Effects of Field Dependence-Independence and Passive Highlights on Comprehension

by

Samuel Dodson

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF LIBRARY AND INFORMATION STUDIES

in

The Faculty of Graduate and Postdoctoral Studies

(Library, Archival and Information Studies)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2016

ABSTRACT

This study explores the effects of relevant and irrelevant highlights on reading comprehension. Participants were divided by their cognitive styles based on their degree of Field Dependence-Independence (Witkin, Dyk, Fattuson, Goodenough, & Karp, 1962). The Construction-Integration model (Kintsch, 1988) was used for the selection of reading tests that are most likely to measure comprehension. As a result, multiple choice, open-ended summary, and Sentence Verification Technique (Royer, Hastings, & Hook, 1979) questions were used.

Passive highlights were found to have significant effects on comprehension. Both Field Independents and Field Dependents were positively affected by relevant highlights and negatively affected by irrelevant ones. Differences were found between measures of comprehension used in the study, suggesting the comprehension tests measure different components of comprehension. These results have implications for the future study of reading.

PREFACE

This thesis is an unpublished work by the author, Samuel Dodson. Drs. Luanne Freund and Rick Kopak provided recommendations for the design of the study and the analysis of data.

This work was approved by the University of British Columbia Behavioural Research Ethics Board under the project title *Effects of Text Annotations in Digital Reading Environments,* with the certificate number H15-00652.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my advisors, Drs. Luanne Freund and Rick Kopak, for their support throughout this project. I am grateful to Tim Rainey, Saguna Shankar, and Abeer Siddiqui for reading drafts of this work. I would also like to thank the faculty and staff at the School of Archival, Library, and Information Studies at the University of British Columbia for their support of my studies.

CHAPTER 1

INTRODUCTION

1.1 PURPOSE OF THE STUDY

Highlighting is one of the most common types of annotation; however, we know little about its effects on reading. Highlights can be divided into active highlights, which are created while reading for a specific task, and passive highlights, which already exist in the text. Active highlights may be a sign of active reading, which has been found to have positive effects on reading (Adler & Van Doren, 1972). Passive highlights, on the other hand, are already in the text and are encountered by the reader regardless of their purpose for examining the text. These highlights may, or may not, be relevant to the reader's task. It is practical – even necessary – to study passive highlights, partly because many used texts contain highlighting (Fowler & Barker, 1974). This work is also relevant to digital reading systems, which frequently allow readers to share their annotations. By comparing the effects of relevant and irrelevant highlights on readers, we may learn whether passive highlights have pos-

itive or negative effects on reading.

1.2 RESEARCH QUESTIONS

Can highlighting support our information processing needs? Cognitive styles, such as Field Dependence-Independence (Witkin et al., 1962), describe how we process information (Messick, 1976). Witkin et al. (1962) state that Field Dependents use external cues to guide their information processing, while Field Independents use internal cues. Because of their need for external cues, it is easier for Field Dependents to process information when they are provided relevant cues (Kent-Davis & Cochran, 1989). The effects of relevant and irrelevant highlights on Field Dependents and Independents have not been studied. For this reason, we do not know if the comprehension of readers with different cognitive styles is affected by passive highlights, and, if so, whether those readers are affected in different ways.

1.3 SIGNIFICANCE OF THE STUDY

Many digital reading systems allow readers to create and share annotations. These services are also able to combine the highlights of readers to identify the passages with the most highlights (F. Shipman, Price, Marshall, & Golovchinsky, 2003). It is thought that the resulting highlights may help guide readers to important passages in the text (Marshall, 1998). When relevant, these highlights may help readers focus on passages that are important to their tasks. These services, however, could

present highlights that are irrelevant to readers' tasks. Both of these situations may have an effect on reading outcomes. In this study, for example, relevant highlights positively affected comprehension, but irrelevant highlights had a negative effect. For these reasons, we should be cautious when using passive highlights in digital reading systems.

The findings of this work may also lead to better practices for studying reading. The results of the study provide an argument for using comprehension tests other than multiple choice questions, which are commonly the only measure used in studies. Differences were found between all of the measures of comprehension used in the study. This suggests that the comprehension tests were measuring different components of comprehension. For this reason, future studies should devise more robust ways of measuring reading comprehension.

1.4 SUMMARY

To increase our knowledge of the effects of passive highlighting, we studied how relevant and irrelevant highlights affect Field Dependent and Field Independent readers. We found that Field Dependents and Field Independents were positively affected by relevant highlights and negatively affected by irrelevant highlights. These results suggest that highlights can support the needs of readers with different cognitive styles. In addition, the results indicate that irrelevant highlights can negatively affect comprehension.

These results build on previous studies of highlighting. The study made two

contributions. First, it measured the effects of relevant and irrelevant highlights on readers with different cognitive styles, specifically Field Dependence and Field Independence. Second, the study used measures of comprehension in accordance with the Construction-Integration model (Kintsch, 1988). These measures took into account the importance of measuring comprehension at a deeper level that integrates information from the text with the reader's knowledge. The study makes a contribution to work on comprehension and reading by providing a greater understanding of the effects of relevant and irrelevant highlights on readers with different cognitive styles. The findings of the study provide a greater understanding of the effects of passive highlights on comprehension.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

The literature review synthesizes existing research in order to lay a foundation for the development of improved systems and theories related to highlighting and reading practices. This chapter is divided into five sections. The first section examines the parts of an annotation as described by Marshall's model of annotation (2009). This section continues with a comparison of the forms and functions of different types of annotations. The second section reviews a selection of studies on the effects of passive highlights on reading. The third and fourth sections introduce the psychological models central to the study. The third section introduces the Construction-Integration model (Kintsch, 1988), which describes how readers create a mental representation of text. The fourth section describes the Field Dependence-Independence (Witkin et al., 1962) cognitive style. The fifth section outlines the research questions and hypotheses of the study.

2.2 ANNOTATION

To understand the functions highlights serve, we must first describe the components of annotations and how they operate. This section is divided into four subsections. The first subsection identifies the universal components of annotation through a model. This allows us to explore how these parts act as building blocks that create different types of annotations. It also hints at how these parts affect forms and functions. The second subsection discusses link associations, or how annotations link to the text. The third subsection moves on to an evaluation of the form of annotations. The fourth subsection reviews the function of annotations.

2.2.1 MODELS OF ANNOTATION

Adding annotations to paper texts using most writing instruments is an easy, straightforward act. As more readers use electronic texts, it is becoming increasingly important that digital reading systems provide easy-to-use annotative tools.[1] While progress has been made on adding annotative functionality to digital reading systems (e.g., Nelson et al., 2009; Pearson, Buchanan, & Thimbleby, 2011, 2013; Pearson, Buchanan, Thimbleby, & Jones, 2012; Price, Schilit, & Golovchinsky, 1998; Schilit, Golovchinsky, & Price, 1998), more work is needed before digital texts are as easy to annotate as analog documents. Key to these efforts are models of annotation, which "create a universal representation that facilitates the creation of

[1] Readers annotate less in electronic texts than print ones (Marshall, 1997; O'Hara & Sellen, 1997; Sellen & Harper, 2001). In their study of electronic document use, Sellen and Harper (2001) found readers believe annotations should exist in a separate layer on top of the text, and are wary of systems that do not make clear distinctions between the text and its annotations.

usable annotations in digital documents" (Pearson et al., 2013, p. 61). Some models of annotation, such as the one provided by Agosti, Ferro, Frommholz, and Thiel (2004), specify the technical requirements for creating digital annotations. This level of technical specificity is not needed in our study, and focusing on how annotations will be represented in a markup language or database schema will likely distract us from our goal in this section – to evaluate the parts, forms, and functions of annotations. We will use a model of annotation to identify the parts of annotations; this is the necessary first step to the larger discussion of the forms and functions of different types of annotations.

When selecting a model, it is important to choose one that is not too technical. Describing how annotations can be represented in a markup language for inclusion in digital reading systems, for example, is not necessary here. The model, however, must be robust enough to apply to the study of annotation. The following section examines the forms and functions of different types of markings through Marshall's model of annotation (2009). Marshall's model identifies the universal parts of annotations. Together, these components control the form of markings. A discussion of the different characteristics of these components of annotation, and their forms, will allow us to deduce how, and why, readers use them. The motivation for reviewing various shapes and functions of annotations is twofold. First, it provides a vocabulary for the comparison of types of annotations. Second, reviewing the form and function of annotations allows us to see how the different characteristics of markings support some kinds of interactions between the reader and text, but not others.

There has been a considerable amount of work on modelling annotations. Much of this work has been done to incorporate annotative functionality into digital reading systems (e.g., Agosti et al., 2004; Agosti & Ferro, 2007; Constantopoulos, Doerr, Theodoridou, & Tzobanakis, 2005; Haslhofer, Simon, Sanderson, & Van de Sompel, 2011; World Wide Web Consortium, 2015). Most models of annotation are built for a specific digital reading system. As a result, Marshall's model (2009), which is system agnostic, is most appropriate for this study.

2.2.2 FORM

Once we identify the components of annotation, we can see how these parts make up the forms and functions of different types of markings. Marshall's model can "distinguish between the different components of annotations" (Pearson et al., 2013, p. 61). In Marshall's model, annotations are composed of three elements: a body, an anchor, and a marker. The first element, body, is any content the reader adds to the text. The content may be verbose, like a lengthy note in the margin, or as succinct as a single character, such as an asterisk, that marks an important passage. The second element, anchor, is the link from the annotation to the text. Anchors may be explicit, like a highlighted passage, which has a clearly delimited start and end, or implicit, such as a note in the margin that is linked to the text through proximity. The scope of an anchor may be broad, referencing larger components of a text, such as a summary of an entire chapter, or narrow and localized, like a translation of a foreign word. The third element, marker, is the display of the annotation. For an example of the model applied to an annotation, see Figure 2.1.

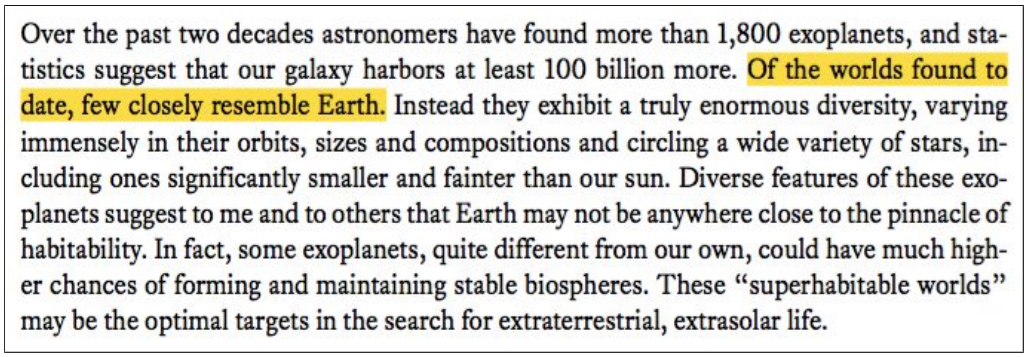

Over the past two decades astronomers have found more than 1,800 exoplanets, and statistics suggest that our galaxy harbors at least 100 billion more. Of the worlds found to date, few closely resemble Earth. Instead they exhibit a truly enormous diversity, varying immensely in their orbits, sizes and compositions and circling a wide variety of stars, including ones significantly smaller and fainter than our sun. Diverse features of these exoplanets suggest to me and to others that Earth may not be anywhere close to the pinnacle of habitability. In fact, some exoplanets, quite different from our own, could have much higher chances of forming and maintaining stable biospheres. These "superhabitable worlds" may be the optimal targets in the search for extraterrestrial, extrasolar life.

Figure 2.1: A passive highlight in "Better than earth" by Heller (2015). This annotation can be evaluated with Marshall's model: there is no body because the reader did not add any content to the text. The anchor, which begins at "Of the worlds" and ends at "closely resemble Earth", is explicitly delimited. The scope of the annotation is narrow, and only references a sentence. The marker is a yellow highlight that overlays the emphasized passage.

