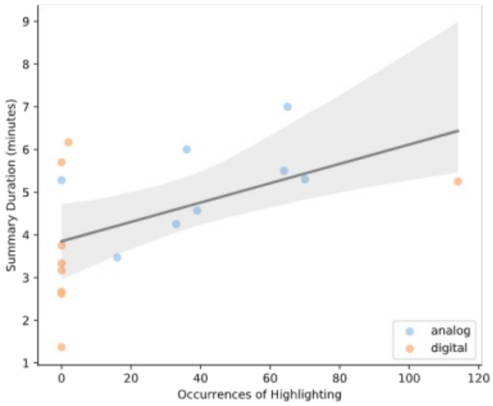


Figure 4.12: Scatterplot showing a moderately strong, positive correlation between the number of highlight occurrences and duration of summaries.

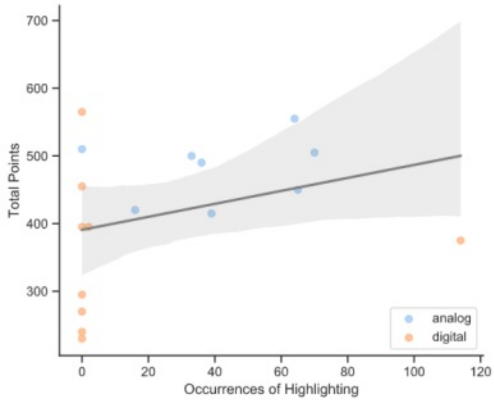


Figure 4.13: Scatterplot showing no significant correlation between the number of highlight occurrences and the achieved total points.

annotations made in the analog and digital condition. In Figure 4.11, participant A2 can be seen while working on the task and highlighting text passages.

We found a positive correlation between the number of highlighting occurrences and the summary durations, $r = .514$, $p = .035$, Figure 4.12. There was no correlation between the number of highlighting occurrences and the achieved total points, $r = .314$, $p = .22$, Figure 4.13. Also, we could not find any significant correlations between the number of highlighting occurrences and the task performance for the analog and digital condition only.

4.2.2 Note-Taking

Figure 4.14 shows the number of words written down in the analog and digital condition. Contrary to our expectation that people take more notes while digital multi-document active reading than while analog multi-document active reading (**H3**), there was no statistically significant difference between the number of words written down by participants of the analog and digital condition ($t(15) = -.753$, $p = .463$, $d = -.366$). However, a small effect has been found according to Cohen's convention. Unsurprisingly, the time spent on writing positively correlates with the number of written words ($r = .945$, $p =< .001$, Figure 4.15). However, there was no correlation between the number of words written down and the task performance (total points, $r = -.081$, $p = .758$, Figure 4.16; points per second, $r = .144$, $p = .582$, Figure 4.17 and the summary duration, $r = .062$, $p = .814$).

When qualitatively analyzing the notes of participants, we identified six different types/-categories over both conditions:

1. **None** (A2, A5, D4, D9)

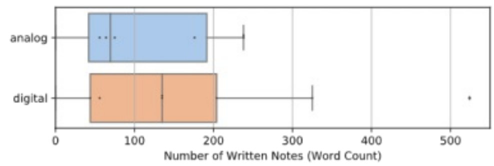


Figure 4.14: Box-plot diagram showing the number of words written down in the analog condition ($n = 8$; $\tilde{\mu} = 69.5$) and the digital one ($n = 9$; $\tilde{\mu} = 135$).

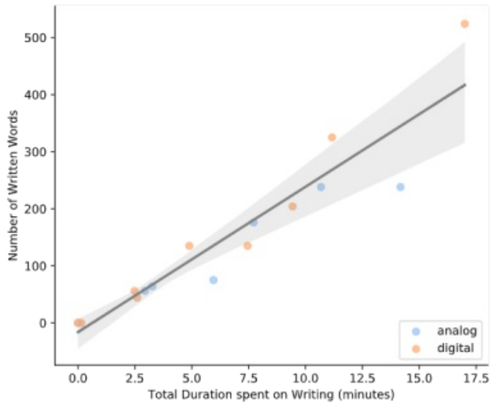


Figure 4.15: Scatterplot showing a strong, positive, linear correlation between the time spent on writing and the number of written words.

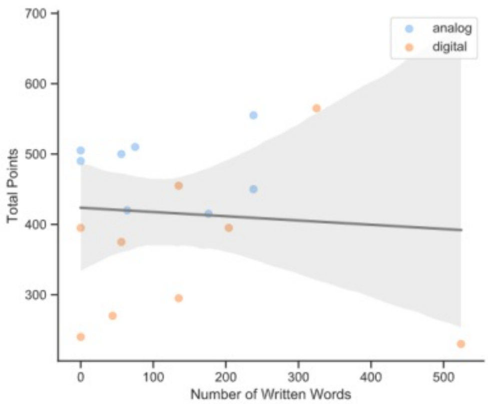


Figure 4.16: Scatterplot showing no significant (negative) correlation between the number of written words and the achieved total points.

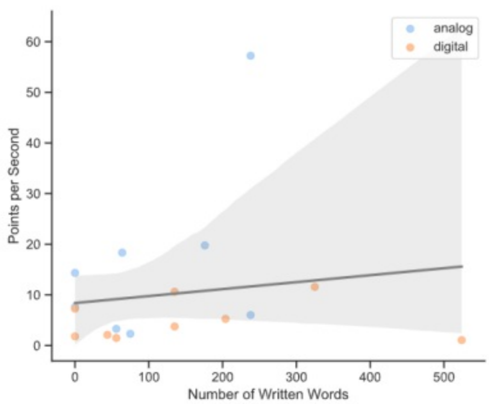
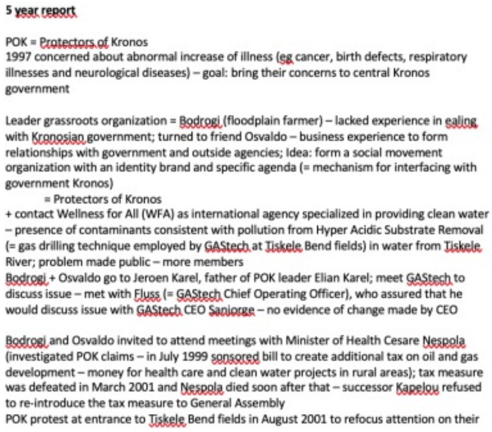


Figure 4.17: Scatterplot showing no significant correlation between the number of written words and the points per second.



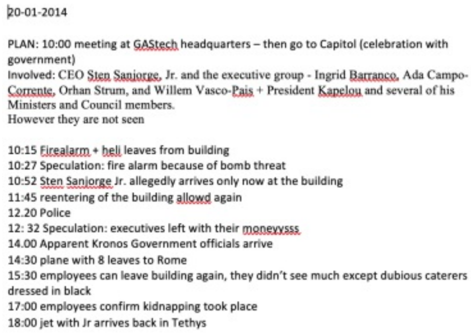
5 year report

POK = Protectors of Kronos
1997 concerned about abnormal increase of illness (eg cancer, birth defects, respiratory illnesses and neurological diseases) – goal: bring their concerns to central Kronos government

Leader grassroots organization = Bodrogi (floodplain farmer) – lacked experience in ealing with Kronosian government; turned to friend Osvaldo – business experience to form relationships with government and outside agencies; Idea: form a social movement organization with an identity brand and specific agenda (= mechanism for interfacing with government Kronos)
= Protectors of Kronos
+ contact Wellness for All (WFA) as international agency specialized in providing clean water
– presence of contaminants consistent with pollution from Hyper Acidic Substrate Removal (= gas drilling technique employed by GAStech at Tiskele Bend fields) in water from Tiskele River; problem made public – more members
Bodrogi + Osvaldo go to Jeroen Karel, father of POK leader Elian Karel; meet GAStech to discuss issue – met with Fluss (= GAStech Chief Operating Officer), who assured that he would discuss issue with GAStech CEO Sanjorge – no evidence of change made by CEO

Bodrogi and Osvaldo invited to attend meetings with Minister of Health Cesare Nespola (investigated POK claims – in July 1999 sponsored bill to create additional tax on oil and gas development – money for health care and clean water projects in rural areas); tax measure was defeated in March 2001 and Nespola died soon after that – successor Kapelou refused to re-introduce the tax measure to General Assembly
POK protest at entrance to Tiskele Bend fields in August 2001 to refocus attention on their

(a) An excerpt of a long outline made by participant D6.

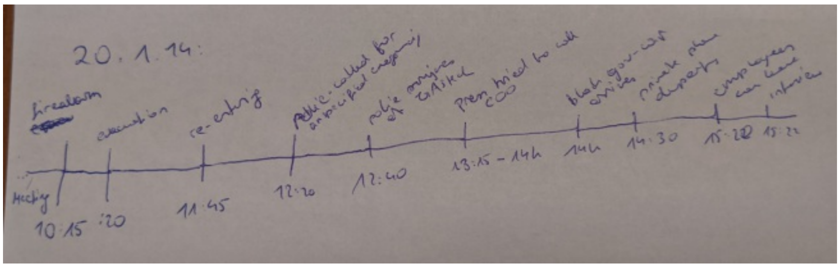


20-01-2014

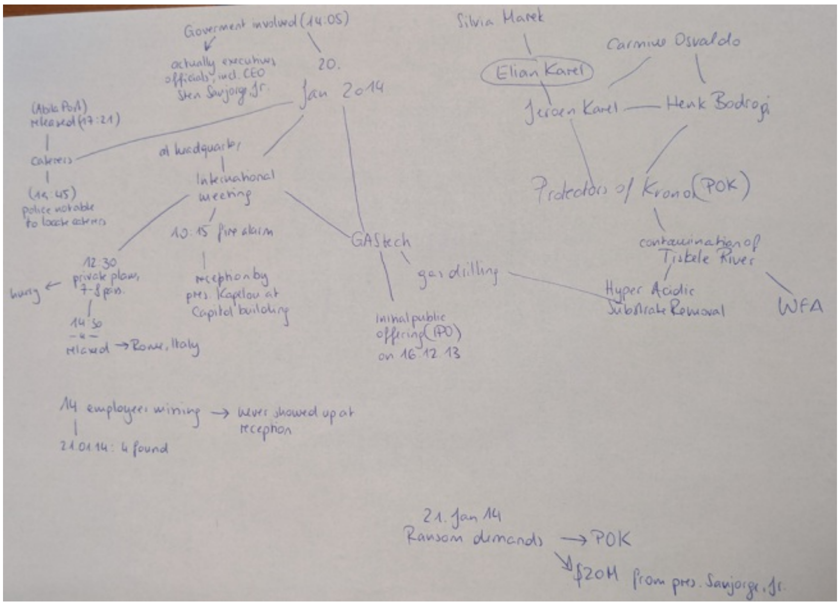
PLAN: 10:00 meeting at GAStech headquarters – then go to Capitol (celebration with government)
Involved: CEO Sten Sanjorge, Jr. and the executive group - Ingrid Barranco, Ada Campo-Corrente, Orhan Strum, and Willem Vasco-Pais + President Kapelou and several of his Ministers and Council members.
However they are not seen

10:15 Firealarm + heli leaves from building
10:27 Speculation: fire alarm because of bomb threat
10:52 Sten Sanjorge Jr. allegedly arrives only now at the building
11:45 reentering of the building allowd again
12.20 Police
12: 32 Speculation: executives left with their moneyyss
14.00 Apparent Kronos Government officials arrive
14:30 plane with 8 leaves to Rome
15:30 employees can leave building again, they didn't see much except dubious caterers dressed in black
17:00 employees confirm kidnapping took place
18:00 jet with Jr arrives back in Tethys

(b) A timeline in form of a list made by participant D3.



(c) The only timeline drawn horizontally during this study by participant A8.



(d) The only concept map constructed during this study by participant A6.

Figure 4.18: Examples of four of the identified type of notes: an outline, a timeline (both in in list form and graphical) and a concept map.

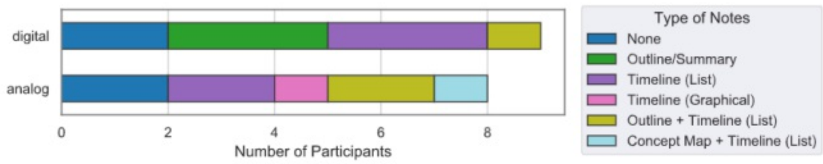


Figure 4.19: Bar-plot showing the usage of the different types of notes in both conditions.

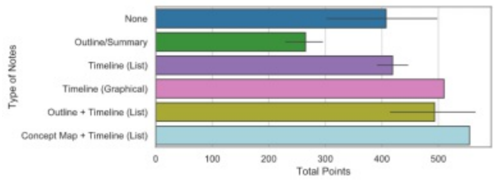


Figure 4.20: Bar-plot showing the achieved total points for each type of notes including error bars.

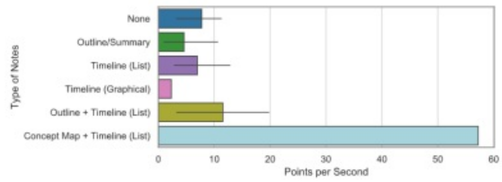


Figure 4.21: Bar-plot showing the achieved points per second for each type of notes including error bars.

2. **Outline/Summary**: An outline/summary of the provided materials, both in keywords as well as complete sentences (D5, D6, D7, Figure 4.18a).

3. **Timeline (List)**: A timeline in form of a list (A1, A4, D2, D3, D8, Figure 4.18b).

4. **Timeline (Graphical)**: A timeline drawn horizontally (A8, Figure 4.18c).

5. **Outline & Timeline (List)**: A combination of a (short) outline and a timeline in form of a list (A3, A7, D1).

6. **Concept Map & Timeline (List)**: A combination of concept map(s) and a timeline in form of a list (A6, Figure 4.18d).

Examples of four of the different note types can be seen in Figure 4.18. A6 was the only one that created a concept map. She said that she *"likes the creation of mind maps or clusters, where information can be added without a specific hierarchy"*. She also mentioned that she would always create those maps analogously as until now she was not able to find a software solution that suits her needs.

Figure 4.19 shows the usage of all six types in both conditions. Even though there was no statistically significant difference between the note types used in the analog and the digital condition, we found that there is a tendency towards achieving more total points (Figure 4.20) as well as points in relation to the participants' response times (Figure 4.21) when putting more effort into the creation of their notes (e.g., by structuring their notes into a concept map or timeline rather than just writing a summary in keywords).

The number of timeline items in the notes positively correlates not only with the total points, $r = .53$, $p = .028$, Figure 4.22, but also with the points per second, $r = .488$,

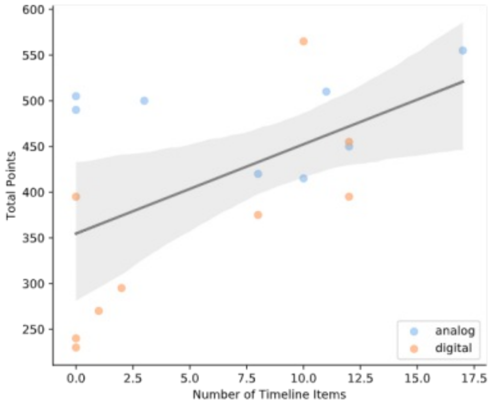


Figure 4.22: Scatterplot showing a moderately strong, positive, linear correlation between the number of timeline items and achieved total points.

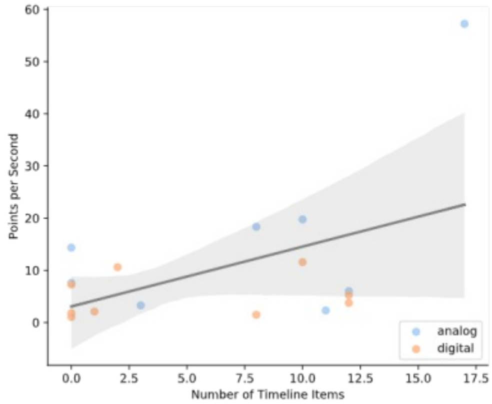


Figure 4.23: Scatterplot showing a moderately strong, positive, linear correlation between the number of timeline items and points per second.

$p = .047$, Figure 4.23, indicating that writing down important events in the correct order does have a positive effect on the task performance. As can be seen in Figure 4.23, participant A6 created the most timeline items and achieved the most points in relation to the participants' response times. There was, however, no statistically significant difference between the number of timeline items made in the analog and the digital condition.

Both A7 and A4 said they were taking notes because otherwise, the information would be scattered over several different pieces of paper, which they would have to look for again. By writing those things down, they have all the most critical information in one place right next to them. A6 also said: *"I remember things better when I write them down and have a better overview."*, while A3 mentioned: *"It takes too long when I write everything down, so at some point, I just started marking things."* This statement confirms our observations that note-taking occurs mainly at the end and could be a possible explanation for this phenomenom.

4.2.3 Spatial Organization & Navigation

Participants of each condition were provided with 15 documents, ranging from short ones containing only a few paragraphs to documents with several pages. Figure 4.24 shows that almost all participants opened all documents during task solving at least once. Those, who were not able to open all documents, mostly ran out of time to do so. There was one participant (A7), who had not seen the pile with the small articles in the upper right corner of the workspace, which is why this participant only had opened six out of the 15 documents.

To test the hypothesis whether the available workspace is used and organized similarly during analog and digital multi-document active reading using a large display (**H4**),

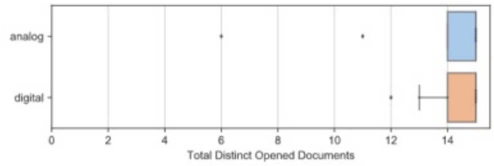


Figure 4.24: Box-plot diagram showing the number of the total distinct opened documents in the analog condition ($n = 8$; $\tilde{\mu} = 15$) and the digital one ($n = 9$; $\tilde{\mu} = 15$).

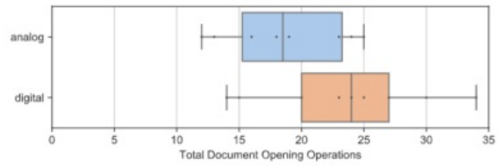


Figure 4.25: Box-plot diagram showing the total number of document opening operations (= select events) of the analog condition ($n = 8$; $\tilde{\mu} = 18.5$) and the digital one ($n = 9$; $\tilde{\mu} = 24$).

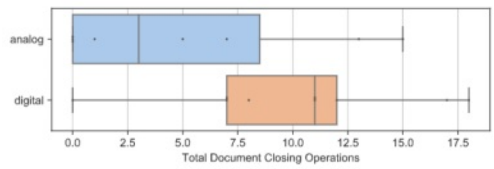


Figure 4.26: Box-plot diagram showing the total number of document closing operations of the analog condition ($n = 8$; $\tilde{\mu} = 3$) and the digital one ($n = 9$; $\tilde{\mu} = 11$).

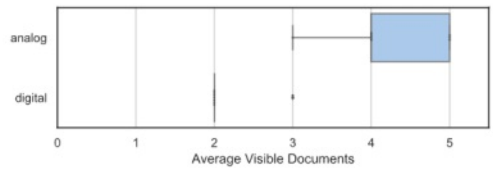


Figure 4.27: Box-plot diagram showing the average number of visible documents of the analog condition ($n = 8$; $\tilde{\mu} = 4$) and the digital one ($n = 9$; $\tilde{\mu} = 2$).

multiple tests have been conducted. Differences in the number of times documents were opened/selected and closed, as well as the average number of documents visible at the same time, are analyzed. In addition, the number of visible documents is set in relation to the total number of "open" documents and any differences are documented. The number of groupings of documents at the end of the task and possible differences also play a crucial role in answering this hypothesis.

A two-tailed t-test showed a large effect, but no statistically significant difference ($t(15) = -1.69$, $p = .112$, $d = -.821$) between the total number of documents (re-)opened/selected by participants of the analog condition ($\mu = 18.8$, $\sigma = 4.95$) and the digital one ($\mu = 23.6$, $\sigma = 6.54$). Figure 4.25 illustrates the distribution of document opening operations (= select events) in each condition.

Analogously, participants of the digital condition did not close significantly more documents than those of the analog condition (Figure 4.26, $\mu_{analog} = 5.13$, $\sigma_{analog} = 6.08$, $\mu_{digital} = 10.1$, $\sigma_{digital} = 5.49$, $t(15) = -1.78$, $p = .096$, $d = -.864$). The large effect in both cases, however, is an indication that people working in analog environments keep their documents in reach rather than putting them back to their original position, while people working in digital environments often close documents after reading.

A Mann-Whitney U test was conducted to compare the average number of visible

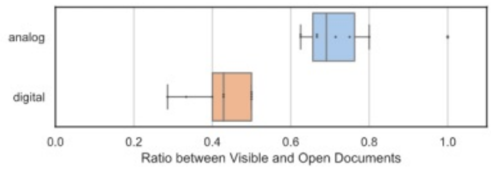


Figure 4.28: Box-plot diagram showing the average ratio of visible and open documents in the analog condition ($n = 8$; $\tilde{\mu} = .69$) and the digital one ($n = 9$; $\tilde{\mu} = .429$).

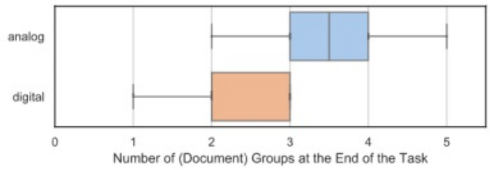


Figure 4.29: Box-plot diagram showing the number of (document) groups at the end of the task of the analog condition ($n = 8$; $\tilde{\mu} = 3.5$) and the digital one ($n = 9$; $\tilde{\mu} = 2$).

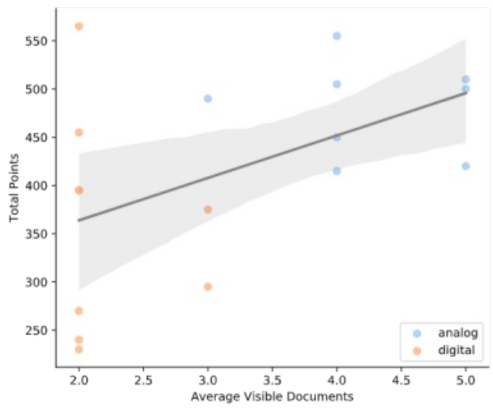


Figure 4.30: Scatterplot showing a strong, positive, linear correlation between the average visible documents and the total points.

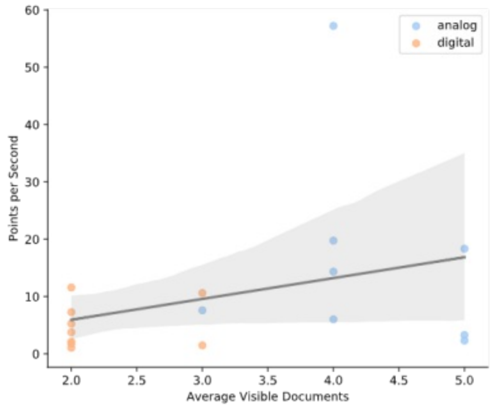


Figure 4.31: Scatterplot showing a strong, positive, linear correlation between the average visible documents and the points per second.

documents in analog and digital conditions. We found a large effect (.817) and a significant difference in the number of visible documents for analog ($\mu = 4.25$, $\sigma = .707$) and digital ($\mu = 2.22$, $\sigma = .441$) conditions; $U = 1$, $z = -3.368$, $p =< .001$. These results suggest that in analog conditions people use the available space to organize their documents in such a way that as many documents as possible are visible. These suggestions can be further strengthened by considering not just the visible documents but rather the ratio between the number of visible and open documents. A two-tailed t-test showed a large effect and a statistically significant difference ($t(15) = 6$, $p =< .001$, $d = 2.91$) between the ratio of the number of visible and open documents of the analog condition ($\mu = .731$, $\sigma = .125$) and the digital one ($\mu = .431$, $\sigma = .08$). In other words, in the analog condition, about 75% of the opened documents were concurrently visible, while in the digital condition only about 45% of them were concurrently visible. The distributions of both the average number of visible documents and the ratio between visible and opened ones can be seen in Figure 4.27 and Figure 4.28 respectively.

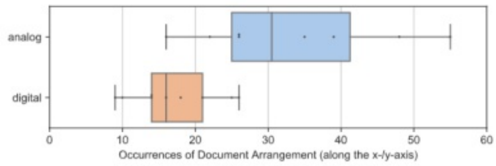


Figure 4.32: Box-plot diagram showing the occurrences of document arrangement (along the x-/y-axis) in the analog condition ($n = 8$; $\tilde{\mu} = 30.5$) and the digital one ($n = 9$; $\tilde{\mu} = 16$).

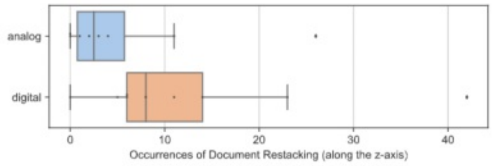


Figure 4.33: Box-plot diagram showing the occurrences of document restacking (along the z-axis) in the analog condition ($n = 8$; $\tilde{\mu} = 2.5$) and the digital one ($n = 9$; $\tilde{\mu} = 8$).