According to the model, highlights have no body because the reader does not use them to add content to the text. They have an explicit anchor, which is visually delimited by the beginning and end of the highlighted passage. The scope of the annotation is usually narrow because highlights must be explicitly delimited, which discourages marking broader passages, like whole sections or chapters, in favour of paragraphs, sentences, or words. Highlights have a marker that is displayed as a coloured overlay that spans the linked content.

The border between form and function is often blurry. For example, the colour of a highlight contributes to both form and function. The colour is a characteristic of form, yet it can also communicate meaning (Gaddy, 1996; MacMullen, 2005; Worley, 1999). A colour scheme can represent different types of information (Marshall, 1997; Schilit, Price, Golovchinsky, Tanaka, & Marshall, 1999; Wolfe, 2002).

However, Marshall (2000, p. 112) found "It was rare to find one of these schemes that lasted throughout a textbook." Highlights allow readers to quickly mark up a passage; selecting the "correct" colour may be too slow of a process, or too cognitively taxing, to sustain for a long text that may require several reading sessions. In paper-based reading environments readers may not have all the coloured markers at their disposal. In a digital reading system, the colour options may be multiple clicks away in a dense menu. In these cases, the value of a colour scheme is outweighed by the inconvenience of its creation. The value of the colour scheme may also be short-lived. Without a key mapping colour to information type, the function, or meaning, of a colour will likely be lost; consequently, even the annotator may forget the meaning of the marking in subsequent readings (Marshall, Price, Golovchinsky, & Schilit, 1999).

2.2.3 LINK ASSOCIATIONS

Annotations allow readers to create links not only within, but also between texts. There are four types of link associations: collection, node-to-annotation, standard, and word-to-word associations (Marshall, 1998). Different types of annotations support different associations. For this reason, readers use different types of markings to serve their expressive needs when interacting with texts.

Each association has a different scope of linked content. Collection associations have the broadest scope because they connect an annotation to larger components of a text, such as chapters or sections. Node-to-annotation and standard level associations connect an annotation to smaller components of a text, usually paragraphs

or sentences. Node-to-annotation level associations have implicit anchors, such as a note in the margin that is only linked to the text by proximity. Conversely, standard links, such as highlights, have explicit anchors with bounded starts and ends. Word-to-word associations have the narrowest scope because they connect annotations to the smallest components of a text – that is, words or characters. Bélanger (2010) adds collapsed association – some markings collapse the body-anchor-marker trio on itself. For example, a highlight acts as the body, anchor, and marker altogether, linking, or referring, to itself (Bélanger, 2010).

Highlights can use three of Marshall's (1998) four link associations: standard, word-to-word, and collection associations. The linking abilities of these associations, however, are limited. Because they can only span linear text, highlights cannot connect non-sequential passages. Some types of annotations can use auxiliary tools, such as arrows and brackets, to create links between one annotation and multiple, non-linear passages. The most common types of highlighting link associations are the standard and word-to-word associations. This may be due to the effort needed to explicitly mark an emphasized passage, which causes readers to highlight smaller passages, often at the word, sentence, or paragraph level of a text, as opposed to extended passages. But, highlights can create collection associations. This is most common when readers highlight page after page. This reading behaviour creates "a visible trace of a reader's attention, a focus on the passing words, and a marker of all that has already been read" (Marshall, 1997, p. 136). Extensive highlighting is usually found in difficult texts, where readers use the highlighter to focus their attention on the text, similarly to how readers use their index fingers to guide

their eyes across the page. The highlighter follows the line of words in the text, allowing readers to keep their place. In this case, the reader uses highlighting as a type of aid to keep their place in the text rather than as a marking tool. When all text is highlighted, no content is emphasized; the large spans of highlighted text are likely of little or no value on subsequent readings.

2.2.4 FUNCTION

Some types of annotation are better suited for a given reading task than others. For example, in a first reading, readers may exert a great amount of cognitive effort to understand the meaning of a text. This may include re-reading difficult sections or taking marginal notes. When readers revisit a text on subsequent readings, they may spend less time reading sequentially and more time reviewing their notes or highlights. These different tasks – reading for understanding and reading for re-membering – result in different reading behaviours, which are served by different types of annotations.

Marshall (1997) identified six functions of annotation by studying the markings she found in used college textbooks. First, annotations act as procedural signals for future use. For example, students highlight assigned content and, conversely, strike out unassigned text. Second, annotations are placemarks for passages to be re-read or referenced later. Third, annotations are *in situ* locations for problem-solving. Students work on practice problems problems where they are presented in the text-book. Marshall (1998), for example, found one student calculated the rotations of molecules beside related figures in a chemistry textbook. Fourth, annotations are a

record of interpretive activity. Notes in the margin are used to paraphrase the text in the annotator's own words. Fifth, annotations are a "visible trace of the reader's attention" (Marshall, 1997, p. 136), especially when the material is difficult. This is usually manifested as extensively highlighted pages of text. Sixth, annotations are "incidental reflections of the material circumstances" (Marshall, 1997, p. 137). Irrelevant doodles and sketches may be a reflection of a reader's disengagement. Readers are affected by distractions that are external to the text and task.

Highlights can serve half of the six functions: they can act as procedural signals for future use, placemarks, and a visual trace of the reader's attention. Marshall's findings are in line with other studies. O'Hara and Sellen (1997), for example, suggest that highlights are especially suited to allow readers to skim from one annotation to the next to remember the main points of a text. Ovsiannikov, Arbib, and McNeill (1999) also found that highlights are used for marking passages for future reference. The ability to serve multiple function make highlights a flexible type of annotation.

2.3 HIGHLIGHTS

While readers have been annotating texts for thousands of years (Jackson, 2001), highlighting is a relatively new practice. In 1962, Yukio Horie created the fiber-tip pen. This eventually became the highlighter a year later, in 1963, when Carter's Ink Company launched the Hi-Lighter (Ward, 2015).

The pen is available in a wide range of colors, although yellows and

> pinks continue to dominate the highlighter market, representing around 85 percent of total sales. Sitting right in the middle of the spectrum of visible light, yellow leaps out from the page and can be seen more easily than any other color (Ward, 2015, p. 179).

Highlights are an easy and quick way to emphasize passages. When a reader highlights a passage, there is no need, or way, to justify the rationale for emphasizing the text. Other types of annotation, such as marginalia, the annotator needs to explicitly express his or her idea. Because of this, a highlight is less cognitively demanding to create. Of the types of annotation, typographical cues, such as bolding, bracketing, italicizing, and underlining, are most similar in nature to highlighting. Like highlights, typographical cues are also used to draw attention to specific passages. However, there is a distinction between these markings and highlights, which is discussed further in section 2.3.1.

Highlights' ease of creation may be a disadvantage in disguise. It is not clear if active highlighting supports as thorough, or thoughtful, of an interaction between reader and text as are possible with other types of annotation, such as marginalia. When a reader creates a highlight, there is no way to express the importance of the linked text. In the case of passive highlights, readers have no opportunity to engage with an emphasized passage, because it may not be clear why the annotation was created. There is no way for a series of subsequent readers to have a dialogue with one another through their annotations, as they can with notes in the margins. Nonetheless, the practice of highlighting was widely adopted by readers, and continues to be one of the most used types of annotation (Baron, 2009).

There have been few studies on the effects of passive highlighting and reading outcomes. Reviewing the existing work provides a foundation for the future studies on the effects of highlighting. This section is divided into three subsections. The first subsection discusses the differences between highlighting and other typographical cues. The following subsections review previous studies on highlighting. The second subsection looks at active and passive highlighting, while the third subsection reviews relevant and irrelevant highlighting.

2.3.1 TYPOGRAPHICAL CUES

Some studies have treated highlighting and other forms of annotation, called typographical cues, such as bolding, bracketing, italicizing, and underlining, as equivalent reading techniques with similar effects. However, there are differences between highlighting and typographical cues, suggesting it should not be assumed they have the same effects on readers. While a highlight and a typographical cue of the same content have the same body and anchor, they have different markers – the highlight would be displayed as a transparent coloured overlay spanning the linked content, while the typographical cue would be displayed as a bold, italic, or underlined passage. Highlighting has a different form and provides different functions than bolding, bracketing, italicizing, and underlining.

Previous studies on typographical cues have mixed results. Some studies found a positive or neutral effect on recall of cued content (e.g., Klare, Mabry, & Gustafson, 1955; Nist & Hogrebe, 1987; Peterson, 1991). Foster and Coles (1977) found, however, that underlining using a black line marker could make text harder to read than

unmarked text. It is difficult to compare studies on typographical cueing because the methodologies used vary greatly from one study to the next, as noted by Hartley, Bartlett, and Branthwaite (1980). For this reason, it is difficult to draw conclusions from the findings, especially when considering that each study may be testing different aspects of reading outcomes. Furthermore, because typographic cues do not share the same form and function, it is not clear if the results of studies on bolding, bracketing, italicizing, or underlining, for example, would be applicable to highlighting. It may be acceptable to group highlighting with other typographical cues, but it is clear from the contradictory results regarding these various cues that each may have different effects on reading. For these reasons, we are justified in studying highlighting and avoiding drawing conclusions from findings on other typographical cues.

2.3.2 ACTIVE & PASSIVE HIGHLIGHTS

It is practical to study passive highlighting, since most used texts have been highlighted by previous readers. Fowler and Barker (1974), for example, found that ninety-two percent of used textbooks were thoroughly highlighted. Marshall (2000) found similarly extensive levels of highlighting in used textbooks. These results may be less representative of used texts outside of college bookstores, however, where books may pass through fewer hands and the texts may be read less critically. Nonetheless, it is likely that many texts contain some highlighting. With the rise of social reading features in digital reading systems, it is likely we will continue to read highlighted text even as some of us move from analog to digital texts.

Fowler and Barker (1974) studied the effects of highlighting on recall of highlighted content. The authors tested if active and passive highlights help readers remember content. Subjects were randomly assigned to one of four conditions. The first condition used active highlighting. Subjects in this condition were allowed to highlight as much as they liked. The second and third conditions used passive highlighting. Subjects in the second condition read highlighted text created by subjects in the first condition. Subjects in the third condition read highlighted text created by the experimenters. This was the experimenter-created passive highlighting condition. The fourth condition was the control condition and had no highlights. Subjects had up to one hour to read two articles. A week later, the subjects reviewed the same articles for ten minutes. Then subjects completed a multiple choice recall test.

The differences between active and passive highlighting were difficult to measure, because the subject-generated highlights varied greatly in length. There was no statistically significant difference in recall scores between any of the conditions. There was a positive effect on recall, however, in the experimenter generated passive highlighting condition when questions were from highlighted passages. Thus, subjects in this condition benefited from highlighting, but this effect was too weak to increase their total score by a statistically significant amount. While not significant, it was found that highly relevant highlights, which emphasized content that appeared on the test, had increased recall scores in active and passive conditions. This effect was strongest in the active highlighting condition.

Lorch (1989) suggests that too much text in the Fowler & Barker study may have

been highlighted. In the experimenter highlighting condition, a quarter of the text was highlighted. Those in the subject-generated highlighting condition highlighted as much as thirty-two percent of the text. Readers may ignore highlighting if too much text is highlighted, because of the great effort needed to process the emphasized content (Lorch, Pugzles Lorch, & Klusewitz, 1995). Lorch et al. (1995) found that subjects had better recall of text when five percent of text was cued as opposed to fifty percent. Lorch notes, "the proportion of cued material may be expected to influence the effectiveness of the cue because the distinctiveness of the cued information decreases as the proportion of cued content increases" (Lorch, 1989, p. 225).

The von Restorff effect (von Restorff, 1933) provides an explanation for the diminishing value of highlighting when too much text is emphasized. The von Restorff effect predicts that an item that stands out from its background, such as a highlighted passage, is more likely to be remembered than items that do not. When applied to reading, the von Restorff effect suggests that readers focus their attention on emphasized content regardless of its relevance (Nist & Hogrebe, 1987; Peterson, 1991). In an eye-tracking study, Chi, Gumbrecht, and Hong (2007) found further evidence of the von Restorff effect. The authors found that readers' eyes "jump" from one highlighted passage to the next. It is not clear if, or how, readers evaluate the relevancy of highlights. For example, are some readers better at filtering relevant and irrelevant highlights than others? Have readers learned through experience to conserve the cognitive effort needed to assess the relevancy of a highlight and just expect semantic value from highlighted text?

2.3.3 RELEVANT & IRRELEVANT HIGHLIGHTS

The relevancy of a highlight is situational – it depends on reader's task. Passive highlights created to support readers' current tasks are relevant, while those produced for different tasks are irrelevant. What one reader may emphasize could be considered irrelevant to another reader. In fact, readers may find their own annotations become irrelevant as their tasks change with subsequent readings. Passive highlights may become distracting when the task for which they were created is no longer relevant to the reader. They may act as "noise", drawing the reader's attention away from the text, as predicted by the von Restorff effect (von Restorff, 1933).

Silvers and Kreiner (1997) studied the effects of relevant and irrelevant highlighting in a two-part study. The first part tested if relevant highlights have a positive effect on subjects' comprehension and irrelevant highlights have a negative effect. The Nelson-Denny Reading Comprehension Test (NDRCT) was used to measure subjects' comprehension. The NDRCT was administered in one of three conditions: no highlighting (control), relevant highlighting, and irrelevant highlighting. Compared to the control condition, the relevant highlighting condition had no effect on comprehension. Irrelevant highlights, however, had a negative effect.

To test if the negative effect of irrelevant highlighting found in the first experiment could be reduced or neutralized, subjects in the second part of the study were warned that the highlights may be irrelevant. Other than this warning, the same methodology as the first experiment was used. Mean comprehension scores were lowest in the irrelevant highlighting condition. However, there was no statisti-

cally significant difference between the control and relevant highlighting conditions. These results may have been affected by the short length of the NDRCT. The effects of relevant and irrelevant highlighting have not been studied on longer texts, which may be more representative of the reading material assigned at universities or read in the real world.

The results of the study suggest that irrelevant highlighting results in lower comprehension, even when subjects are warned that highlights may be irrelevant to their task. This suggests that readers may not be able to easily identify or ignore irrelevant highlighting. This is concerning, given the regular occurrence of pre-existing highlights in used texts.

2.3.4 KEYWORD & PASSAGE HIGHLIGHTS

Cao (2006) studied the effects of cognitive style and passive highlighting on reading. Subjects' degree of Field Dependence-Independence (see section 2.5) was measured. Subjects were randomly assigned to one of three conditions. Each condition used the same text, but varied in the type of highlighting used. The first condition highlighted keywords, the second highlighted passages, and the third had no highlighting (control). Subjects were given twenty-five minutes to read the text, before completing a multiple-choice comprehension test. Subjects were not allowed to return to the text once they had started the test.

The mean comprehension scores of Field Independents were significantly higher than Field Dependents in the control and keyword conditions. There was not a significant effect of highlighting on Field Dependents' comprehension scores. A sig-

nificant difference, however, was found between the comprehension scores of Field Independents in the control condition and in the two other highlighted keyword and passage conditions. This suggests that highlights presents "noise" that distracts Field Independents and impairs their comprehension. These results suggest that highlighted keywords and passages do not help Field Dependents' comprehension, and can hurt Field Independents' comprehension.

It is surprising that there was no statistical difference between the two highlighting conditions. Lorch (1989) suggests that highlighting smaller pieces of a text would result in better recall. There are, however, notable differences between recall and comprehension.[2] The difference in the lengths of the keywords and passages may also be too small for a difference to be found – Lorch et al. (1995) found longer cued passages were ten times larger than the shorter passages. These results suggest that, at best, passive highlights have no effect on comprehension, but may hurt readers' understanding of a text.

The previous studies on highlights contribute to our understanding of the practice, but each study has limitations. First, each of the studies measured reading outcomes, such as comprehension, with multiple choice tests. These are simple evaluation methods that are unlikely to assess the effects of annotation on different levels of understanding. A number of studies have shown that multiple choice tests are poor measures of comprehension (Drum, Calfee, & Cook, 1981; Ozuru,

[2] Recall is the ability to remember information from the text, while comprehension is the capacity to understand that information. Comprehension tests should be designed to ensure that questions do not demand recall skills. "The issue of comprehension has not been as fully researched as one might expect, perhaps in no small way due to the difficulty of devising a suitable means of quantification; that is, how does one measure a reader's comprehension?" (Dillon, 2003, p. 42).

Briner, Kurby, & McNamara, 2013; Tuinman, 1973). Many of these testing methods may assess other abilities, such as intelligence, in addition to comprehension (Royer, Greene, & Sinatra, 1987; Royer et al., 1979). A failure to isolate the controlled variable, which is in this case comprehension, seriously compromises the validity of the results.

2.4 CONSTRUCTION & INTEGRATION

The Construction-Integration model creates a mental model of the text in two steps (Kintsch, 1998). In the first step, the reader creates nodes for all meaning in a text. These nodes are either derived from the text (i.e. the textbase) or the reader's knowledge (i.e. the situation model). In the second step, these nodes are either joined into the reader's mental model of the text or removed. Nodes that are relevant to the textbase or situation model are added to the mental model of the text, which is made up of a connected network of nodes (Kintsch, 1998). Kintsch notes, "The reader must add nodes and establish links between nodes from his or her own knowledge and experience to make the structure coherent, to complete it, to interpret it in terms of the reader's knowledge, and last but not least integrate it with knowledge" (1998, p. 103). These two steps create a mental model of the text. Comprehension occurs if, and only if, the majority of the relevant nodes are connected together and the irrelevant nodes are removed from the model.

The distinction between understanding at microstructure and macrostructure is critical to the study of comprehension. A reader can establish a microstructure of

a text that is sufficient to answer some questions about the text without comprehending its higher-level meaning. Previous highlighting studies, however, have all used measures of comprehension that evaluate shallow, not deep, levels of the mental representations of comprehension. This is why more robust testing methods are needed to determine if readers' have achieved a deeper level of comprehension.

2.4.1 MICROSTRUCTURE & MACROSTRUCTURE

The microstructure and macrostructure represent nodes of the meaning of the text at shallow and deep levels. The local elements of meaning form the microstructure. These are local elements in the sense that they are close to each other in the text. The macrostructure lifts the most important nodes up and out of the text and connects them together in a network of nodes. The macrostructure forms a global, overall meaning of the text (Kintsch, 1998). Bottom-up processes allow the reader to interpret meaning of sentences and break the text into local meaning units (Leighton & Gierl, 2011). The meaning units that are most important to the overall meaning of the text are then connected to form a network, also known as the macrostructure, which represents the gist of the text (Kintsch & Van Dijk, 1978; Kintsch, 1998). The macrostructure is generated by bottom-up and top-down processes that select the most relevant elements from the microstructure (Butcher & Kintsch, 2003; Kintsch & Van Dijk, 1978).

2.4.2 TEXTBASE & SITUATION MODEL

Kintsch notes, "the mental representation of a text a reader constructs includes the textbase … plus varying amounts of knowledge elaboration and knowledge-based interpretations of the text – the situation model" (1998, p. 50). The distinction between the textbase and situation model refers to the origin of the meaning units in the mental representation of the text (Kintsch, 1998). The textbase meaning units are all derived from the text, while meaning units in the situation model are knowledge-derived. Kintsch states, "the mental text representation is a mixture of text-derived and knowledge-derived information" (1998, p. 104). It is unlikely that all the meaning units necessary to understand a text will be provided in the text (Kintsch, 1998). Because of this, the reader may need to fill in information gaps in the textbase with their knowledge. Kintsch notes, "knowledge may be...needed to complement the textual information and to transform what by itself is only an isolated memory structure into something that relates to and is integrated with the reader's personal store of knowledge" (Kintsch, 1998, p. 103). However, in some cases the reader may lack the topic knowledge to build an effective situation model. Kintsch states this "typically occurs when a reader lacks the background knowledge necessary for a full understanding of the text" or "when a reader has relevant background knowledge but does not use it during comprehension. Passive readers are not rare, and to ensure learning from text, such readers have to be jolted out of their passivity" (1998, p. 232).

2.4.3 CONSTRUCTION & INTEGRATION

The Construction-Integration model creates a mental model of the text in two steps (Kintsch, 1998). In the first step, the reader creates nodes for all meaning in a text. These nodes are either derived from the text (i.e. the textbase) or the reader's knowledge (i.e. the situation model). In the second step, these nodes are are either joined into the reader's mental model of the text or removed. Nodes that are relevant to the textbase or situation model are added to the mental model of the text, which is made up of a connected network of nodes (Kintsch, 1998). Kintsch notes, "The reader must add nodes and establish links between nodes from his or her own knowledge and experience to make the structure coherent, to complete it, to interpret it in terms of the reader's knowledge, and last but not least integrate it with knowledge" (Kintsch, 1998, p. 103). These two steps create a mental model of the text. Comprehension occurs if, and only if, the majority of the relevant nodes are connected together and the irrelevant nodes are removed from the model.

The distinction between understanding at shallow and deep levels is critical to the study of comprehension. The Construction-Integration model shows that a reader can establish a microstructure representation of a text that is sufficient to answer some questions about the text without comprehending its higher-level meaning. Previous highlighting studies, however, have all used measures of comprehension that evaluate shallow levels of the mental representations of comprehension. This is why more robust testing methods are needed to determine if readers' have achieved a deeper level of comprehension.

2.5 FIELD DEPENDENCE-INDEPENDENCE

Our cognitive styles control how we process information (L. J. Ausburn & Ausburn, 1978; Messick, 1976; Tinajero & Páramo, 1997). Tinajero and Páramo (1997, p. 200) note "depending on their cognitive style, subjects appear to pay attention to different aspects of information, to encode, store and recall information differently, and in general to think and comprehend in different ways." Field Dependence-Independence (Witkin et al., 1962) is one of the most studied cognitive styles (S. Shipman & Shipman, 1985; Witkin & Goodenough, 1981; Witkin, Moore, Goodenough, & Cox, 1975). Messick describes the difference between Field Dependents and Field Independents as follows:

> The field independent person tends to articulate figures as discrete from their backgrounds and to easily differentiate objects from embedding context, while the field dependent person tends to experience events globally in an undifferentiated fashion. Field independent (or analytical) individuals have more facility with tasks requiring differentiation and analysis (1976, p. 5).

While Field Dependence-Independence was originally a measure of perceptual ability (to structure or restructure visual fields), it has been found that this ability affects other cognitive tasks, such as problem solving (Witkin & Goodenough, 1981), cognitive restructuring ability (Goodenough, 1976), active hypothesis testing (Davis & Frank, 1979), and attention to relevant cues (Berger & Goldberger, 1979).

Most people are not fully Field Dependent or Independent, and show characteristics of both styles. The term "Field Mixed" is used to describe those in the middle of the Field Dependent-Independent continuum (Liu & Reed, 1994). For this reason, we should review the characteristics of both cognitive styles.

2.5.1 FIELD DEPENDENCE

Field Dependents use external cues to guide their information processing, while Field Independents use internal ones. Field Dependents are more likely to use the existing organization of a field (Witkin, Goodenough, & Oltman, 1979). It is difficult for Field Dependents to focus on the most important information, especially when they are presented with distracting cues (Kent-Davis & Cochran, 1989). Field Dependents count on external cues to guide their attention. For these reasons, Witkin and Goodenough (1977, p. 8) note "[a Field Dependent] is likely to have difficulty with that class of problems…where the solution depends on taking some critical element out of the context in which it is presented and restructuring the problem material so that the item is now used in a different context." In addition to having difficulty structuring visual stimuli, Field Dependents find it challenging to solve problems that require information be separated from its context and used in other contexts (Witkin & Goodenough, 1981). This suggests that "individual differences in expressions of articulated function in one area are related to expression in other areas" (Goodenough, 1976, p. 676). When a field is well-structured, however, Field Dependents can perform as well as Field Independents (Witkin et al., 1979).