Furthermore, a large effect and a statistically significant difference was observed for the number of groups built at the end of the task (Figure 4.29, $\mu_{analog} = 3.5$, $\sigma_{analog} = .926$, $\mu_{digital} = 2.11$, $\sigma_{digital} = .782$, $t(15) = 3.35$, $p = .004$, $d = 1.63$). The results show that users working in the analog environment built more (document) groups than those working on computers.

A strong positive correlation between the average number of visible documents and the total points, $r = .564$, $p = .018$, Figure 4.30, as well as the points per second, $r = .513$, $p = .035$, Figure 4.31, could be found.

A two-tailed t-test showed a large effect and a statistically significant difference ($t(15) = 3.32$, $p = .005$, $d = 1.61$) between the number of occurrences of document arrangements along the x-/y-axis of the analog condition ($\mu = 33.4$, $\sigma = 13.4$) and the digital one ($\mu = 17$, $\sigma = 6.06$). Besides, a Mann-Whitney U test revealed a small effect (.467) and a nearly statistically significant difference in the number of occurrences of document restacking (along the z-axis) for analog ($\mu = 4$, $\sigma = 4.54$) and digital ($\mu = 12.8$, $\sigma = 12.8$) conditions; $U = 16$, $z = -1.925$, $p = .059$. The results show that participants in the digital condition often need to reorganize overlapping documents (along the z-axis) while in the analog condition documents are more likely to be pushed to the side (along the x- and y-axis) than stacked on top of each other. The distributions of both the number of occurrences of document arrangement and restacking can be seen in Figure 4.32 and Figure 4.33 respectively.

Even though we expected that spatial organization consumes more time during digital multi-document active reading than during analog multi-document active reading (**H5**), the opposite was actually the case. Figure 4.34 shows the total time participants spent on moving and arranging documents. A Mann-Whitney U test showed a moderate effect (.677) and a statistically significant difference ($U = 7$, $z = -2.791$, $p = .004$) between the total duration of document arrangement of the analog condition ($\mu = 2.35$, $\sigma = 1.147$) and the digital one ($\mu = 1.096$, $\sigma = .486$). Figure 4.35, in comparison, shows the mean duration of how long such arrangement operations lasted in both conditions. A small effect ($d = .255$) and no statistically significant difference could be found here. The result of the mean document arrangement duration shows that the operation itself is not more

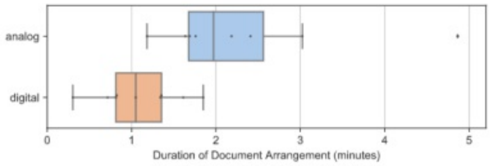


Figure 4.34: Box-plot diagram showing the total duration spent on document arrangement in the analog condition ($n = 8$; $\tilde{\mu} = 1.967$) and the digital one ($n = 9$; $\tilde{\mu} = 1.05$).

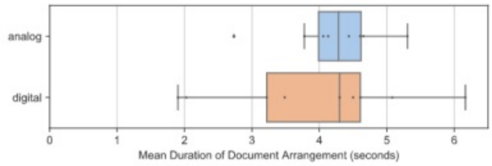


Figure 4.35: Box-plot diagram showing the mean duration spent on document arrangement in the analog condition ($n = 8$; $\tilde{\mu} = 3.96$) and the digital one ($n = 9$; $\tilde{\mu} = 4.3$).

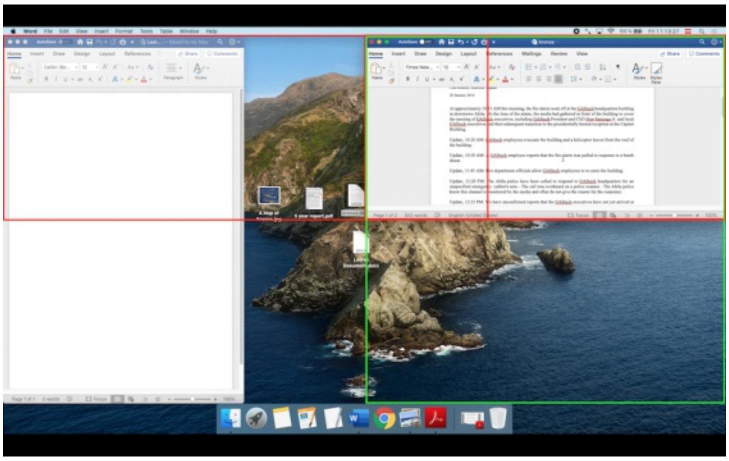


Figure 4.36: Participant D3 trying to move the document to the right half of the screen (= green border), resulting in moving it only to the top (right corner; red borders).

time-consuming. Nevertheless, only about half as much time was spent on arranging documents in the digital condition compared to the analog one. This is not surprisingly as there occurred also about half as much arranging operations in the digital condition compared to the analog one.

We noticed that most of the analog condition participants spread out the documents in their field of vision. A6, who explicitly stated: *"I had everything in view and ready to compare."*, while D6 mentioned: *"I would have laid out all analog documents in front of me, to have a look at them all at once for comparing. In the digital world, on the other hand, I read one document after the other without being able to make cross-references."* A8 was also the only one who explicitly stated that she used the existing color codes of the documents to determine, which of them were relevant to her. For the digital condition participants, we noticed that not much time was invested into the windows' spatial organization. In some cases, windows were only opened, sometimes maximized, and then minimized or closed again. Sometimes, the windows' arrangement did not result in the users' expectations, which was very annoying for the users. D3, for example, tried

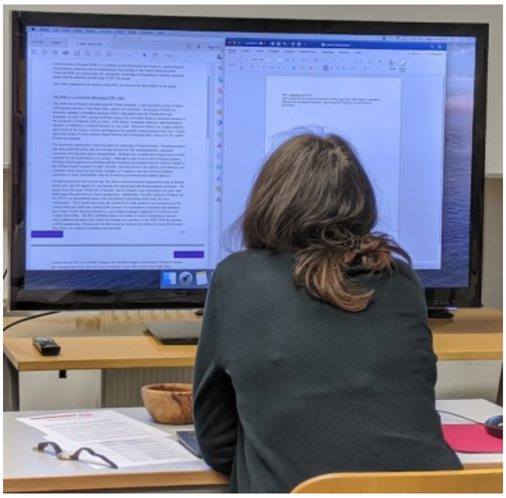


Figure 4.37: Participant D6 while working on the task, showing the divided display space into halves.

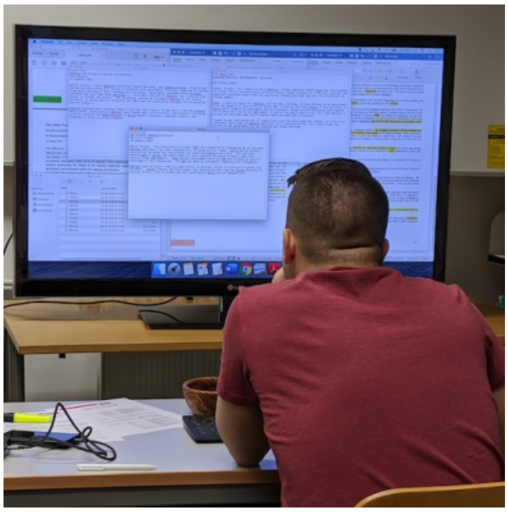


Figure 4.38: Participant D2 while working on the task, showing the divided display space into thirds (in the background with the main articles) and a few small articles (in foreground).

to put one document on the right half of the screen. However, she was only able to move it to the upper right corner (see Figure 4.36). In the end, she closed the document again. Most of the participants have halved the display space and arranged the windows side by side (see Figure 4.37 as an example). A few have even divided the space into three to compare the three primary documents optimally (see Figure 4.38 as an example). For those who took notes, the window or sheet of paper for taking notes was in both conditions in all cases (almost) always visible/in front of them (see Figure 4.39 as an example). A7, D1 and D3 all stated: *"My notes are always in front of me and within reach."*

Six of the total 17 participants stated in the interview that their typical work environment consists of a desktop setting with two to three monitors. The actual display space of such a multi-monitor setting would be quite similar to the provided one by the large display. When asking about the reasons for their preference, participants of the digital condition explained that it is not about the size or space of the display, but rather about the number of different contexts they can establish. Each monitor (or digital workspace) provides a new context and enables more accessible context switches (e.g., between reading and writing).

In order to be able to make statements about the usage of the available space in the digital condition compared to the one available in the analog condition, it is necessary to compare the actual available space for solving the task. In the analog condition, participants were provided with an empty table of 1.6 x 1.2 meters resulting in 1.92m^2.

Figure 4.39: Participant A4 right after working on the task for 45 minutes, showing her notes right in front of her and the articles spread out in reach.

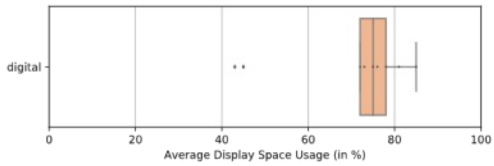


Figure 4.40: Box-plot diagram showing the average usage of the display space in the digital condition ($n = 9$; $\tilde{\mu} = 75$).

The display of the digital condition was 1.2 x 0.69 meters, resulting in a physical space of $0.83m^2$, which is not even half of the space available in the analog condition. The difference becomes even more striking when looking at the number of documents that can be placed next to each other at the same time: In the analog condition, 56 whole A5 documents in landscape format (21 x 14.8 cm) can be spread out on the available workspace. The digital counterpart – the document printed on an A5 piece of paper in plain text format, opened with TextEdit, the text editor of macOS X, with the default window size preferences of 96 characters and 30 lines when using the font Menlo Regular 11pt (resulting in a window size of 684 x 420 pixels on our study setup with a resolution of 1920 x 1080 pixels) – leads to only 4 completely visible, simultaneously arranged windows. Thus, 14 times more documents could be simultaneously visible in the analog condition compared to the digital one.

Figure 4.40 shows that, on average, 69.8% of the available display space was used by participants of the digital condition, with 85% being the maximum and 43% the minimum. For the analog condition, we had no instrument for analyzing the usage of the available physical space. Therefore, we can not compare the actual (display) space usages between conditions to make further statements about space management. However, according to observations, it can be assumed that a relatively large amount of physical space was unused. Most of the time, not even half of the available space provided by the table was completely covered with the documents and utensils.

4.2.4 Strategies & Approaches

When analysing the coded temporal sequences of participants' active reading behaviours (see Figure 4.6 and Figure 4.7), we identified six main approaches that participants of both conditions employed during task solving:

1. **Reading only** (D4, D5, D9)
2. **Writing at the end**: Writing happens at the end of reading (all) documents/materials (D3).
3. **Writing only**: Writing happens while reading without highlighting anything (A8, D1, D6, D7, D8).
4. **Highlighting only**: Reading is interwoven with highlighting without writing anything down (A2, A5).
5. **Writing after highlighting**: Reading and highlighting are interwoven and happen simultaneously, while writing starts at the end of reading (all) documents/materials (A1, A3, A7, D2).
6. **Highlighting and writing**: Reading is interwoven with highlighting and writing (A4, A6).

While in the digital condition only one participant (D2) highlighted text passages, the complete opposite was the case for the analog condition: All participants of the analog condition, except for A8, highlighted text passages. Almost the same amount of participants did not take any notes in both conditions (A2 and A5 of the analog condition; D4, D5 and D9 of the digital condition). Due to the time constraint of 45 minutes, A2 was not able to take notes, but he said he *"might have made some later"*. Participant A5 explained his approach in the interview as follows: *"I never read through the notes again anyway. I always only look at the highlighted passages."* Without further explanation, participant A8 stated that she never highlights text passages and would rather add annotations to the original text than highlights.

In the analog condition, all participants engaged in other reading-related activities such as highlighting and note-taking in addition to reading. In the digital condition, on the other side, there were three participants (D4, D5, and D9) who concentrated purely on the reading task itself, rather than additionally highlighting text passages or taking notes. D5, for example, took just a few notes in the beginning but mostly stopped after the first document. He stated that he *"wanted to look through all materials"* and *"tried to remember things rather than writing them down."* D9 did not see any use in taking

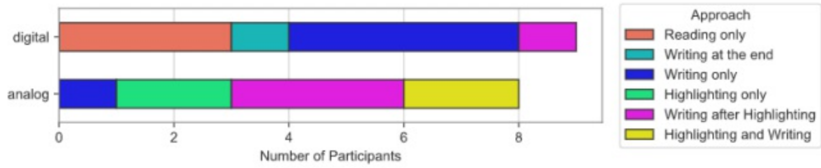


Figure 4.41: Bar-plot showing the usage of the different approaches used for task solving in both conditions.