2.5.2 FIELD INDEPENDENCE

Field Independents distinguish elements from the background of a given field to organize fields that lack structure (Witkin et al., 1979). Field Independents also have the ability to restructure a field, while Field Dependents have difficulty structuring or restructuring a field (Witkin & Goodenough, 1981). Field Independents are better at focusing their attention on relevant information and ignoring distractions than Field Dependents (Kent-Davis & Cochran, 1989). These abilities come from their use of internal references (Witkin et al., 1962). Field Independents "are more likely to be aware of needs, feelings, attributes, which they experience as their own, and as distinct from those of others. These distinctive needs, feelings, and attributes in effect provide internal frames of reference to which the person may adhere in dealing with external social referents" (Witkin et al., 1975, p. 19).

2.6 SUMMARY

Few studies on relevant and irrelevant highlighting have considered the effects of cognitive style. The literature could be improved by the integration of psychological models of comprehension, such as the Construction-Integration model (Kintsch, 1988), and cognitive styles, for example, Field Dependence-Independence (Witkin et al., 1962). The effects of relevant and irrelevant highlights on Field Dependents and Field Independents has not been studied. Are Field Dependents, who rely on the given structure of a stimulus, guided to important passages by relevant highlights? Are they hopelessly mislead by irrelevant highlights? Are Field Indepen-

dents, who easily differentiate stimuli from context, able to assess the relevancy of passive highlights? These questions are addressed in the study.

2.6.1 RESEARCH QUESTIONS

Based on a review of the literature, several research questions emerge concerning the influence of cognitive styles and passive highlighting on reading outcomes. Two research questions motivated the study. First, are readers able to identify relevant highlights and use them to guide their attention to the most important information in a text? Conversely, are readers able to ignore irrelevant highlights? Second, are Field Dependents and Field Independents affected by relevant and irrelevant highlighting to the same degree? In other words, are readers' abilities, or inabilities, to identify relevant and irrelevant highlights related to their field structuring capacities?

2.6.2 HYPOTHESES

Four hypotheses were formed based on the research questions. First, relevant highlights will positively affect Field Dependents' comprehension. Second, relevant highlights will have no effect on Field Independents' comprehension. Third, irrelevant highlights will negatively affect Field Dependents' comprehension. Fourth, irrelevant highlights will not affect Field Independents' comprehension.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

This chapter describes the subjects, materials, procedures, and data analysis used in the study. The first section discusses the design of the study. This includes a review of the dependent and independent variables. The second section describes the subjects. The third section describes the materials, including the three texts and corresponding quizzes. The fourth section lists the procedures of the study. The fifth section describes the data analysis.

3.2 SUBJECTS

Twenty-nine undergraduates (fourteen males and fifteen females) from the University of British Columbia participated in the study.[1] Participants were recruited

[1] One subject, and her data, was excluded from the study, because her Group Embedded Figures Test score of zero was abnormally low. This subject also did not attempt to complete all of the comprehension quizzes or the post-session questionnaire.

through advertising at the Irving K. Barber Learning Centre at the University of British Columbia.

Consent forms, approved by the University of British Columbia Behavioural Research Ethics Board, were signed prior to beginning the study. As part of their consent, all subjects reported being able to read proficiently in English. The study took approximately one and a half hours. Subjects were paid a $20.00 honorarium for their time.

A pre-session questionnaire was used to obtain an understanding of the subjects' demographic information. The subjects form a representative group of undergraduate university students. Of the twenty-nine subjects, twenty-eight were in the age range of eighteen to twenty-four, and one was twenty-five to twenty-nine years old. Thirteen freshman, five sophomores, seven juniors, and four seniors participated. Twelve subjects were from the Faculty of Arts, sixteen from Sciences, and one was undeclared. The questionnaire was also used to learn about the subjects' reading behaviours and use of electronic texts. The mean usage scores show that subjects divide their reading time between analog (M=50.414) and electronic (M=49.586) texts.

3.3 MATERIALS

3.3.1 GROUP EMBEDDED FIGURES TEST

The Group Embedded Figures Test, or GEFT, (Witkin, Oltman, Raskin, & Karp, 1971) was used to measure subjects' degree of Field Dependence-Independence.

A pencil and paper GEFT was administered. It took about fifteen minutes to complete the test. The completed GEFTs were scored based on the answer key provided by Demick (2014). Field Dependents tend to have more difficulty and spend more time finding the embedded figure than Field Independents, hence Field Dependents' scores are lower.

There are several ways to classify Field Dependence and Field Independence. Witkin et al. (1971) divided subjects' scores into quartiles. Subjects who score in the lower quartile are Field Dependents, those who score in the upper quartile are Field Independents, and those in the middle quartiles are Field Mixed. Another acceptable method divides subjects into Field Dependent and Field Independent based on a median split; those subjects who score less than the median score are classified as Field Dependent, those who score greater than the median score are classified as Field Independent, and those at the median are removed from the data analyses (Demick, 2014). A variant of this method was used in the study: subjects who scored at, or below, the median GEFT score were classified as Field Dependents; and those who scored above the median were labelled as Field Independents. This method resulted in fifteen subjects being labelled Field Dependents and fourteen Field Independents.

There was neither a bias of year nor academic discipline between the two cognitive styles. The Field Dependents included seven freshmen, four sophomores, three juniors, and one senior. Five Field Dependents were from the Faculty of Arts, nine from the Sciences, and one was undeclared. The Field Independents were composed of six freshman, one sophomore, four juniors, and three seniors. Seven of the Field

Independents were studying in the Faculty of Arts and seven in the Sciences.

3.3.2 TEXTS

The three texts were articles from *Scientific American* (Heller, 2015; Ricard, Lutz, & Davidson, 2014; Summa & Turek, 2015). The texts were selected for their potential general interest to subjects from various backgrounds. The texts were neither difficult to read nor did they require specific subject knowledge. Subjects with a science background, however, may have had enough knowledge about the topics to answer the comprehension questions without fully understanding the text. In a pilot study, both subjects agreed that the texts were easy to read and interesting. The articles were marked up in the Hypertext Markup Language. Explanatory aids, such as figures, were removed to reduce the effect of confounding variables. The title, byline, and headings, however, were retained. Each article was about three thousand words in length.

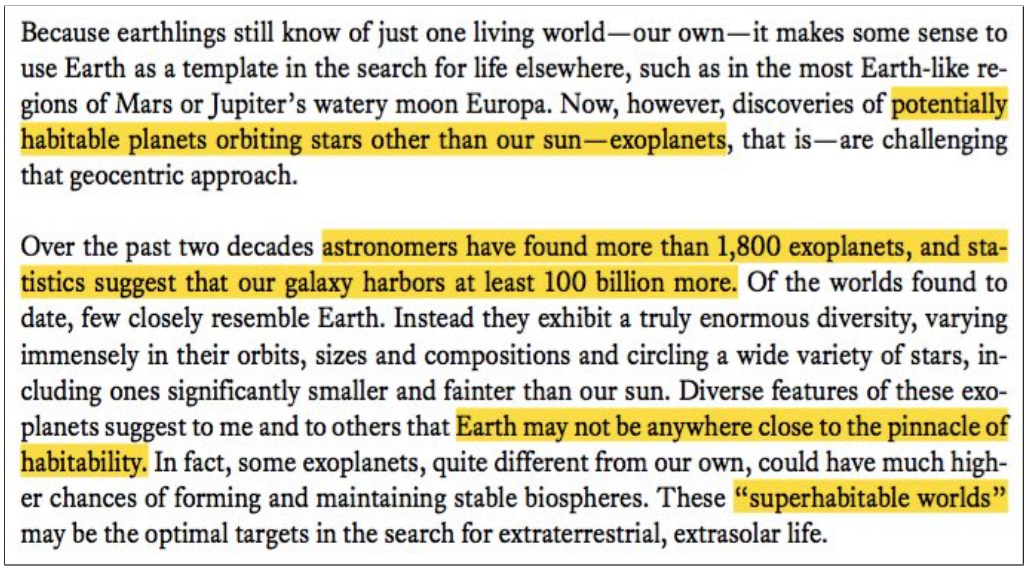
3.3.3 CONDITIONS

Three conditions were used: no highlighting (control), relevant highlighting, and irrelevant highlighting. The highlights in the relevant and irrelevant highlighting conditions were created by the three experimenters. Figures 3.1 and 3.2 provide examples of relevant and irrelevant highlights. Each experimenter highlighted the three texts separately and then assessed the annotations of the other experimenters. The goal of the relevant highlighting condition was to emphasize passages that con-

tained important concepts or facts to the overall meaning of the text. A relevant highlight was created when at least two of the experimenters highlighted the same passage. In this way, the experimenters' consensus of what was an important passage produced each relevant highlight. The irrelevant highlights were similarly created by experimenters: those that focused on points peripheral to the main themes of the text, but were not obviously irrelevant, were selected. Relevant and irrelevant highlights are mutually exclusive – that is, a highlighted sentence was either relevant or irrelevant. Because the initial consensus highlights emphasized too much of the text, they were then trimmed down. The trimmed highlights emphasized the same concepts, but de-emphasized content that was not essential; usually the beginning or ending of the highlight. For example, rather than emphasize a whole sentence, a highlight was reduced to a phrase. Ten to fifteen percent of the text was highlighted in the final relevant and irrelevant conditions in accordance with guidelines noted in Lorch (1989), Lorch et al. (1995).

3.3.4 QUIZZES

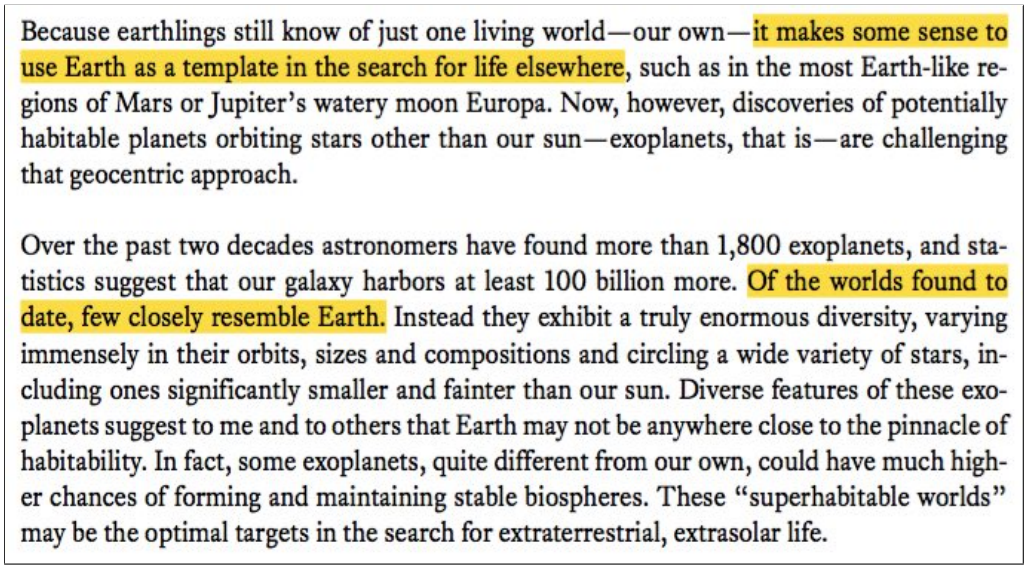
A comprehension quiz was created for each text. An example quiz is provided by Appendix A. The quizzes included nine multiple choice questions, eight Sentence Verification Technique, or SVT, (Royer et al., 1987) questions, and an open-ended summary question. Half the SVT questions tested relevantly highlighted content, and the other half tested non-highlighted content. The SVT was developed to create a method that accurately measured comprehension (Royer et al., 1987; Royer et al., 1979). The SVT assumes, like the Construction-Integration model, that com-



Because earthlings still know of just one living world—our own—it makes some sense to use Earth as a template in the search for life elsewhere, such as in the most Earth-like regions of Mars or Jupiter's watery moon Europa. Now, however, discoveries of potentially habitable planets orbiting stars other than our sun—exoplanets, that is—are challenging that geocentric approach.

Over the past two decades astronomers have found more than 1,800 exoplanets, and statistics suggest that our galaxy harbors at least 100 billion more. Of the worlds found to date, few closely resemble Earth. Instead they exhibit a truly enormous diversity, varying immensely in their orbits, sizes and compositions and circling a wide variety of stars, including ones significantly smaller and fainter than our sun. Diverse features of these exoplanets suggest to me and to others that Earth may not be anywhere close to the pinnacle of habitability. In fact, some exoplanets, quite different from our own, could have much higher chances of forming and maintaining stable biospheres. These "superhabitable worlds" may be the optimal targets in the search for extraterrestrial, extrasolar life.

Figure 3.1: A relevantly highlighted passage in "Better than earth" by Heller (2015).



Because earthlings still know of just one living world—our own—it makes some sense to use Earth as a template in the search for life elsewhere, such as in the most Earth-like regions of Mars or Jupiter's watery moon Europa. Now, however, discoveries of potentially habitable planets orbiting stars other than our sun—exoplanets, that is—are challenging that geocentric approach.

Over the past two decades astronomers have found more than 1,800 exoplanets, and statistics suggest that our galaxy harbors at least 100 billion more. Of the worlds found to date, few closely resemble Earth. Instead they exhibit a truly enormous diversity, varying immensely in their orbits, sizes and compositions and circling a wide variety of stars, including ones significantly smaller and fainter than our sun. Diverse features of these exoplanets suggest to me and to others that Earth may not be anywhere close to the pinnacle of habitability. In fact, some exoplanets, quite different from our own, could have much higher chances of forming and maintaining stable biospheres. These "superhabitable worlds" may be the optimal targets in the search for extraterrestrial, extrasolar life.

Figure 3.2: An irrelevantly highlighted passage in "Better than earth" by Heller (2015).

prehension is a process of construction that can be "measured by determining if readers or listeners remembered the meaning of something read or heard" (Royer et al., 1987, p. 417).

An SVT test contains a text and a set of sentences. There are four types of SVT sentences: original, paraphrase, meaning change, and distractor. Originals are exact copies of passages from the text. Paraphrases have the same meaning as a passage from the text, but most of the words have been changed. Meaning changes contain most of the same words as a sentence from the text, but mean something else. A distractor is on the same topic, but differs in meaning and wording from any passage in the text. A subject reads the text and then, without looking at the text, "judges each of the text sentences to be 'old' or 'new"' (Royer et al., 1987, p. 415). Old sentences are the same or have the same meaning as the text, while new sentences have a different meaning than the text. If readers have comprehended the meaning of a text, they should be able to judge the test sentences as original, paraphrases, meaning changes, or distractors. Conversely, if readers have not understood the text, they should find this task difficult.

3.3.5 PRE- & POST-TASK QUESTIONNAIRES

A pre-session questionnaire was administered to collect demographic information on the subjects. The subjects' sex, age, year of enrolment, field of study, reading habits, and familiarity with digital reading systems were recorded. A post-session questionnaire was used to ask subjects about their experiences in the study and to specifically ask about how they completed the tasks, in addition to their thoughts

on the highlighting. All questionnaires are included in Appendices B and C.

3.4 PROCEDURE

Subjects were tested in a classroom in small groups of less than ten. Subjects were seated next to each other at a desk in rows of three or four. Each subject had a desktop computer, monitor, and mouse. First, the subjects completed a paper pre-session questionnaire, then the GEFT to measure their degree of Field Dependence-Independence.

Then subjects were told their task was to read three articles on the monitor and complete a paper quiz on each text. The texts were displayed with the Firefox web browser on a twenty-seven inch liquid-crystal display monitor. Subjects were given a task, or scenario: to imagine the articles had been assigned for an upcoming class discussion for which they had limited time to prepare. Subjects were only given five minutes to read each article, which was likely to be insufficient time to read each article line by line from start to finish. This was done to encourage subjects to employ efficient reading strategies, including the use of highlights.

In the pilot study, subjects were given ten minutes to read the texts. With double the reading time, subjects in the pilot study were able to read the whole texts line by line more than once. In a post session interview, the pilot subjects said that they had so much time that they did not feel the need to skim the text using the highlights. With less time, however, they explained that they would be more likely to use the highlights. Because we wanted to study the effects of highlighting, we lowered the

time limit to encourage subjects to look at the highlights.

Subjects were presented the relevantly highlighted article first, the control – without highlights – second, and the irrelevantly highlighted article third. The conditions were not counterbalanced because we wanted to study the subjects' ability to recognize bad highlighting and filter or ignore the poor markings. By presenting the subject with good highlighting first, we hoped they would recognize the value of the annotations. The control acted as a kind of palate cleanser. This order of conditions was also more likely to encourage subjects to use, or at least consider, the highlights. The texts were counterbalanced for ordering effects; subjects were randomly self-assigned to one of six different measure orders. In the pilot study the order of conditions was counterbalanced. Subjects who were presented the irrelevant highlights first found them so bad that they said they either did not use the good highlights in the subsequent condition or were highly skeptical of them. For these reasons, the relevant highlights were presented first.

After reading the article, the subjects were allowed up to seven minutes to complete the comprehension test. Subjects were not allowed to look at the test until they had finished reading the article and were not able to refer to the article once they started the test. Subjects proceeded to the next condition after completing the test. After finishing the three articles and corresponding quizzes, the subjects completed a post-session questionnaire. A complete session usually lasted less than an hour and a half.

3.5 DATA ANALYSIS

The dependent variable was subjects' comprehension, which was measured after reading a text in the three conditions. There were two independent variables. The first was the between-subject variable of Field Dependence-Independence. The second was the within-subjects variable of highlighting condition. Given that comprehension was measured using three types of questions, each was treated as a separate measure and also calculated a summative overall comprehension score.

The study collected qualitative and quantitative data. The pre-session questionnaire collected demographic information about the subjects as well as their reading habits and familiarity with digital reading systems. The post-session questionnaire provided open-ended responses about the study and the highlighting (i.e. qualitative data). The quantitative data analysis consisted of the GEFT and comprehension scores.

All quantitative data were analyzed using the R programming language (R Core Team, 2015). Descriptive statistics were calculated to describe the characteristics of subjects' performance across the conditions. Data were tested for normality and scores were compared using the relevant parametric or nonparametric tests.

The qualitative data were analyzed manually. The summary responses were rated separately by three experimenters. The main points of each article were identified with the help of summaries provided in the original articles. Subjects' summary scores were based on the percentage of these points addressed in their responses. After several rounds of assessment, unanimous agreement was reached

Measure	Significance
Multiple Choice	p=.358
Summary	.112
SVT	.016
Overall	.642

Table 3.1: Bartlett's test of homogeneity of variances for Field Dependents.

Measure	Significance
Multiple Choice	p=.358
Summary	.112
SVT	.016
Overall	.642

Table 3.2: Bartlett's test of homogeneity of variances for Field Independents.

Measure	Significance
Multiple Choice	p=.731
Summary	.039
SVT	.748
Overall	.642

Table 3.3: Bartlett's test of homogeneity of variances for all subjects.

between the three experimenters for all eighty-seven summary responses. The post-session questionnaires were read and comments were grouped by common themes.

Measure	Significance
Multiple Choice	p=.008
Summary	.069
SVT	.014
Overall	.259

Table 3.4: Shapiro-Wilk test of normality for Field Dependents.

Measure	Significance
Multiple Choice	p=.015
Summary	.056
SVT	.019
Overall	.906

Table 3.5: Shapiro-Wilk test of normality for Field Independents.

Measure	Significance
Multiple Choice	p<.001
Summary	.007
SVT	<.001
Overall	.787

Table 3.6: Shapiro-Wilk test of normality for all subjects.

CHAPTER 4

RESULTS

4.1 INTRODUCTION

This chapter presents the results of the data analysis. Comprehension scores are compared for each condition to measure the effects of highlighting on comprehension. The study used three comprehension tests: multiple choice, open-ended summary, and the Sentence Verification Technique (SVT). Because it is unclear which test, or combination of tests, provides the most accurate evaluation of comprehension, the data analysis used the tests individually and together. This resulted in four measures: Overall, Multiple Choice, Summary, and SVT.

This chapter is divided into sections that evaluate the four measures for Field Dependent, Field Independent, and all subjects. Each section presents descriptive statistics, tests the assumptions of t-tests and ANOVAs, and presents the results of the parametric, or nonparametric, tests used. This chapter also examines the post-session questionnaire. Patterns emerged from the assessment of the post-session

questionnaire responses, specifically the division of subjects into two groups: anti- and pro-highlighting.

4.2 OVERALL MEASURE

The Overall comprehension scores are normally distributed (Tables 3.4 to 3.6) and have equal variances (Tables 3.1 to 3.3) for Field Dependents, Field Independents, and all subjects. As a result, parametric tests are used for this measure.

Mean comprehension scores were highest in the control and lowest in the irrelevant highlighting condition for Field Dependents, Field Independents, and all subjects (Tables 4.1 to 4.3). Figures 4.2 and 4.3 provide scatter plots of comprehension in each condition for all subjects.

Figure 4.1 shows that, for Field Dependents, there was a concentration of scores around the median in the control condition. This condition had the smallest difference between the upper and lower quartiles scores, however, the violin plot is elongated because of a score below the lower quartile.

One-way within-groups ANOVAs show differences in comprehension between conditions were not significant for Field Dependents ($F(2,13)=2.984$, $p=.086$), Field Independents ($F(2,12)=0.634$, $p=.547$), or all subjects ($F(2,27)=1.166$, $p=.327$).

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	35.19	53.70	60.65	62.70	77.96	85.00
Control	30.09	59.26	65.28	64.06	71.34	78.33
Inappropriate	34.72	43.98	53.43	54.30	62.04	78.33

Table 4.1: Overall descriptive statistics for Field Dependents.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	28.24	45.37	51.39	55.31	69.44	93.33
Control	20.37	49.42	60.74	57.71	69.63	80.56
Inappropriate	28.24	46.41	51.99	51.80	58.91	74.17

Table 4.2: Overall descriptive statistics for Field Independents.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	28.24	48.61	58.80	59.13	74.17	93.33
Control	20.37	56.39	62.04	60.99	70.93	80.56
Inappropriate	28.24	46.30	53.15	53.09	62.04	78.33

Table 4.3: Overall measure descriptive statistics for Field Dependents and Field Independents.

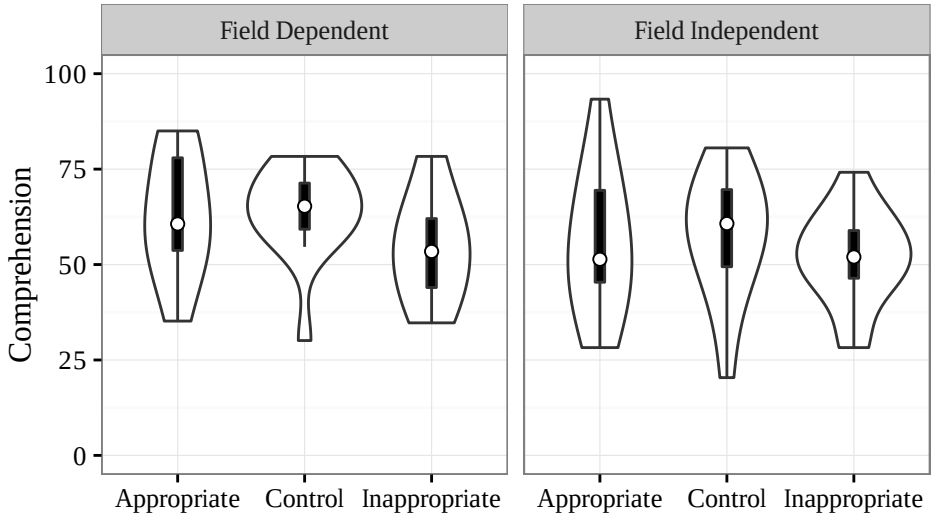


Figure 4.1: Violin plots of comprehension for Field Dependents (left) and Field Independents (right) as scored by the Overall measure. The violin plots include box plots, with a white dot indicating the median comprehension score.