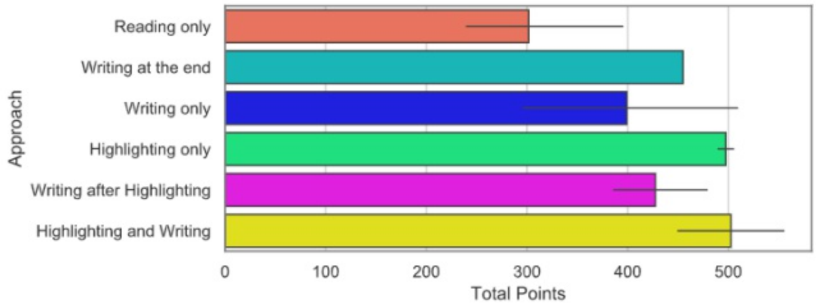


Figure 4.42: Bar-plot showing the achieved total points for each approach used for solving the task including error bars.

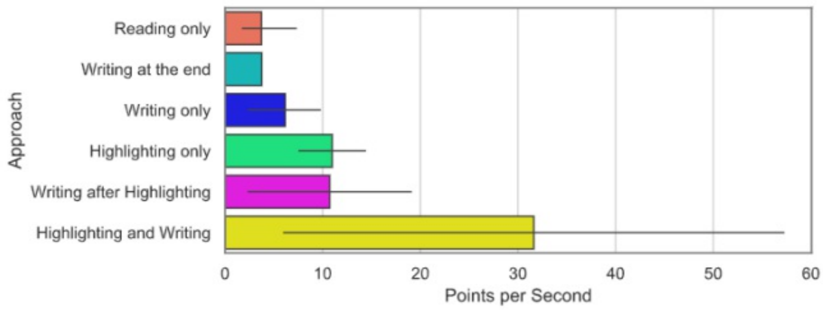


Figure 4.43: Bar-plot showing the achieved points per second for each approach used for solving the task including error bars.

notes at all and does not digitally highlight text passages in general. D4 explained that highlights *"are only a reminder [for him] that something is good again to read"*, for example, when studying for an exam. He further explained that he only takes notes if he sees *"that something is not clear [to him] or might be forgotten [by him]."* Since he *"just needed to explain"*, he did neither highlight nor took notes.

To test the hypothesis that peoples' approaches during analog multi-document active reading differ from the ones used during digital multi-document active reading (**H6**), a X^2-test was conducted. An almost large effect and a nearly statistically significant difference ($X^2(5) = 10.8$, $p = .056$, $V = .796$) were obtained. Figure 4.41 illustrates the

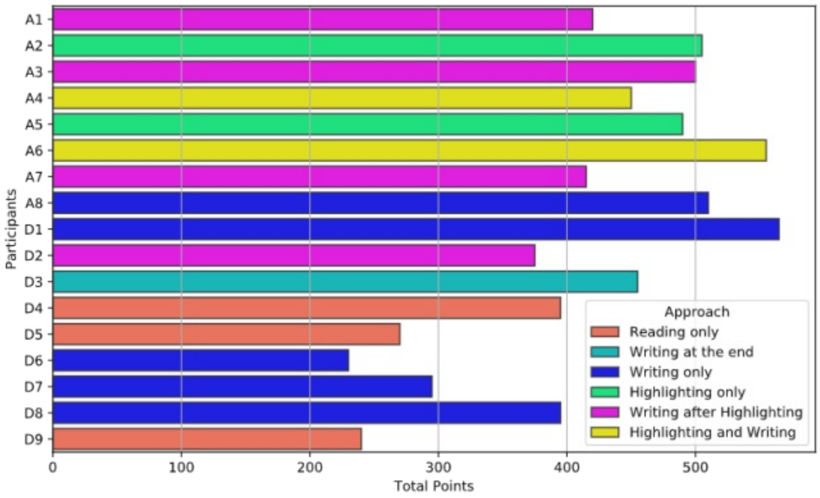


Figure 4.44: Bar-plot showing the achieved points per participant and their used approach.

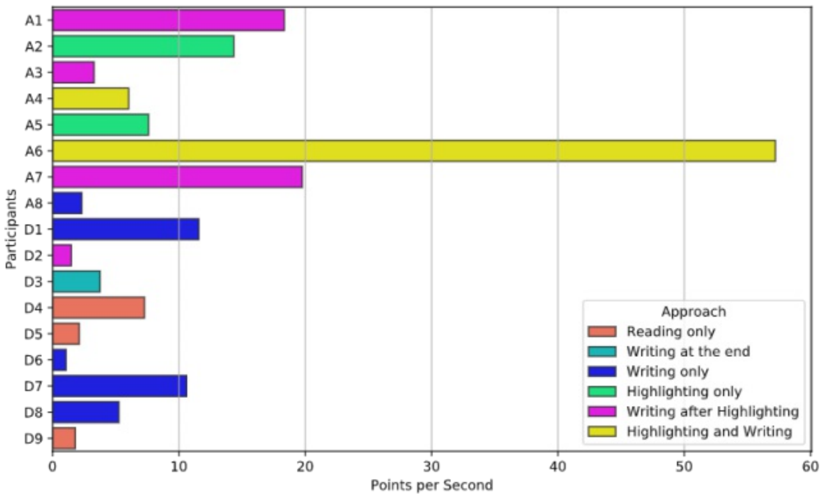


Figure 4.45: Bar-plot showing the participants' achieved points in relation to their response times and their used approach.

usage of all six approaches across conditions.

Even though the Kruskal Wallis tests revealed no statistically significant difference between the approaches used by participants for solving the task and the total achieved points ($H(5) = 5.84$, $p = .322$, $\varepsilon^2 = .365$, Figure 4.42) nor the achieved points in relation to the participants' response times ($H(5) = 4.05$, $p = .542$, $\varepsilon^2 = .253$, Figure 4.43), we found a small effect in both cases.

Figure 4.44 illustrates the achieved points per participant including their used approaches for solving the task, while Figure 4.45 shows the achieved points per participant in

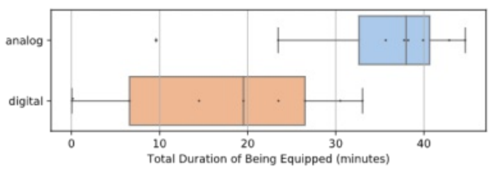


Figure 4.46: Box-plot diagram showing the durations (in minutes) of people being equipped with pens and markers during the task in the analog condition ($n = 8$; $\tilde{\mu} = 38$) and the digital one ($n = 9$; $\tilde{\mu} = 19.517$).

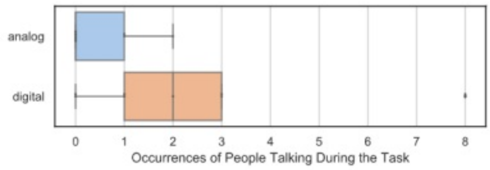


Figure 4.47: Box-plot diagram showing the frequencies of people talking (with themselves) during the task in the analog condition ($n = 8$; $\tilde{\mu} = 0$) and the digital one ($n = 9$; $\tilde{\mu} = 2$).

relation to their response times. When comparing the approaches with the achieved points in relation to the response time, it seems that the fourth (highlighting only) and fifth approach (writing after highlighting) are the most effective ones in the analog condition. A3 is an exception of the fourth approach, as he has switched from his initial approach of only taking notes (third approach) to highlighting due to speed differences. For the digital condition, it seems that the third approach (writing only) is the most effective one. Only D6 scored relatively poorly in the comparison despite the use of the first approach. She has written the most words (over both conditions!), but as she said herself, she *"also usually adopts the wording of the original text or [she] just rewrites a bit"*. Even though D4 did only employ the first approach (reading only), he ranked third in the comparison of the achieved points per second in the digital condition.

Participants of the analog condition were considerably longer equipped with pens and markers during task solving than those of the digital condition. As illustrated in Figure 4.46, equip durations of the analog condition ranged from 9.617 to 44.683 minutes ($\mu = 34.033$, $\sigma = 11.767$), whereas digital conditions were from .046 to 33.033 ($\mu = 17.15$, $\sigma = 12.567$). A two-tailed t-test showed a large effect and a statistically significant difference between the equipment durations of analog and digital groups ($t(15) = 2.85$, $p = .012$, $d = 1.38$).

Moreover, people in the digital condition talked/mumbled more often during task solving than those in the analog condition. A Mann-Whitney U test revealed a small but statistically significant effect (.478) in talking occurrences for analog ($\mu = .5$, $\sigma = .756$) and digital ($\mu = 2.89$, $\sigma = .3.1$) conditions; $U = 15.5$, $z = -1.973$, $p = .044$. Figure 4.47 illustrates the distribution of peoples' talk frequencies in the analog and digital condition.

4.3 Potential Confounding & Random Effects

We explored whether a large difference within the population can be expected. For that, we ran similar tests as previously for the condition, covering the areas of task performance, annotating, highlighting, note-taking, spatial organization and navigation as well as the strategies and approaches. However, we only found two statistically significant differences. Figure 4.48 shows the distribution of words written down by males and females. A Mann-Whitney U test revealed a small, nearly moderate, effect (.49) and a statistically significant difference ($U = 15$, $z = -2.021$, $p = .047$) for the number of words written down by males ($\mu = 75.6$, $\sigma = 103$) and females ($\mu = 199$, $\sigma = 155$), suggesting that females take considerably more notes than males. In Figure 4.49, the distribution of the number of timeline items made in the notes by males and females is shown. A two-tailed t-test showed a large effect and a statistically significant difference for this variable ($t(15) = -2.31$, $p = .035$, $d = -1.12$). We did not find any statistically significant differences between genders and the type of notes nor their approaches.

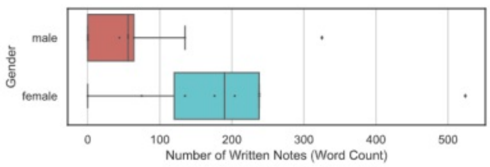


Figure 4.48: Box-plot diagram showing the number of words written down by males ($n = 9$; $\tilde{\mu} = 56$) and females ($n = 8$; $\tilde{\mu} = 190$).

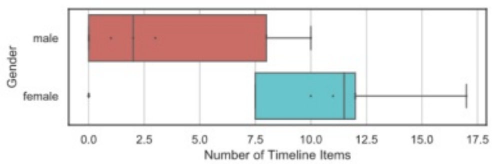


Figure 4.49: Box-plot diagram showing the number of timeline items used in notes by males ($n = 9$; $\tilde{\mu} = 2$) and females ($n = 8$; $\tilde{\mu} = 11.5$).

Even though females seem to write down more information, we could not find any statistically significant difference between the duration of oral summaries by males and females. Figure 4.50 illustrates the duration distributions of their oral reports. A two-tailed t-test did not reveal any statistically significant difference besides a very small effect ($t(15) = .335$, $p = .742$, $d = .163$) between the total points achieved by males ($\mu = 424$, $\sigma = 100$) and females ($\mu = 406$, $\sigma = 117$). The distribution of the total points by males and females is shown in Figure 4.51. We also conducted a Mann-Whitney U test for the points per second for which we, too, did not find any statistically significant difference or effect (Figure 4.52, $\mu_{male} = 8.51$, $\sigma_{male} = 5.75$, $\mu_{female} = 12.2$, $\sigma_{female} = 19.2$, $U = 30$, $z = -.577$, $p = .606$, no effect $= .14$). Those results confirm that there is no difference between genders in relation to task performance.

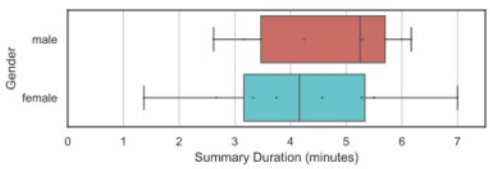


Figure 4.50: Box-plot diagram showing the summary durations (in minutes) by males ($n = 9$; $\tilde{\mu} = 5.25$) and females ($n = 8$; $\tilde{\mu} = 4.16$).

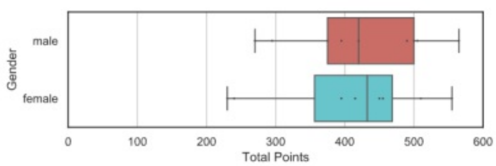


Figure 4.51: Box-plot diagram showing the total points achieved on the task by males ($n = 9$; $\tilde{\mu} = 420$) and females ($n = 8$; $\tilde{\mu} = 433$).

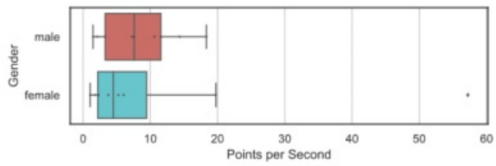


Figure 4.52: Box-plot diagram showing the points per second achieved by males ($n = 9$; $\tilde{\mu} = 7.6$) and females ($n = 8$; $\tilde{\mu} = 4.51$).

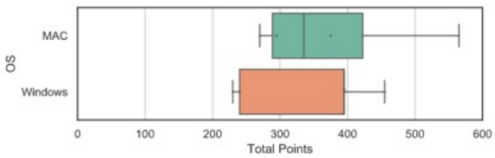


Figure 4.53: Box-plot diagram showing the total achieved points of participants of the digital condition who are usually MacOS users ($n = 4$; $\tilde{\mu} = 49.9$) and Windows users ($n = 5$; $\tilde{\mu} = 121$).

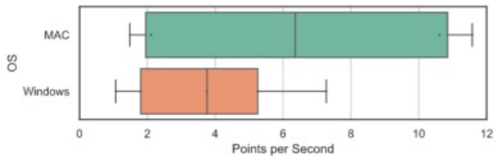


Figure 4.54: Box-plot diagram showing the achieved points in relation to participants' total response times of the digital condition who are usually MacOS users ($n = 4$; $\tilde{\mu} = 5.29$) and Windows users ($n = 5$; $\tilde{\mu} = 3.33$).

Since the digital condition performed significantly lower than the analog condition, we performed some further tests considering only the digital condition participants. We did not find any statistically significant difference between the total points or the points per second of the digital condition participants in relation to their normally used operating system (macOS or Windows). However, the effect size for the analysis of the points per second, depending on the preferred operating system of digital condition participants ($d = .651$), was found to be of a moderate effect according to Cohen's convention ($d = .50$). The distributions of both total achieved points and the points per second divided into the participants' usual operating system can be seen in Figure 4.53 and Figure 4.54 respectively.

We did not find any statistically significant difference in document arrangement frequencies split by the preferred operating system of the digital condition participants. Therefore, it seems that having no prior knowledge about macOS did not influence the performance of arranging documents. However, a two-tailed t test shows a moderate effect of .584 for the total duration spent on arranging documents ($t(7) = .87$, $p = .413$) between Windows users ($\mu = 58.1$, $\sigma = 33.8$) and macOS users ($\mu = 75.4$, $\sigma = 22.84$).

Moreover, we did not find any statistically significant difference between the task performance (total points and points per second) of participants of the digital condition who copy-pasted text and non-copy-pasters, nor was there a statistically significant difference between the number of words in their final notes. Even though we did not find a statistically significant difference ($t(15) = -1.02$, $p = .342$) between the duration

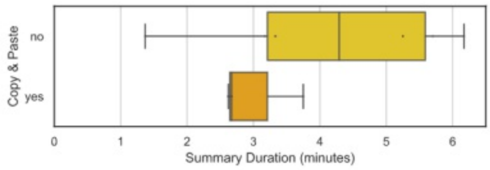


Figure 4.55: Box-plot diagram showing the summary duration (in minutes) of participants in the digital condition that used copy-paste functionalities ($n = 3$; $\tilde{\mu} = 2.67$) and those who did not ($n = 6$; $\tilde{\mu} = 4.29$).

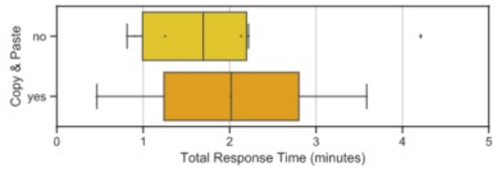


Figure 4.56: Box-plot diagram showing the total response time of participants of the digital condition who copy-pasted text ($n = 3$; $\tilde{\mu} = 2.02$) and non-copy-pasters ($n = 6$; $\tilde{\mu} = 1.7$).

of the summary report of copy-paste users ($\mu = 3.01$, $\sigma = .638$) and "normal typing" users ($\mu = 4.17$, $\sigma = 1.85$), we found a moderate to large effect ($d = -.721$). Figure 4.55 illustrates that participants who used copy-paste functionalities reported fewer (or faster) in their oral summary. Intuitively, participants using copy and paste would also have to search longer for the necessary information to be able to answer the questions. However, no significant difference or effect was found between the total response time of copy-paste users and non-copy-pasters ($t(7) = .1$, $p = .92$, $d = .07$, Figure 4.56). Particularly striking in Figure 4.56 is the outlier of the non-copy-pasters: While D2 did not copy any text portions, he highlighted a great deal of them. Nevertheless, it seems that he did not find the necessary information particularly fast.

We could observe some problems with the digital condition apparatus, which were confirmed by almost half (four out of nine) of the participants in the interview without asking for it. The main problem was the combination of a Mac computer with a Windows keyboard, which caused confusion, especially with keyboard shortcuts, and was not only unintuitive for Windows users but also for established Mac users. Additionally, the use of an older mouse in combination with the large screen caused a small but noticeable delay of the mouse pointer. For Mac users, the setup was also very unusual because of the missing gesture control (via touchpad), and the limitation to one workspace (Mac OS X usually supports the usage of multiple virtual desktops, called workspaces). All these problems were a hurdle for participants of the digital condition when arranging windows.

They could also have influenced their general approach: Due to the discomfort caused by the setup participants might have worked differently than usual. For example, participants might have not highlighted text portions digitally because they did not feel comfortable using the mouse, although they would otherwise have highlighted text portions very well digitally. To a certain extent, all these factors could be at least partially responsible for the little interaction of participants with the documents and as a consequence of the comparatively low scores achieved.

CHAPTER 5

Discussion

Based on a combination of a confirmatory and exploratory, controlled, qualitative user study with a total of 17 participants, users' multi-document active reading behaviors (and performances) in analog and digital environments were compared. The study's evaluation shows that despite the use of a large screen, analog multi-document active reading is still clearly more efficient than digital multi-document active reading. The study reveals differences in behaviors and the used strategies between participants of the analog and digital groups, which can serve as explanations for the performance differences. In general, we observed that people externalized their internal thinking processes more easily when working in the analog world. In the digital world, those externalization processes seemed to happen less frequently: People tried to remember what they have read rather than making externalizations that could potentially have helped them remember. In the few cases where externalization(s) did occur in the digital world, they were mainly in the simplest form of written words and a few highlighted text passages.

5.1 Task Performance

The expectation that analog multi-document active reading is (still) more efficient than digital multi-document active reading using a large display, considering there is better performance with similar time investment (**H1**), was confirmed. Users of the analog condition performed considerably better on the post-test, achieving higher total scores and points in relation to their response times. These results suggest that working with analog tools and materials does have a large effect on task performance and, therefore, is more effective. Specifically, the positive correlation between the condition and the total points suggests that when humans perform active reading activities with analog tools, their total points on questions about the materials in an immediate post-test increase.

Furthermore, the oral reports of users of the analog condition lasted longer, suggesting that people engaging in analog active reading can report more details about the read materials than those engaging in digital active reading. Similar results about performance differences between analog and digital active reading are reported by different studies in the literature [SA17, KSZ18, Cli19, DVAS18, GCR+19, MWB13]. Our study reinforces these results.

5.2 Behaviour Differences and Links to Performance

5.2.1 Annotating & Highlighting

As a possible explanation for the large performance differences, we see differences in participants' typical active reading activities and behaviors. For example, much highlighting was done by the analog group, whereas this functionality was hardly used in the digital condition (**H2**). The reason for the (very) low usage might be mainly the higher effort and the inaccuracy of digital highlighting (see Table 2.1 for comparison). Participants themselves reported that the functionality was unfamiliar, impractical and imprecise. Prior studies [OS97, MBM07, GCR+19] made the same observations, and therefore back up our findings.

Our study also shows that oral reports last longer the more highlighting is used during task solving, which suggests that people remember more details when they highlight materials. The positive effect of the oral reports' length on the task performance further suggests that the number of highlights has an indirect effect on the task outcome. Unlike a previous study [GCR+19], we did not find a correlation between digital highlighting and the task performance, but this may be due to the fact that in our study, only one out of the nine participants of the digital group used this functionality to a greater extent.

5.2.2 Note-Taking

Although differences between analog and digital note-taking during multi-document active reading would have been expected, no significant ones could be found here. Contrary to our hypothesis (**H3**), there is no difference between the number of written words in the analog and the digital group, which indicates that there is no compensation for missing functionalities or (poorly) supported active reading activities (such as highlighting) by taking more notes. Also, the number of words written has no impact on task performance.

Besides, notes were prepared similarly in both conditions, as no discernible differences were found here either, which could also be an indicator for the similar number of written words. The way in which notes are prepared is much more crucial for performance outcomes than the amount of notes. For example, there is evidence that concept mapping is slightly more effective than other constructive activities such as writing summaries and outlines [NA06]. In this study, the timeline is particularly of importance as the task, and the content of the provided materials, are very time- and event-heavy. Thus, it cannot be generalized that timelines always lead to a better result than, for example, summaries, but very well that the type of notes plays a crucial role and is an essential factor for the determination of performance outcomes. In addition, it can also be assumed that the effectiveness of the notes increases the more constructively and visually information is prepared.

In the interviews, participants of both conditions described typing as *"faster"*, notes as *"easier to edit and remove"* and typed words as *"better readable by others (and yourself)"* than handwritten ones. However, almost all of the participants stated that they usually would always have pen and paper ready, as it is easier and faster for them to write unstructured notes per hand. They also have the feeling that they memorize handwritten information to a higher degree, which has already been proven by other studies [SO18, SMR09]. In case of requiring a more illustrative approach (such as concept maps or diagrams), people would also always prefer pen and paper, as sketching and drawing images or diagrams (similar to writing down complex mathematical formulas) requires less effort and is more intuitive on paper with a pen than on computers with mouse and keyboard. All the above mentioned properties have been noted several times in the literature [MBM07, SO18] (see Table 2.1 for comparison). In the case of this study, however, all these characteristics do not seem to have had any particular effect on the task performance, since, there are no differences in either the number of words written or the type of the notes between participants in the analog and digital condition.

5.2.3 Spatial Organization & Navigation

At the outset, it was assumed that the available workspace would be used and organized similarly during analog and digital multi-document active reading if a large display was provided for the digital condition (**H4**). However, this expectation was disproved in several aspects. The results of the study show that in the analog condition, more documents are visible at the same time on average, also when putting them in relation to the average number of currently opened documents. Besides, participants of the digital group restacked documents more often while in the analog condition, documents are moved more along the x- and y-axis. Those results indicate that documents in analog environments are close at hand and visible most at the time, while digital documents often slip "into the background" and have to be refocused/-selected.

Also, during analog multi-document active reading, the available workspace is organized into more groups than during digital multi-document active reading despite using a large display. This behavior is not surprising since grouping and arranging are more effortless in the analog world than the window management in the digital one [MBM07]. The easiest way is to divide the screen into halves, which is also reflected by our results and could be a possible explanation for the low number of groups, the average number of documents visible simultaneously, and the higher amount of document opening operations.

The positive correlation between the average visible documents and the achieved total points further shows that the visibility of documents (and information) is an essential factor for determining the task outcome. A possible explanation for this could be the increased simultaneous access to information [BEK+13, AEN10] when more documents are visible simultaneously, which enables an easier comparison of the contents. Both A1 and D6 confirmed this explanation as they stated that they (would have) had everything in view and ready to compare. Especially in the analog group, where there was a lot of highlighting, the combination of highlighting and the visibility of the documents may also have had an impact on the results, as highlighting creates additional visual structure that helps to quickly retrieve critical points during the reviewing process [LTC16].

Furthermore, it was assumed that spatial organization consumes more time during digital multi-document active reading than during analog multi-document active reading (**H5**). However, this expectation was also disproved. Contrary to our expectation, it has been shown that more time is spent on arranging documents during analog multi-document active reading. One possible explanation could be that, due to the complicated window management, users of the digital condition have limited themselves to the most straightforward and least time-consuming window management operations, such as bisecting windows. Besides, our spatial organization results are more similar to those of single monitor studies. For example, more display space should lead to a higher effort in managing windows and optimizing the window layout to improve participant's workflow [BB09]. The physical space provided by a large display should also bias the user towards working spatially, leading to increased externalization of the user's synthesis [AEN10]. In our case, however, participants in the digital group spent little time arranging and had, on average, only two to a maximum of three documents visible at a time, while some windows were on top of each other.

Even though only A8 explicitly mentioned the usage of the existing color codes – that have been added in order to be able to better distinguish documents in the recordings – of the documents for determining their relevancy, it is likely, that other participants felt the same. The color tags could have helped the participants to find documents faster. Without them, participants might have found it more difficult to search for a specific document or they might have had to developed a similar scheme themselves.