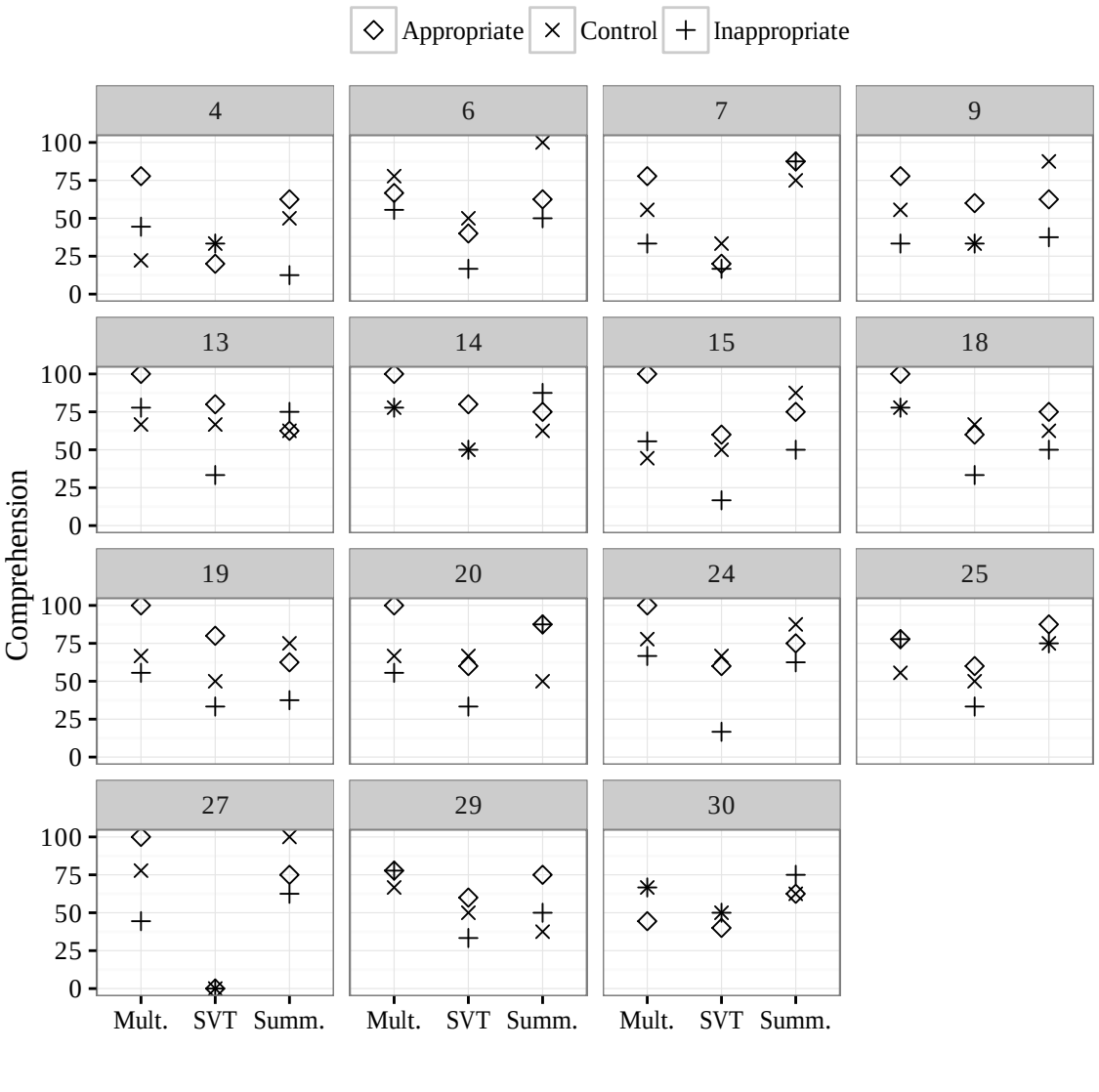


Figure 4.2: A scatter plot of each Field Dependents' comprehension scores across conditions.

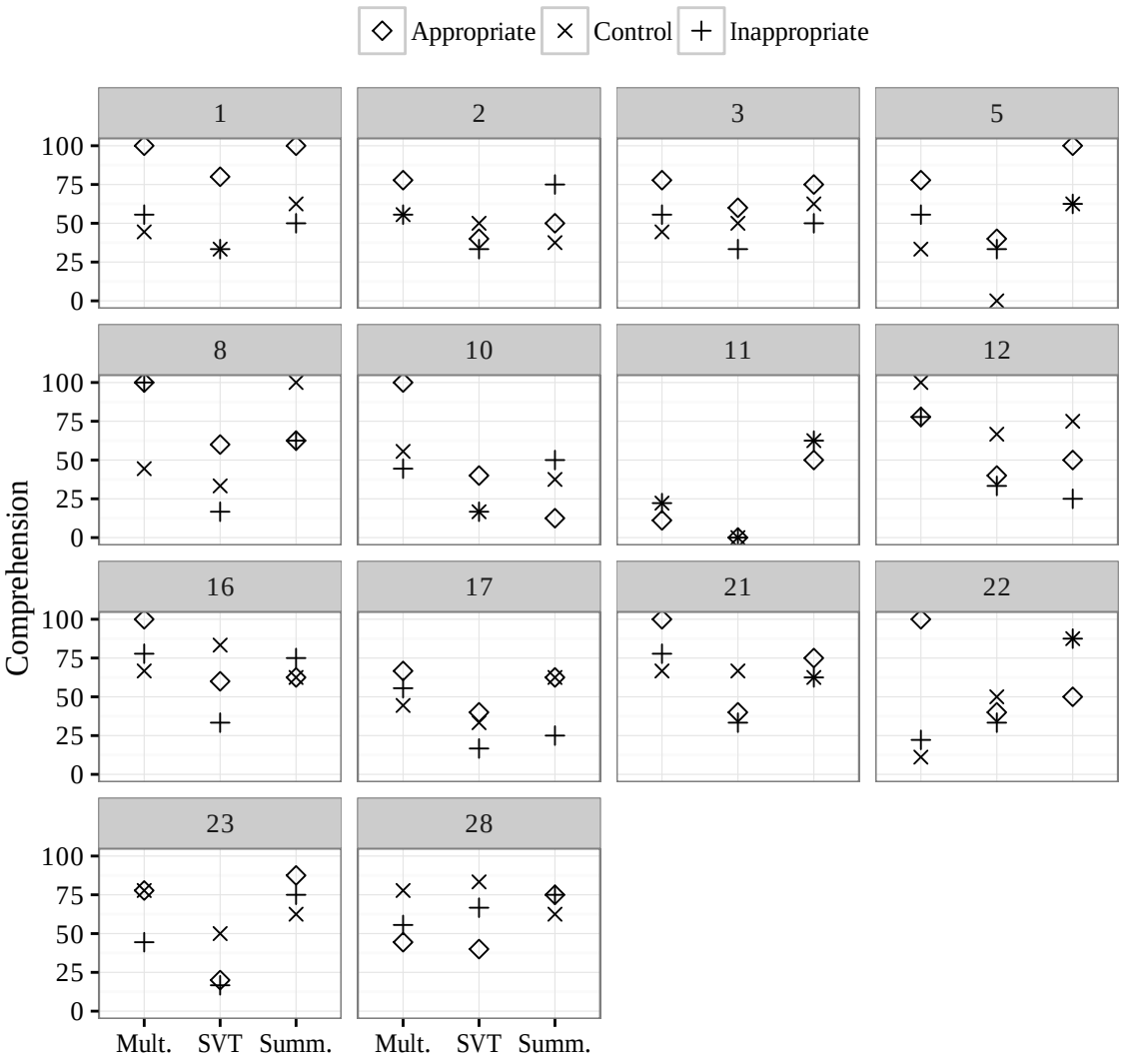


Figure 4.3: A scatter plot of each Field Independents' comprehension scores across conditions.

4.3 MULTIPLE CHOICE MEASURE

The scores of Field Dependents, Field Independents, and all subjects are not normally distributed (Tables 3.4 to 3.6), but have equal variances (Tables 3.1 to 3.3). Because these scores fail to meet the assumptions of parametric tests, nonparametric tests are used for this measure.

Comprehension was highest in the relevant condition for Field Dependents, Field Independents, and all subjects (Tables 4.4 to 4.6). Comprehension was lowest in the irrelevant condition for Field Dependents and in the control condition for Field Independents. For all subjects, there was no difference between mean comprehension scores in the control and irrelevant conditions.

Figure 4.4 shows comprehension was concentrated around the median in all conditions for Field Dependents and Field Independents. Scores below the lower quartile elongated the violin plots of the relevant condition, for Field Dependents and Field Independents, and, for Field Dependents, the control condition. The lowest quartile of the relevant condition is higher than all other conditions for both Field Dependents and Field Independents.

Friedman tests show that condition had a significant effect on comprehension for Field Dependents ($\chi^2(2)=13.345$, $p=.001$), Field Independents ($\chi^2(2)=7.098$, $p=.029$), and all subjects ($\chi^2(2)=19.415$, $p<.001$).

For Field Dependents, *post hoc* analysis with Wilcoxon signed rank tests, with the Bonferroni correction applied, show significant differences between the relevant and control ($p=.017$) and relevant and irrelevant ($p=.015$) conditions, but not

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	44.44	77.78	100	86.67	100	100
Control	22.22	55.56	66.67	63.71	77.78	77.78
Inappropriate	33.33	50.00	55.56	60.00	77.78	77.78

Table 4.4: Multiple choice descriptive statistics for Field Dependents.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	11.11	77.78	77.78	79.37	100	100
Control	11.11	44.44	50.00	53.17	66.67	100
Inappropriate	22.22	47.22	55.56	57.14	72.22	100

Table 4.5: Multiple choice descriptive statistics for Field Independents.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	11.11	77.78	77.78	83.14	100	100
Control	11.11	44.44	66.67	58.62	77.78	100
Inappropriate	22.22	44.44	55.56	58.62	77.78	100

Table 4.6: Multiple choice measure descriptive statistics for Field Dependents and Field Independents.

between the control and irrelevant conditions (p=1.0).

For Field Independents, there were also significant differences between the relevant and control (p=.052) and relevant and irrelevant conditions (p=.022), but not between the control and irrelevant conditions (p=1.0).

A Wilcoxon signed rank test, with the Bonferroni correction applied, shows significant differences between the relevant and control (p=.001) and relevant and irrelevant conditions (p<.001) for all subjects. There was not, however, a statistically significant difference between the control and irrelevant conditions (p=1.0).

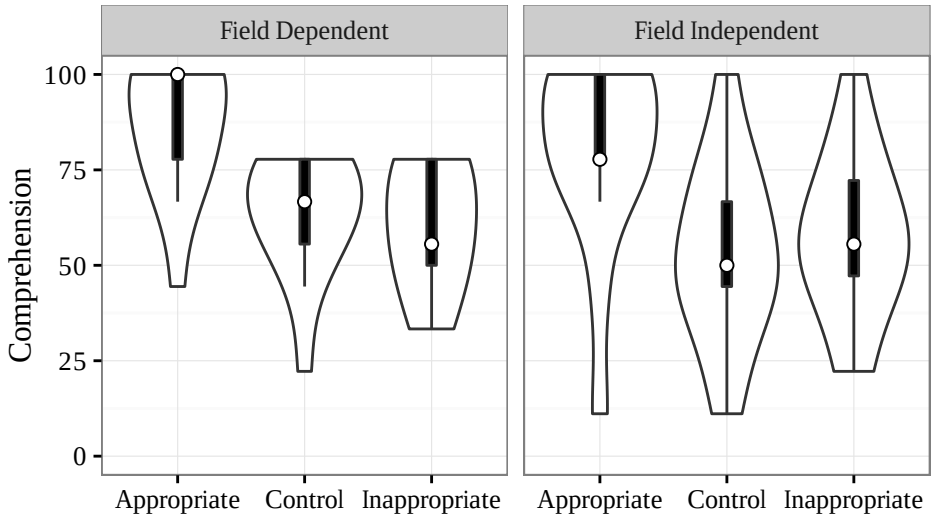


Figure 4.4: Violin plots of comprehension for Field Dependent (left) and Field Independent (right) subjects as scored by the Multiple Choice measure.