In addition, participants reported that they prefer a multi-monitor setup over a large screen because the different monitors can be used to structure windows (semantically). The literature reinforces this statement, as there it is also stated that most multi-monitor configurations encourage users to think in terms of separate workspaces, usually associated with a distinct application or task [AN12]. However, this result is inconsistent with Bi & Balakrishnan [BB09], where people preferred to work with large displays rather than multi-monitor displays. One possible explanation for these contradicting results could be the lack of a direct comparison between a multi-monitor setup and the large display in our study, causing participants to believe only that they would like the multi-monitor setup better. In contrast, if a direct comparison would have been provided, then the results could possibly be different. However, another possible and much more likely explanation could be that the screen provided was not yet big enough (comparing our screen of around 1.2 x 0.68m with a resolution of 1920 x 1080 pixels to Bi & Balakrishnan's screen of 4.88 x 1.82m with a resolution of 6144 x 2304 pixels [BB09]). This would also explain why our results are similar to those of a "normal" single monitor setup and do not show the benefits of a large high-resolution display.

5.2.4 Strategies & Approaches

To determine what adjustments users are making to their active reading strategies, based on the availability and accessibility of operations and tools, we examined users' strategies and approaches during analog and digital multi-document active reading. The study results confirm the assumption that peoples' approaches during analog and digital multi-document active reading differ (**H6**), despite the usage of a large display. The reasons for the different strategies are possibly due to the characteristics of typical active reading activities. For example, digital highlighting is (more) complicated and imprecise [OS97, GCR+19], while analog highlighting is fast and interwoven in the reading process [OS97]. The same is true for spatial organization and navigation in both conditions (see Table 2.1 for comparison).

Writing plays a vital role in both conditions: On the one hand, it enables a central collection of essential information, and on the other hand, it helps to memorize this information better. Still, writing usually happens at the end (after highlighting) in the analog world while it happens simultaneously during digital reading. It is plausible that the primary reason why analog writing is not as interwoven with reading as highlighting and happens more often at the end of the initial reading is that handwriting takes longer and more effort than grabbing a marker and highlighting text portions [MBM07].

In addition, our study also reveals that the people's chosen approach has an impact on their performance during multi-document active reading. This means that in the end, the differences of the individual activities are decisive for the approach, while the chosen strategy then determines the (performance) outcome of multi-document active reading. For example, in the analog world, it seems to be more efficient to highlight text passages only or to write notes afterward. Doing all three activities (reading, highlighting, and writing; approach six) at once might be too much overload for people like A4 (except for A6, who seemed to switch between these three activities effortlessly, see Figure 4.43 for comparison) while taking notes only in addition to reading led to inferior task performance. In the digital condition, writing while reading seemed to be the best approach, possibly since it involves the highest engagement with the material.

Despite using this superior approach (writing only) for digital multi-document active reading, D6 scored considerably poorly on the post-questionnaire. She mentioned that she usually adopts the wording of the original text or rewrites it a bit, indicating that copy and paste information and transcribed information will not be memorized as well as typed information in own words. This finding is confirmed by the literature, as usually more verbatim notes are less effective [MO14, Kob05]. Furthermore, D6 stated that she did not highlight text passages as she has written down the information anyway, indicating the double effort for digital highlighting and writing. A possible explanation for why D4 ranked third in comparison in the digital condition even though he only read through the materials might be his advantage of being an English native speaker and understanding the contents of the materials more easily.

The significant difference in the equipment duration between analog and digital multi-document active reading suggests that people working with analog tools can engage more quickly (and longer) in reading-related activities and enable them, for example, to read and write/highlight simultaneously. Moreover, possible explanations for the higher talking frequencies during digital multi-document active reading compared to the analog one might be problems and annoyances with the digital setup (macOS operating system with a Windows keyboard), which led to expressions about their displeasure.

5.3 Potential Confounding & Random Effects

Hardly any significant differences were found with respect to potential confounding or random effects. The only noteworthy finding is that although females write slightly more notes than males, they do not perform better on post-questionnaires. There are also no differences in the oral report's length, indicating that females do not remember more details than males. Overall, there are no differences in task performance between males and females.

A detailed analysis of the digital group further shows that people who copy-pasted reported less in the oral report, or at least, they did not take as long to do so, which could indicate that they have remembered less information. Prior knowledge of the operating system (macOS) does not appear to have affected performance outcomes. MacOs users, however, spent slightly more time arranging documents. One explanation for this could be that they were not familiar with the window management and any related keyboard shortcuts due to the Windows keyboard and therefore took longer. Another explanation might be that they put more time and effort into arranging the documents, while Windows users left the documents as they were after opening them and did not bother with them any further.

5.4 Limitations

Due to the nature of the research questions and the topic itself, this research was based on qualitative research methods. The reader is cautioned that the findings reported here are qualitative to a large extent. Qualitative data has been quantified in a coding process. The findings should be viewed as working hypotheses, subject to quantitative validation. In addition, the insufficient number of participants for statistical measurements and the large number of exploratory comparisons create a risk of Type I (the rejection of a true null hypothesis; false positive) and Type II (the acceptance of a false null hypothesis; false negative) errors. To avoid a Type I error, the α-level would have to be reduced from .05 to at least .001 when testing 60 or more measurements (Bonferroni's adjustment). To avoid a Type II error in the number of words written, for example, one would need 56 participants per group, resulting in a total of 112 participants (Power Analysis), which is more than six times the number of participants used in this study.

Due to time constraints, we also opted for a between-subject design for the study. Although previous work indicated that individual variability would be high, given the diverse range of strategies for sensemaking and externalization behaviors [RHRH+19], a within-subject design was always out of the question. In a within-subject design, participants would have had very little time to solve the task or had to spend several hours completing the study tasks of both conditions. The between-subject design enabled them to work almost one hour intensively on the task, which in some cases was still too little time to really do everything as they would do it at their usual workstation (and without time limits).

Furthermore, the target group is also relatively limited (and not truly random). Participants in the study are all roughly the same age and have a similar background. For example, no students or adults who have been active in the workforce for some time and whose training dates back several years were included in the study. Moreover, our methods of measurements were limited: We had no instrument for analyzing the users' eye movements (such as eye tracking). Therefore, we could not determine the actual document in focus of participants to be able to explain more precisely any of their behavior.

Moreover, any kind of "switching costs" were generally not considered in this study. The documents to be read were already printed out for the analog group, the workstation was empty and only equipped with the most necessary materials. If the analog group had also started the task with digital materials, the overall task performance would probably have been much worse. In addition, participants in the digital group were only equipped with materials that were already available digitally. Again, this avoided any "switching cost" from analog to digital (e.g., digitizing a text that was only analogously available using an OCR scan to enable digital search), although this type of "switching cost" is probably less common in the practice of knowledge workers.

CHAPTER 6

Conclusion

This chapter summarizes the presented work and concludes the findings of the conducted user study. Additionally, design implications for (creating better) user interfaces of (future) applications to support and improve digital (multi-document) active reading processes are given based on these findings. Finally, an outlook of possible future research is given.

6.1 Summary

The aim of this diploma thesis is a direct comparison of users' classic active reading behaviors (annotating, highlighting, and note-taking) and their spatial organization approaches when confronted with multiple documents in analog and digital environments. Despite the improvements made in the digital domain, knowledge workers still often switch between analog and digital tools while performing their (sensemaking) tasks. A better understanding of possible differences between active reading behaviors and approaches in these two environments should help in supporting and improving active reading processes even more in the digital context. By doing so, it would no longer be necessary for knowledge workers to have to accept the "switching costs" between digital and analog materials for carrying out their tasks carefully and successfully.

For the comparison, we conducted a controlled, qualitative user study with a total of 17 participants. Eight of them were provided with analog materials and tools, while the remaining nine were only allowed to work with digital materials and tools on a computer with a large display provided by us. Users then had to work on a text analysis task using the materials and tools provided. They had 45 minutes to complete the task. Afterward, they gave a short oral summary of the events they considered most important. The study was completed with a qualitative, semi-structured, open-ended interview in which participants commented on their strategies and approaches for solving the task.

To assess the effectiveness of both analog and digital multi-document active reading, all participants were required to answer six questions about the provided materials' contents. The study results support previous findings of analog (multi-document) active reading being superior to digital (multi-document) active reading. Participants of the analog condition performed considerably better on the questionnaire about the contents of the materials, achieving higher total scores and points in relation to their response times while investing a similar amount of time in task solving. The use of the large display in the digital condition does not seem to be able to compensate for the advance of analog active reading.

As a possible explanation for the large differences in performance, we see differences in participants' typical active reading activities and behaviors. In the analog environment, a lot of text has been highlighted using markers, whereas this activity was hardly used in the digital environment. Despite the lack of text highlighting, participants of the digital condition did not take more notes than those of the analog one. Also, participants structured their notes similarly in both conditions. Differences were particularly noticeable in the area of spatial organization: While in the analog condition, documents are kept at reach and insight, digital documents often slip into the background and have to be brought to the foreground again. This can be seen in the average number of visible documents (also in relation to the number of simultaneously opened documents), which is significantly higher during analog multi-document active reading than during digital multi-document active reading using a large display. The positive correlation between the visibility of the documents and the points achieved (see Figure 4.30) is particularly noteworthy at this point, too. Besides, during analog multi-document active reading, participants spend more time arranging and structuring their documents. In contrast, during digital multi-document active reading, participants resort to the most straightforward and quickest operations – such as splitting the available display space into halves – and put minimal effort into a spatial organization.

The study results also show very well that depending on the availability and accessibility of operations and tools in the different environments, users adapt their strategies for (multi-document) active reading. While for participants in the analog condition highlighting is an essential part of their (multi-document) active reading approach for solving the task, writing notes is of enormous importance for participants in the digital condition. Although participants of the digital condition did not take more or detailed notes compared to those of the analog one, note-taking is the most common (and almost only) reading-related activity employed by users of the digital condition. In some cases (three out of nine), participants in the digital condition also chose just to read and memorize the contents without using any additional or more complex strategy.

Results and observations of the exploratory analysis were discussed in detail in Chapter 5. In summary, these lead to the following (new) hypotheses, from which appropriate follow-up studies can be derived, which are discussed in more detail in Section 6.3:

- Highlighting during (analog) active reading leads to a better performance.

- The creation of a timeline has a positive impact on (active reading) outcomes and reading comprehension when analyzing events.
- More space, and therefore a higher number of visible documents at the same time, leads to a better performance, but only when text portions can be highlighted.
- The chosen approach (and thus the sequence of typical active reading activities) influences the (multi-document) active reading performance of participants.

This work's main contributions are the observations and findings from the user study, which provide a direct comparison of participants' active reading behaviors with multiple documents in analog and digital environments. Learning from these findings, design implications for user interfaces to tackle (and solve) identified (multi-document) active reading issues on computers are defined in Section 6.2. They are a further contribution of this work on the topic of (multi-document) active reading.

6.2 Design Implications

Learning from (multi-document) active reading in the analog world, we found that improvements for digital multi-document active reading are necessary, especially in the area of spatial organization, but also for note-taking and highlighting. The window management needs to be better supported, especially for single monitor setups and large displays, such as the one used in this study. Measures should also be taken to ensure that the users' digital note-taking and highlighting experiences are similar to the analog experiences and that they show similar benefits.

(Pre-defined) Window Management Layouts: In digital environments, little time is usually spent on arranging windows. Windows are mainly maximized or halved to make comparisons between two documents possible. Anything beyond that becomes tedious for users, which is why the idea of (predefined) "window layouts" came to life. Users should be provided with (more complex) layouts, allowing them to easily and quickly arrange a bunch of windows. Users should also have the possibility to create such layouts themselves (similar to a customizable grid). The respective layout areas should also be nameable and assignable with keyboard shortcuts. Thus individual windows can be assigned to a certain screen region defined by the layout. The predefined layouts are especially advantageous when frequently working with many windows and when similar complex window arrangements are needed several times. The task management system *Scalable Fabric* of Robertson et al. [RHC+04] already comprises similar concepts. Agarawala and Balakrishnan also developed a prototype, called *BumpTop* [AB06a], that supports a new style of desktop organization. The idea is similar to a website builder or a photo book software: Users define placeholders/layouts and then insert corresponding (not yet used) components at the placeholders provided or automatically arrange a selection of components based on the defined layout.

Window Grouping/Categorization: Often, several windows of different applications are open about one specific topic or for solving a specific task. Especially when working with only one screen, where the available display space is running out quite quickly, a simple mechanism is needed to bring relevant windows back to the foreground without searching for or refocusing every single one of them. Therefore it would be useful to be able to categorize or group windows of different applications, similarly to GroupBar [SBR+03] or the "Window Group" prototype of Lischke et. al. [LMH+17]. By doing so, users would be able to apply window management operations to all elements of these specific window groups, such as (re-)focusing, minimizing but also moving the windows altogether. This can be compared to having different piles of documents and moving all at once by grabbing the whole pile in analog environments.

Quick writing mode: In analog environments, users are always ready to take notes when they have a pen in hand and a blank sheet of paper in reach. The corresponding document for taking notes must always be focused in digital environments before thoughts can be written down. This costs time on the one hand and (a lot of) space on the other hand if trying to avoid this time-cost by keeping the note-taking window always in the foreground (and focused). A variant would be to always open documents (and applications, in general) in a "read-only" mode so that as soon as typing occurs, those pieces of information are automatically written to a (new empty or a previously selected) note document. While writing, the writing document could be automatically displayed, for example, at the bottom of the screen and brought to the foreground. When writing is finished, this window should be automatically hidden to display the actual information/documents fully. Another variant would be to activate (and exit) the "writing mode" using a keyboard shortcut, which automatically gives the note document focus and allows beginning to write without interruptions. Jourknow [BVKKS08], for example, offers already similar functionality to the one proposed here.

Quick highlighting mode: Similar to writing, highlighting requires a mechanism to enter the "highlighting mode" quickly and without interruptions. Currently, switching to the "highlighting mode" requires several clicks, often in places far from the relevant information. For example, Adobe Acrobat Reader offers the possibility to switch to the "highlighting mode" via keyboard shortcuts. However, this functionality is hidden in the settings, has to be activated first, and does not provide any information about the usage of keyboard shortcuts for specific actions. These keyboard shortcut instructions should be clearly labeled at the corresponding action and defined uniformly across applications. In addition, the highlighting itself should work much more roughly and similarly like on paper (such as ScreenCrayons [OTF04] or XLibris [PSG98]). Currently, digital highlighting is based on individual letters and words and requires very precise input. Otherwise it often leads to inaccuracies and annoying errors.

Interestingly, the aforementioned prototypes and solutions have not really become established. One possible explanation is that haptics and optics play a decisive role for active reading and, therefore, should be kept in mind even more when implementing the guidelines. In general, a solution is needed that supports highlighting as well as note taking and spatial organization for commodity monitors.

6.3 Follow-Up Studies

A list of potential specialized confirmative follow-up studies has been established during the execution and analysis of this thesis' user study. Those follow-up studies should investigate potentially interesting partial aspects or questions of (multi-document) active reading that could not be answered by our study design.

Assessing the influence of highlighting: Especially in the analog world, highlighting plays an important part while multi-document active reading. Therefore, the influence of highlighting needs to be further analyzed. This could be done with a study comparing participants multi-document active reading outcomes in an analog setting, where one group of participants is allowed to highlight text passages and the other one is not.

Assessing the influence of timelines for the analysis of events: In this study, the creation of a timeline was particularly of importance as the task and the materials' content were very time- and event-heavy. As it cannot be generalized from those results, that timelines always lead to a better result during multi-document active reading, the influence and role of timelines for the analysis of events needs to be assessed in more detail.

Assessing the influence of visibility in combination with highlighting: The results of this study show that the visibility of documents had a decisive influence on the outcome of participants. In addition, our results suggest that this influence is amplified when participants additionally highlighted text passages, as they add visually distinguishable structure to documents and catch ones attention to important parts. Accordingly, the influence of the combination of document visibility and highlighting should be assessed in more detail in a follow-up study. For this purpose, a study with a two-factorial design could be conducted: providing much/-less space (and thus more/less documents visible at the same time) with/without the possibility of highlighting.

Assessing the influence of multi-document active reading strategies: In the course of this study, we identified six different strategies that participants employed during multi-document active reading to solve the given task. In order to find out how (and to what extend) these strategies really influence the result, a study needs to be conducted that compares the strategies of the users in the same condition or alternatively dictates a specific strategy to users.

Detailed analysis of multi-document active reading behaviors via eye-tracking: In order to analyze not only how many documents are visible on average during multi-document active reading, but also which of them were viewed or in focus (e.g., own notes, original documents), it requires a thorough investigation and recording of the behavior using appropriate measurement tools, such as eye-tracking technologies. The use of eye-tracking methods would also allow for better categorization of when participants are engaged in reading or other hard-to-recognize activities (e.g., searching for something or checking the time).

Assessing the influence of even more digital space: Since the large screen used in this study did not show the expected success of large high-resolution displays, it is still necessary to evaluate the influence of even more digital space on digital multi-document active reading behaviors and approaches. Comparing analog multi-document active reading to digital using an even larger screen of a truly comparable size (e.g., there is space for three documents next to each other and two on top of each other in both conditions) should serve to answer whether or not more digital space yields to more similar behaviors in both environments.

Assessing the influence of the window management: Windows enables users to arrange windows easily via "drag and drop" (e.g., to halve or quarter them) and simple keyboard shortcuts (e.g., Windows and left arrow). Additional tools such as "Rectangle" or "Hammerspoon" are required to accomplish similar window management on macOS. In return, macOS offers excellent gesture control via the touchpad for navigating between multiple windows and workspaces (= multiple virtual desktops). Future research could compare the window management (tools) of both operating systems to investigate whether they really provide such different window management strategies. In addition, the standard window management systems provided by operating systems should be compared to the proposed custom solutions concerning the window management and grouping of the previous subsection (Subsection 6.2).

Assessing the influence of virtual desktops: Since operating systems now also support virtual desktops very well, it would also be interesting to evaluate their influence on multi-document active reading. Based on the results of Ringel's study [Rin03], the assumption is that users behave differently in such a setup as in a multi-monitor setup. A comparison of these approaches (large display, single monitor with multiple virtual desktops, and multi-monitor setup) in relation to users' behaviors, task performances, and personal preferences would provide additional insight into the spatial organization during multi-document active reading. It further provides insights onto whether the task of active reading is better suited for either virtual desktop or multiple monitor environments.

Assessing gender influences on note-taking: Further research on (possible) gender influences on typical active reading activities, especially note-taking, could be done outside the field of computer science. For example, a comparison of note-taking strategies between males and females could help better understand both genders' needs and ways of thinking in this area.

List of Figures

List of Tables

Bibliography

[AB06a] Anand Agarawala and Ravin Balakrishnan. Keepin' it real: Pushing the desktop metaphor with physics, piles and the pen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 1283–1292. Association for Computing Machinery, 2006.

[AB06b] Anand Agarawala and Ravin Balakrishnan. Keepin' it real: Pushing the desktop metaphor with physics, piles and the pen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 1283–1292. Association for Computing Machinery, 2006.

[AD72] Mortimer J. Adler and Charles Van Doren. *How to read a book: The classic guide to intelligent reading*. Simon and Schuster, 1972.

[AEN10] Christopher Andrews, Alex Endert, and Chris North. Space to think: Large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 55–64. Association for Computing Machinery, 2010.

[AGH+98] Annette Adler, Anuj Gujar, Beverly L. Harrison, Kenton O'Hara, and Abigail Sellen. A diary study of work-related reading: Design implications for digital reading devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, page 241–248. ACM Press/Addison-Wesley Publishing Co., 1998.

[Amb20] Amberscript. AmberScript. `https://www.amberscript.com/de/`, 2020. Accessed: 2020-08-15.

[AN12] C. Andrews and C. North. Analyst's workspace: An embodied sensemaking environment for large, high-resolution displays. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 123–131, Oct 2012.

[AN13] C. Andrews and C. North. The impact of physical navigation on spatial organization for sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2207–2216, Dec 2013.

[Appa] Leopard "Stacks" Implement Ages-Old GUI Concept "Piles", author=Pete Mortensen, year=2007, howpublished = https://www.wired.com/2007/06/leopard-stacks-/, note = Accessed: 2020-08-19.

[Appb] The secret history of Apple's Stacks, author=Andrew Orlowski, year=2018, howpublished = https://www.theregister.com/2018/10/05/the_secret_history_of_apples_stacks/, note = Accessed: 2020-08-19.

[BB09] Xiaojun Bi and Ravin Balakrishnan. Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 1005–1014. Association for Computing Machinery, 2009.

[BEK+13] Lauren Bradel, Alex Endert, Kristen Koch, Christopher Andrews, and Chris North. Large high resolution displays for co-located collaborative sensemaking: Display usage and territoriality. *International Journal of Human-Computer Studies*, 71(11):1078 – 1088, 2013.

[BGLG11] Joeran Beel, Bela Gipp, Stefan Langer, and Marcel Genzmehr. Docear: An academic literature suite for searching, organizing and creating academic literature. In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, JCDL '11, page 465–466. Association for Computing Machinery, 2011.

[BIC06] Eric A. Bier, Edward W. Ishak, and Ed Chi. Entity workspace: An evidence file that aids memory, inference, and reading. In Sharad Mehrotra, Daniel D. Zeng, Hsinchun Chen, Bhavani Thuraisingham, and Fei-Yue Wang, editors, *Intelligence and Security Informatics*, pages 466–472, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[BJ05] Olha Bondarenko and Ruud Janssen. Documents at hand: Learning from paper to improve digital technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, page 121–130. Association for Computing Machinery, 2005.

[BL11] Joeran Beel and Stefan Langer. An exploratory analysis of mind maps. In *Proceedings of the 11th ACM Symposium on Document Engineering*, DocEng '11, page 81–84. Association for Computing Machinery, 2011.

[BPP+19] S. Briggs, M. Perrone, M. Peveler, J. Drozdal, L. Balagyozyan, and H. Su. Multimodal, multiuser immersive brainstorming and scenario planning for intelligence analysis. In *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–4, Nov 2019.

[BVKKS08] Michael Bernstein, Max Van Kleek, David Karger, and M. C. Schraefel. Information scraps: How and why information eludes our personal information management tools. *ACM Trans. Inf. Syst.*, 26(4), October 2008.

[BYEA14] Gal Ben-Yehudah and Yoram Eshet-Alkalai. The influence of text annotation tools on print and digital reading comprehension. pages 28–35, 01 2014.

[Car83] Carol A. Carrier. Notetaking research. *Journal of instructional development*, 6(3):19, 1983.

[CHC+04] Alberto J Cañas, Greg Hill, Roger Carff, Niranjan Suri, James Lott, Gloria Gómez, Thomas C Eskridge, Mario Arroyo, and Rodrigo Carvajal. Cmaptools: A knowledge modeling and sharing environment. 2004.

[Cli19] Virginia Clinton. Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading*, 42(2):288–325, 2019.

[Coh88] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Routledge, 1988.

[Com] AbiSource Community. AbiWord - Word Processing for Everyone. `https://www.abisource.com`. Accessed: 2020-07-11.

[Com15] Visual Analytics Community. VAST Challenge 2014. `http://www.vacommunity.org/VAST+Challenge+2014`, Apr 2015. Accessed: 2020-05-12.

[CT79] Carol A. Carrier and Amy Titus. The effects of notetaking: A review of studies. *Contemporary Educational Psychology*, 4(4):299 – 314, 1979.

[Cyr97] Thomas E. Cyrs. Visual thinking: Let them see what you are saying. *New Directions for Teaching and Learning*, 1997(71):27–32, 12 1997.

[ddpG20] "dr. dresing & pehl GmbH". f4x Spracherkennung für Interviews. `https://www.audiotranskription.de/f4x`, 2020. Accessed: 2020-08-15.

[DVAS18] Pablo Delgado, Cristina Vargas, Rakefet Ackerman, and Ladislao Salmerón. Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review*, 25:23 – 38, 2018.

[Epp06] Martin J Eppler. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information Visualization*, 5(3):202–210, 2006.

[FG16] Olivier Friard and Marco Gamba. Boris: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11):1325–1330, 2016.

[FH73] Judith L Fisher and Mary B Harris. Effect of note taking and review on recall. *Journal of Educational Psychology*, 65(3):321, 1973.

[FHH02] Paul Farrand, Fearzana Hussain, and Enid Hennessy. The efficacy of the 'mind map' study technique. *Medical Education*, 36(5):426–431, 2002.

[FMS] Inc. FMS. The Next Generation of Bi Data Visualization. `http://www.fmsasg.com/Products/SentinelVisualizer/`. Accessed: 2020-07-11.

[Gau15] Wilhelm Gaus. Interpretation of statistical significance - exploratory versus confirmative testing in clinical trials, epidemiological studies, meta-analyses and toxicological screening (using ginkgo biloba as an example). *Clinical & Experimental Pharmacology*, 05, 01 2015.

[GCR+19] Amanda P. Goodwin, Sun-Joo Cho, Dan Reynolds, Katherine Brady, and Jorge Salas. Digital versus paper reading processes and links to comprehension for middle school students. *American Educational Research Journal*, Dec 2019.

[GWLS17] Thomas Geymayer, Manuela Waldner, Alexander Lex, and Dieter Schmalstieg. How sensemaking tools influence display space usage. In *Proceedings of the EuroVis Workshop on Visual Analytics*, EuroVA '17, page 7–11. Eurographics Association, 2017.

[Ham20] Hammerspoon. Hammerspoon. `https://www.hammerspoon.org`, 2020. Accessed: 2020-08-12.

[HBPB12] Ken Hinckley, Xiaojun Bi, Michel Pahud, and Bill Buxton. Informal information gathering techniques for active reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 1893–1896. Association for Computing Machinery, 2012.