4.4 SUMMARY MEASURE

Parametric and nonparametric tests were used for the Summary measure. The comprehension scores for Field Dependents and Field Independents are normally distributed (Tables 3.4 to 3.6) and have equal variances (Tables 3.1 to 3.3). The group composed of all subjects, however, are neither normally distributed nor homogeneous variances.

Comprehension was highest in the relevant condition for Field Dependents and all subjects (Tables 4.7 to 4.9). For Field Independents comprehension was highest in the control condition. Comprehension was lowest in the irrelevant highlighting condition for Field Dependents, Field Independents, and all subjects.

Figure 4.5 shows that the relevant and control condition violin plots for Field De-

pendents are elongated because of a comprehension score below the lower quartile. The relevant and irrelevant condition violin plots for Field Independents were also elongated. The relevant condition was affected by a score above the highest quartile and one below the lowest quartile. The irrelevant condition was also affected by a score above the highest quartile.

One-way within-groups ANOVAs show differences in comprehension between conditions were significant for Field Dependents ($F(2,12)=24.72$, $p<.001$) and Field Independents ($F(2,11)=6.163$, $p=.016$). A Friedman test shows that there was also a significant effect of condition on comprehension for all subjects ($\chi^2(2)=23.529$, $p<.001$).

Post hoc tests, using the Bonferroni correction, show a significant difference, for Field Dependents, between the relevant and irrelevant ($p=.001$) and control and irrelevant ($p=.002$) conditions, but not between the relevant and control conditions ($p=.938$).

For Field Independents, *post hoc* tests, using the Bonferroni correction, show significant differences between the relevant and irrelevant ($p=.043$) and control and irrelevant ($p=.039$) conditions, but not the relevant and control conditions ($p=1.0$).

A Wilcoxon signed rank test, with the Bonferroni correction applied, shows significant differences between the relevant and irrelevant ($p<.001$) and the control and irrelevant ($p=.001$), but not the relevant and control conditions ($p=1.0$) for all subjects.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	0.00	40.00	60.00	52.00	60.00	80.00
Control	0.00	41.66	50.00	47.78	58.34	66.67
Inappropriate	0.00	16.67	33.33	28.89	33.33	50.00

Table 4.7: Summary measure descriptive statistics for Field Dependents.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	0.00	40.00	40.00	42.86	55.00	80.00
Control	0.00	33.33	50.00	44.05	62.50	83.33
Inappropriate	0.00	16.67	33.33	28.57	33.33	66.67

Table 4.8: Summary measure descriptive statistics for Field Independents.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	0.00	40.00	40.00	47.59	60.00	80.00
Control	0.00	33.33	50.00	45.98	66.67	83.33
Inappropriate	0.00	16.67	33.33	28.73	33.33	66.67

Table 4.9: Summary measure descriptive statistics for Field Dependents and Field Independents.

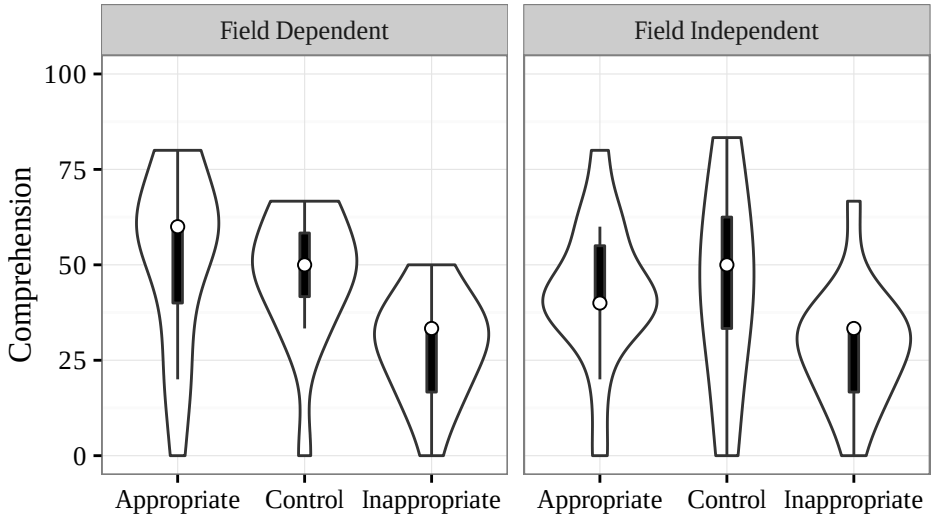


Figure 4.5: Violin plots of comprehension for Field Dependents (left) and Field Independents (right) as scored by the Summary measure.

4.5 SENTENCE VERIFICATION TECHNIQUE MEASURE

The SVT scores were not normally distributed for any of the groupings (Tables 3.4 to 3.6). Only the comprehension scores for Field Dependents and the group composed of all subjects had equal variance (Tables 3.1 to 3.3). Non-parametric tests were used for this measure, because scores failed to meet the assumptions of parametric tests.

Comprehension was highest in the relevant condition and lowest in the irrelevant condition for Field Dependents, Field Independents, and the group composed of all subjects (Tables 4.10 to 4.12). Figure 4.6 shows that Field Dependents' comprehension scores were concentrated around the median in the relevant condition. Comprehension scores had a much greater distribution in the control and irrelevant conditions for Field Dependents. Field Independents' comprehension scores in the control condition were even more heavily concentrated at the median – which was also the upper, middle, and lower quartiles.

Friedman tests show that these differences were not significant for Field Dependents (χ^2(2)=2.655, p=.265), Field Independents (χ^2(2)=0.5, p=.779), or all subjects (χ^2(2)=2.582, p=.275).

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	62.50	62.50	75.00	72.50	75.00	87.50
Control	37.50	62.50	75.00	71.67	87.50	100
Inappropriate	12.50	50.00	62.50	60.00	75.00	87.50

Table 4.10: SVT descriptive statistics for Field Dependents.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	12.50	50.00	62.50	65.18	75.00	100
Control	37.50	62.50	62.50	64.29	62.50	100
Inappropriate	25.00	50.00	62.50	59.82	75.00	87.50

Table 4.11: SVT measure descriptive statistics for Field Independents.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	12.50	62.50	75.00	68.97	75.00	100
Control	37.50	62.50	62.50	68.10	75.00	100
Inappropriate	12.50	50.00	62.50	59.91	75.00	87.50

Table 4.12: SVT measure descriptive statistics for Field Dependents and Field Independents.

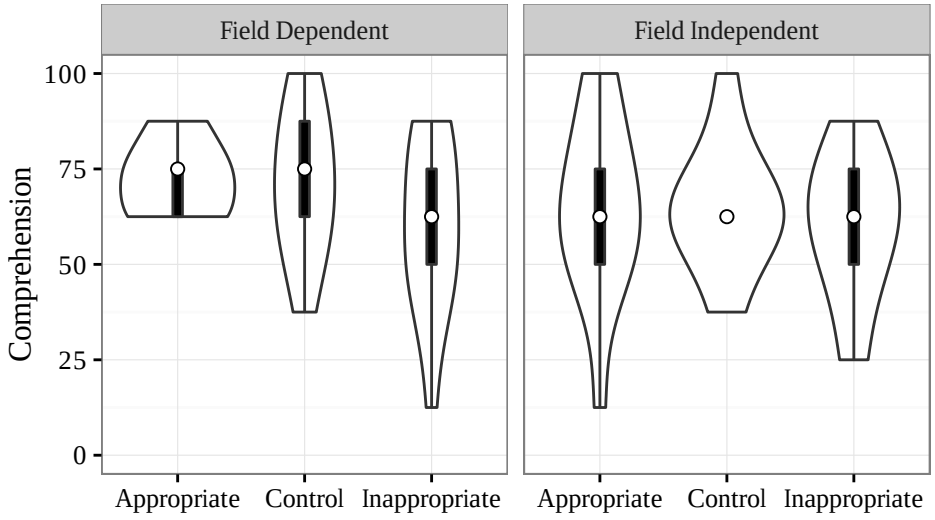


Figure 4.6: Violin plots of comprehension for Field Dependents (left) and Field Independents (right) as scored by the SVT measure. Note that the box plot in the control condition for Field Independents is concealed by the white median score dot. In this condition, the value of the upper quartile, median, and lower quartile were the same.

4.6 POST-SESSION QUESTIONNAIRE

The responses from the post-session questionnaire were analyzed for opinions regarding the passive highlighting used in the relevant and irrelevant conditions. Twenty one of the subjects (ten Field Dependents) said some, or all, of the highlights were helpful. These subjects formed the pro-highlighting group. Eight (five Field Dependents) of the twenty nine subjects said that none of the highlights were helpful. These subjects formed the anti-highlighting group.

The comprehension scores for both the anti- and pro-highlighting groups had homogeneous variances and were normally distributed (Tables 4.13 and 4.14).

4.6.1 ANTI-HIGHLIGHTING SUBJECTS

Eight subjects (five Field Dependents) said none of the highlights were helpful. Within this group there was a split between those that claimed to have ignored all

Group	Significance
Anti-Highlighting	p=.120
Pro-Highlighting	.748

Table 4.13: Bartlett's test of homogeneity of variances for the Anti- and Pro-Highlighting groups.

Group	Significance
Anti-Highlighting	p=.676
Pro-Highlighting	.217

Table 4.14: Shapiro-Wilk test of normality the Anti- and Pro-Highlighting groups.

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	35.19	39.70	49.54	55.57	64.31	93.33
Control	30.09	44.33	63.52	56.82	67.80	75.93
Inappropriate	34.72	49.70	52.41	52.32	57.80	63.43

Table 4.15: Descriptive Statistics for the Anti-Highlighting group.

the highlights and those that looked at them, but did not find them useful. Several subjects used words such as "annoying" and "distracting" to describe the highlights.

Two subjects, both Field Independents, said that they do not value passive highlighting. Subject 10 said "I didn't really notice the highlights, because I was focused on reading. I believe highlights are most helpful, when done by the reader." Subject 23 added, "I didn't find the highlighting useful. If I wasn't the one to highlight, the highlight just gets in the way".

Four of the subjects in this group (three Field Dependents) said they ignored the highlights. Subject 9, a Field Dependent, said "did not pay much attention to the highlights, felt they were distracting as they pulled my focus away from the article when I was reading." Subject 4 "noticed them but didn't really analyze them".

For subjects in the anti-highlighting group, mean comprehension scores were highest in the control and lowest in the irrelevant highlighting condition (Table 4.15). A one-way within-groups ANOVA shows the differences in comprehension scores across conditions were not significant ($F(2,6)=0.151$, $p=.863$).

4.6.2 PRO-HIGHLIGHTING SUBJECTS

Of the subjects in the pro-highlighting group, thirteen (six Field Dependents) found that the quality of the highlights varied. Subject 21, a Field Independent, said, "I thought the highlights were distributed between helpful and useless". Subject 14, a Field Dependent, added, "I found the highlights to be very helpful, especially in the first article [the relevant condition], the third article highlights [the irrelevant condition] made me skim the surrounding information". A Field Independent, subject 5, said "While skimming, I felt like I had to read them [the highlights], which was obnoxious when they weren't helpful".

Eight subjects (four Field Dependents) made no distinction between the quality of the highlights across the relevant and irrelevant conditions. Subject 28, a Field Independent, reflected "Reading the highlights helped to get the gist of what [the] article was talking about but didn't help with the little details". Subject 15, a Field Dependent, added "I tended to focus on the highlights".

Mean comprehension scores were highest in the control and lowest in the irrelevant highlighting condition for subjects in the pro-highlighting group (Table 4.16). A one-way within-groups ANOVA shows a significant difference in comprehension between conditions for subjects in the pro-highlighting group ($F(2,19)=3.787$, $p=.041$). *Post hoc* tests, using the Bonferroni correction, show significant differences between the control and irrelevant conditions ($p=.039$), but not between the relevant and control ($p=1.0$) or the relevant and irrelevant conditions ($p=.235$).

Condition	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Appropriate	28.24	50.46	60.65	60.49	74.17	85.00
Control	20.37	58.33	62.04	62.58	71.67	80.56
Inappropriate	28.24	45.83	53.15	53.39	62.04	78.33

Table 4.16: Descriptive Statistics for the Pro-Highlighting group.