[HPW+12] Matthew Hong, Anne Marie Piper, Nadir Weibel, Simon Olberding, and James Hollan. Microanalysis of active reading behavior to inform design of interactive desktop workspaces. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '12, page 215–224. Association for Computing Machinery, 2012.

[HRL17] Jinghui (Jove) Hou, Justin Rashid, and Kwan Lee. Cognitive map or medium materiality? reading on paper and screen. *Computers in Human Behavior*, 67:84–94, 02 2017.

[HSM+04] Dugald Ralph Hutchings, Greg Smith, Brian Meyers, Mary Czerwinski, and George Robertson. Display space usage and window management operation comparisons between single monitor and multiple monitor users. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, page 32–39. Association for Computing Machinery, 2004.

[IBM] IBM. IBM Security i2 Analyst's Notebook. https://www.ibm.com/at-de/products/i2-analysts-notebook. Accessed: 2020-07-11.

[IF09] Petra Isenberg and Danyel Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. *Computer Graphics Forum*, 28(3):1031–1038, 2009.

[IFP+12] Petra Isenberg, Danyel Fisher, Sharoda A. Paul, Meredith Ringel Morris, Kori Inkpen, and Mary Czerwinski. Co-located collaborative visual analytics around a tabletop display. *IEEE Transactions on Visualization and Computer Graphics*, 18(5):689–702, May 2012.

[ITC08] Petra Isenberg, Anthony Tang, and Sheelagh Carpendale. An exploratory study of visual information analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 1217–1226. Association for Computing Machinery, 2008.

[J.01] Brown Gary J. Beyond print: reading digitally. 19(4):390–399, 2020/05/27 2001.

[JH14] Mikkel R. Jakobsen and Kasper HornbÆk. Up close and personal: Collaborative work on a high-resolution multitouch wall display. *ACM Trans. Comput.-Hum. Interact.*, 21(2), February 2014.

[JLI17] Renée S. Jansen, Daniel Lakens, and Wijnand A. IJsselsteijn. An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review*, 22:223 – 233, 2017.

[Jon07] William Jones. Personal information management. *Annual Review of Information Science and Technology*, 41(1):453–504, 2007.

[JPM99] Mento Anthony J., Martinelli Patrick, and Jones Raymond M. Mind mapping in executive education: applications and outcomes. 18(4):390–416, 1999.

[KHA00] Ian M. Kinchin, David B. Hay, and Alan Adams. How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, 42(1):43–57, 2000.

[Kir95] David Kirsh. The intelligent use of space. *Artificial Intelligence*, pages 31–68, 02 1995.

[Kir10] David Kirsh. Thinking with external representations. *AI & SOCIETY*, 25(4):441–454, 2010.

[Kob05] Keiichi Kobayashi. What limits the encoding effect of note-taking? a meta-analytic examination. *Contemporary Educational Psychology*, 30(2):242 – 262, 2005.

[KS11] Y. Kang and J. Stasko. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 21–30, Oct 2011.

[KSZ18] Yiren Kong, Young Sik Seo, and Ling Zhai. Comparison of reading performance on screen and on paper: A meta-analysis. *Computers and Education*, 123:138 – 149, 2018.

[Lad80] Harold S. Ladas. Note taking on lectures: An information-processing approach. *Educational Psychologist*, 15(1):44–53, 1980.

[LBG+08] Marieke Longcamp, Céline Boucard, Jean-Claude Gilhodes, Jean-Luc Anton, Muriel Roth, Bruno Nazarian, and Jean-Luc Velay. Learning through hand- or typewriting influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. *Journal of Cognitive Neuroscience*, 20(5):802–815, 2008.

[LGHH08] Chunyuan Liao, François Guimbretière, Ken Hinckley, and Jim Hollan. Papiercraft: A gesture-based command system for interactive paper. *ACM Trans. Comput.-Hum. Interact.*, 14(4), January 2008.

[LMH+17] Lars Lischke, Sven Mayer, Jan Hoffmann, Philipp Kratzer, Stephan Roth, Katrin Wolf, and Paweł Woundefinedniak. Interaction techniques for window management on large high-resolution displays. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, MUM '17, page 241–247. Association for Computing Machinery, 2017.

[LTC16] Liang-Yi Li, Shu-Ting Tseng, and Gwo-Dong Chen. Effect of hypertext highlighting on browsing, reading, and navigational performance. *Computers in Human Behavior*, 54:318 – 325, 2016.

[Mal83] Thomas W. Malone. How do people organize their desks? implications for the design of office information systems. *ACM Trans. Inf. Syst.*, 1(1):99–112, January 1983.

[Man20] Miro Mannino. Miro Windows Manager. `https://github.com/miromannino/miro-windows-manager`, apr 2020. Accessed: 2020-08-12.

[MBM07] M. R. Morris, A. J. B. Brush, and B. R. Meyers. Reading revisited: Evaluating the usability of digital display surfaces for active reading tasks. In *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)*, pages 79–86, Oct 2007.

[MDR19] Kayla Morehead, John Dunlosky, and Katherine A. Rawson. How much mightier is the pen than the keyboard for note-taking? a replication and extension of mueller and oppenheimer (2014). *Educational Psychology Review*, 31(3):753–780, 2019.

[Mei20a] MeisterLabs. jamovi (Version 1.2) [Computer Software]. `https://www.jamovi.org`, 2020. Accessed: 2020-08-24.

[Mei20b] MeisterLabs. MindMeister - Create Your Mind Map Online and Collaborate. `https://www.mindmeister.com`, 2020. Accessed: 2020-05-30.

[MN12] Fabrice Matulic and Moira C. Norrie. Supporting active reading on pen and touch-operated tabletops. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, page 612–619. Association for Computing Machinery, 2012.

[MO14] Pam A. Mueller and Daniel M. Oppenheimer. The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science*, 25(6):1159–1168, 2014.

[MSW92] Richard Mander, Gitta Salomon, and Yin Yin Wong. A "pile" metaphor for supporting casual organization of information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, page 627–634. Association for Computing Machinery, 1992.

[MT14] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1633–1642, Dec 2014.

[MWB13] Anne Mangen, Bente R. Walgermo, and Kolbjørn Brønnick. Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58:61 – 68, 2013.

[NA06] John C. Nesbit and Olusola O. Adesope. Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76(3):413–448, 2006.

[NC07] Joseph Novak and Alberto Cañas. Theoretical origins of concept maps, how to construct them, and uses in education. *Reflecting Education*, 3:29–42, 01 2007.

[OIN+19] Kiyoyuki Osugi, Aya S. Ihara, Kae Nakajima, Akiyuki Kake, Kizuku Ishimaru, Yusuke Yokota, and Yasushi Naruse. Differences in brain activity after learning with the use of a digital pen vs. an ink pen - an electroencephalography study. *Frontiers in Human Neuroscience*, 13:275, 2019.

[oRdJ20] Pontifical Catholic University of Rio de Janeiro. Lua: about. `http://www.lua.org/about.html`, jun 2020. Accessed: 2020-08-12.

[OS97] Kenton O'Hara and Abigail Sellen. A comparison of reading paper and on-line documents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, page 335–342. Association for Computing Machinery, 1997.

[OTF04] Dan R. Olsen, Trent Taufer, and Jerry Alan Fails. Screencrayons: Annotating anything. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, page 165–174. Association for Computing Machinery, 2004.

[PAM+16] Alexandre Porion, Xavier Aparicio, Olga Megalakaki, Alisson Robert, and Thierry Baccino. The impact of paper-based versus computerized presentation on text comprehension and memorization. *Computers in Human Behavior*, 54:569 – 576, 2016.

[PC05] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. 01 2005.

[PML+18] Héctor R. Ponce, Richard E. Mayer, María Soledad Loyola, Mario J. López, and Ester E. Méndez. When two computer-supported learning strategies are better than one: An eye-tracking study. *Computers & Education*, 125:376 – 388, 2018.

[PSG98] Morgan N. Price, Bill N. Schilit, and Gene Golovchinsky. Xlibris: The active reading machine. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, CHI '98, page 22–23. Association for Computing Machinery, 1998.

[RHC+04] George Robertson, Eric Horvitz, Mary Czerwinski, Patrick Baudisch, Dugald Ralph Hutchings, Brian Meyers, Daniel Robbins, and Greg Smith. Scalable fabric: Flexible task management. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, page 85–89. Association for Computing Machinery, 2004.

[RHRH+19] Hugo Romat, Nathalie Henry Riche, Ken Hinckley, Bongshin Lee, Caroline Appert, Emmanuel Pietriga, and Christopher Collins. Activeink: (th)inking with data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13. Association for Computing Machinery, 2019.

[Rin03] Meredith Ringel. When one isn't enough: An analysis of virtual desktop usage strategies and their implications for design. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, page 762–763. Association for Computing Machinery, 2003.

[SA17] Lauren M. Singer and Patricia A. Alexander. Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research*, 87(6):1007–1041, 2017.

[SBR+03] Greg Smith, Patrick Baudisch, George Robertson, Mary Czerwinski, Brian Meyers, Daniel Robbins, and Donna Andrews. Groupbar: The taskbar evolved. 09 2003.

[SBW02] Peter Sutherland, Richard Badger, and Goodith White. How new students take notes at lectures. *Journal of Further and Higher Education*, 26(4):377–388, 2002.

[SF11] X. Shihu and S. Fang. Optimization for information visualization based on visual thinking. In *Proceedings of 2011 International Conference on Electronic Mechanical Engineering and Information Technology*, volume 8, pages 4243–4247, Aug 2011.

[SGL08] John Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

[SMR09] Timothy J. Smoker, Carrie E. Murphy, and Alison K. Rockwell. Comparing memory for handwriting versus typing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 53(22):1744–1747, 2009.

[SO18] Hirohito Shibata and Kengo Omura. Reconsideration of the effects of handwriting:. *ITE Transactions on Media Technology and Applications*, 6(4):255–261, 2018.

[Sof20] NCH Software. Express Scribe Transkriptionssoftware for Mac. `https://www.nch.com.au/scribe/de/index.html`, 2020. Accessed: 2020-08-15.

[SPG+99] Bill N. Schilit, Morgan N. Price, Gene Golovchinsky, Kei Tanaka, and Catherine C. Marshall. As we may read: The reading appliance revolution. *Computer*, 32(1):65–73, January 1999.

[SSH13] Szu-Yuan Sun, Chich-Jen Shieh, and Kai-Ping Huang. A Research on comprehension differences between print and screen reading. *South African Journal of Economic and Management Sciences* , 16:87 – 101, 00 2013.

[TE11a] Craig S. Tashman and W. Keith Edwards. Active reading and its discontents: The situations, problems and ideas of readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’11, page 2927–2936. Association for Computing Machinery, 2011.

[TE11b] Craig S. Tashman and W. Keith Edwards. Liquidtext: A flexible, multitouch environment to support active reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 3285–3294. Association for Computing Machinery, 2011.

[Tec] Palantir Technologies. Palantir. `https://www.palantir.com`. Accessed: 2020-07-11.

[TSO15] Kentaro Takano, Hirohito Shibata, and Kengo Omura. Effects of paper on cross-reference reading for multiple documents: Comparison of reading performances and processes between paper and computer displays. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, page 497–505. Association for Computing Machinery, 2015.

[Wee71] PD Weener. The effects of recall mode and recall interval expectancies on note-taking and recall. *Instructional Strategies, Semi-annual Report, Department of Educational Psychology, Pennsylvania State University*, pages 59–74, 1971.

[WGSS11] Manuela Waldner, Raphael Grasset, Markus Steinberger, and Dieter Schmalstieg. Display-adaptive window management for irregular surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, page 222–231. Association for Computing Machinery, 2011.

[WGSS21] Manuela Waldner, Thomas Geymayer, Dieter Schmalstieg, and Michael Sedlmair. Linking unstructured evidence to structured observations. *Information Visualization*, 20(1):47–65, 2021.

[WHP+18] J. Walny, S. Huron, C. Perin, T. Wun, R. Pusch, and S. Carpendale. Active reading of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):770–780, Jan 2018.

[WSP+06] William Wright, David Schroh, Pascale Proulx, Alex Skaburskis, and Brian Cort. The sandbox for analysis: Concepts and methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 801–810. Association for Computing Machinery, 2006.

[Zha97] Jiajie Zhang. The nature of external representations in problem solving. *Cognitive Science*, 21(2):179 – 217, 1997.

[Zim05] Liu Ziming. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. 61(6):700–712, 2020/05/27 2005.

[ZS16] Pengyi Zhang and Dagobert Soergel. Process patterns and conceptual changes in knowledge representations during information seeking and sense-making: A qualitative user study. *Journal of Information Science*, 42(1):59–78, 2016.