4.7 SUMMARY

There were significant effects of condition on comprehension scores of both Field Dependents and Field Independents in the Multiple Choice and Summary measures.

In the Multiple Choice measure, comprehension scores in the relevant highlighting condition were significantly higher than the control for Field Dependents, Field Independents, and all subjects, indicating that the relevant highlights supported comprehension of factual information. Comprehension scores in the irrelevant highlighting condition, however, were significantly lower than the control for Field Dependents, Field Independents, and all subjects. This suggests that irrelevant highlights impair comprehension.

The post-session questionnaire responses can be used to group subjects into anti- and pro-highlighting groups. Using the Overall comprehension measure, there was not a statistically significant difference in comprehension scores across conditions for anti-highlighting subjects. For pro-highlighting subjects, however, comprehension scores in the irrelevant condition were significantly lower than the control.

CHAPTER 5

DISCUSSION

5.1 INTRODUCTION

This chapter reviews the results of the study to confirm or reject the four *a priori* hypotheses. This chapter also reviews two *post hoc* hypotheses, which were defined after reviewing the post-session questionnaire responses. These responses were used to form four *post hoc* groups, anti-highlighting and pro-highlighting. The first and second *post hoc* hypotheses concern the anti-highlighting group. The first is that the reading comprehension of these would not be affected by relevant highlights. The second *post hoc* hypothesis states that subjects in the pro-highlighting group would also be unaffected by irrelevant highlights. Subjects in the pro-highlighting group would be more likely to focus on the highlights and, perhaps, less likely to question their relevance. The third and fourth *post hoc* hypotheses concern the pro-highlighting group. The third *post hoc* hypothesis is that the reading comprehension of subjects in the pro-highlighting group would be positively affected by

relevant highlights. The fourth *post hoc* hypothesis states that irrelevant highlights would negatively affect comprehension. Finally, this chapter discusses the difficulty of measuring comprehension. The study used three different comprehension tests; each test found different results, suggesting that they may be measuring different levels of comprehension or, perhaps, different reading outcomes.

5.2 EFFECTS OF PRE-EXISTING HIGHLIGHTING

The study used three comprehension tests: multiple choice questions, open-ended summary questions, and Sentence Verification Technique (SVT) questions. Because it is unclear which of these tests are best suited to measure comprehension, four measures (the individual test scores and the total score of all three) are used in the study. These measures are used to compare comprehension across all conditions.

The first two *a priori* hypotheses concern relevant highlights. The first states that the relevant highlighting condition would have a positive effect on reading comprehension for Field Dependents. Since they rely on external structuring, Field Dependents would most likely have followed the relevant highlighting, guiding them to the right information within the document. The second hypothesis is that Field Independents would be unaffected by relevant highlighting, not needing them because they use internal processes to structure and make sense of information. These individuals should be able to identify the most relevant passages in a text without passive highlighting, so their comprehension scores should not have differed significantly between the relevant highlighting and control conditions.

The third and fourth *a priori* hypotheses concern irrelevant highlights. The third states that Field Dependents would be negatively affected by irrelevant highlights. These subjects are more likely to rely on all external cues, including irrelevant highlights that draw their attention away from the right information. The fourth *a priori* hypothesis states that Field Independents will be unaffected by irrelevant highlights. They are able to assess external cues and ignore poor ones.

Only the multiple choice measure found that relevant highlighting had a statistically significant positive effect on reading comprehension. This effect was found for both Field Dependents and Field Independents. The three other measures did not find an effect of relevant highlighting on comprehension for either Field Dependents or Field Independents. The result of the multiple choice measure provides weak support for the first *a priori* hypothesis that relevant highlighting could increase comprehension for Field Dependents. It also suggests that relevant highlights may not provide enough guidance for Field Dependents to find the right information. The failure of the other three measures to find an effect on Field Independents' comprehension provides strong support for the second hypothesis that Field Independents will be unaffected by relevant highlights.

The third and fourth *a priori* hypotheses relate to the effects of irrelevant highlights. The third hypothesis states that Field Dependents' comprehension would be negatively affected by irrelevant highlighting. Similar to the first hypothesis, we expected these subjects to focus on highlighted content. Unlike the relevant highlighting, however, irrelevant highlights would guide subjects away from important and towards unimportant content. The fourth *a priori* hypothesis predicted that

Field Independents would be unaffected by the irrelevant highlighting. While these subjects would notice the highlights, their ability to cognitively restructure and attend to relevant cues would cause them to suspect their usefulness. As a result, these individuals would ignore the highlights and create their own interpretations of the text.

Only the summary measure found that that irrelevant highlighting had a statistically significant negative effect on comprehension. This effect was found for both Field Dependents and Field Independents. The other three measure did not find an effect of irrelevant highlighting on comprehension for either Field Dependents or Field Independents. The result of the summary measure provides weak support for the third hypothesis. This may suggest that Field Dependents used the irrelevant highlights as external cues to structure their reading. Three of the four measures failed to find a significant effect of irrelevant highlighting on comprehension. The failure of the other three measures to find an effect on Field Independents' comprehension provides strong support for the fourth hypothesis.

5.3 MEASURES OF COMPREHENSION

The distinctions between the different components of the reader's mental representation of a text are important when considering how to evaluate comprehension. Kintsch (1998) suggests that tests should be directed at these different components of comprehension. Many of the previous studies on highlighting, however, fail to use a theory of comprehension. As a result, they used tests that may insufficiently

measure comprehension, measure different aspects of comprehension, or measure, possibly, other abilities or reading outcomes (Drum et al., 1981; Ozuru et al., 2013; Royer et al., 1987; Tuinman, 1973). To avoid these limitations, the study used three comprehension tests: multiple choice questions, an open-ended summary question, and SVT questions.

Ozuru et al. (2013) suggest that the tasks the various tests entail rely on different aspects of comprehension. Simple tests, such as multiple choice, may measure comprehension at shallow levels. These tests use questions that rely heavily on the reader's ability to recall relevant information rather than understand it. Complex comprehension tests, such as open-ended summary questions, requires the reader to connect multiple text-derived idea units together with their knowledge to understand the text.

In the study, there were differences between the comprehension tests. The multiple choice questions found that relevant highlighting had a positive effect on all subjects, both Field Dependent and Field Independent. The SVT found no significant effect of relevant or irrelevant highlighting on comprehension. The open-ended summary found a significant negative effect of irrelevant highlighting on comprehension for both Field Dependents and Field Independents.

5.4 ANTI- & PRO-HIGHLIGHTING

The subjects were divided into anti- and pro-highlighting groups based on their responses in the post-session questionnaire. Subjects in the anti-highlighting group

found that the passive highlights had no value, while the subjects in the pro-highlighting group felt that the passive highlights had at least some value. Each group included both Field Dependents and Field Independents. There was no significant effect of relevant or irrelevant highlights on subjects in the anti-highlighting group. This supports the first and second *post hoc* hypotheses. Relevant highlights had no effect on the comprehension of subjects in the pro-highlighting group, which fails to support the third *post hoc* hypothesis. Irrelevant highlights, however, had a negative effect, which fails to support the fourth *post hoc* hypothesis.

5.5 SUMMARY

Relevant highlights had a positive effect on comprehension in one of the four measures. This effect was found for both Field Dependents and Field Independents. The summary measure found that irrelevant highlights had a negative effect on comprehension. This effect was found for both Field Dependents and Field Independents. These findings suggests that relevant highlights have limited value as cues for comprehension. Passive highlights were found to have both positive and negative effects on comprehension. For this reason, readers should be wary of texts with passive annotations.

CHAPTER 6

CONCLUSION

6.1 SUMMARY

This study furthers our knowledge of the effects of highlighting on reading comprehension. Participants were divided by their cognitive styles based on their degree of Field Dependence-Independence (Witkin et al., 1962). The study found that passive highlights have significant effects. Both Field Dependents, who rely on external cues to structure and process information, and Field Independents, who use internal cues, were positively affected by relevant highlights and negatively affected by irrelevant highlights. This study contributed to the theory of reading within the field of library and information studies. The Construction-Integration model (Kintsch, 1988) informed the selection of reading tests that are most likely to measure comprehension. Differences were found between measures of comprehension used in the study. In addition to theoretical contributions, the findings that passive highlights affect readers also have practical applications in the design of digital reading

systems.

6.2 LIMITATIONS

A significant limitation of the study is that it only measured reading comprehension. Measuring reading processes could have provided more insight into how passive highlights affect readers. While significant effects of passive highlighting on comprehension were found for both Field Dependents and Field Independents, it is unclear how relevant and irrelevant highlights affect reading behaviours. The study did not measure reading processes, such as through eye-tracking. For this reason, it is difficult to discern how passive highlights affected subjects' reading behaviour.

The within-subjects design of the study required that a limited selection of comprehension tests could be used, so that the study did not go on for too long. Using a between-subjects design could have benefits. For example, more comprehension tests could be used. This would provide more data on subjects' comprehension. Also, longer texts could be used, and subjects could be given more time to complete the task to simulate reading situations other than the one used in the study.

6.3 FUTURE WORK

The results of the study suggest that highlights could improve reading comprehension. It is, however, unclear how passive highlights affect reading processes. Measuring subjects' eye movements would provide a great amount of data that could

show how passive highlights affect reading behaviours. Measuring other reading processes and outcomes should also be considered.

Research on passive highlighting could dovetail with work on collecting and analyzing passive highlights in digital reading systems, such as Marshall (2000) and F. Shipman et al. (2003). Future work could study if, and how, passive highlights could be classified by reading task. If an algorithm or heuristic could identify which highlights are relevant or irrelevant for a given task, we could direct readers to the highlights most likely to aid their understanding of the material for that task.

The Construction-Integration model suggests that multiple components support the process of comprehension. It is unclear, however, which components are being measured by which tests of comprehension. A thorough guide to measuring comprehension is needed.

6.4 IMPLICATIONS

The study could inform other attempts to measure comprehension. The study found that Field Dependents and Field Independents were positively affected by relevant highlights and negatively affected by irrelevant highlights. These results suggest that highlights can support the information processing needs of readers with different cognitive styles. The study made two contributions. First, no previous work on relevant and irrelevant highlights had studied their effects on readers with different cognitive styles. Second, the study used measures of comprehension in accordance with the Construction-Integration model (Kintsch, 1988) to measure different com-

ponents of comprehension. The results of the study provide an argument for using comprehension tests other than multiple choice, which has been used in most previous studies on this topic. The findings may also supplement best practices in the design of digital reading systems. When relevant to a reader's task, it appears that appears that highlighting may be a useful reading tool to improve comprehension.

REFERENCES

Adler, M. J. & Van Doren, C. (1972). *How to read a book*. New York, NY: Simon & Schuster.

Agosti, M. & Ferro, N. (2007). A formal model of annotations of digital content. ACM *Transactions on Information Systems*, *26*(1), 244–255.

Agosti, M., Ferro, N., Frommholz, I., & Thiel, U. (2004). Annotations in digital libraries and collaboratories – facets, models and usage. In *Research and advanced technology for digital libraries* (pp. 244–255). Berlin, Germany: Springer.

Ausburn, L. J. & Ausburn, F. B. (1978). Cognitive styles: Some information and implications for instructional design. *Educational Communication and Technology*, *26*(4), 337–354.

Baron, D. (2009). *A better pencil: Readers, writers, and the digital revolution*. Oxford, England: Oxford University Press.

Bélanger, M. (2010). *The annotative practices of graduate students: Tensions & negotiations fostering an epistemic practice* (Master's thesis, University of Toronto, Toronto, Canada).

Berger, E. & Goldberger, L. (1979). Field dependence and short-term memory. *Perceptual and Motor Skills*, *49*, 87–96.

Butcher, K. R. & Kintsch, W. (2003). Text comprehension and discourse processing. *Handbook of Psychology*, *21*, 575–595.

Cao, Y. (2006). *Effects of field dependent-independent cognitive styles and cueing strategies on students' recall and comprehension* (Doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA).

Chi, E. H., Gumbrecht, M., & Hong, L. (2007). Visual foraging of highlighted text: An eye-tracking study. In *Human-computer interaction:* HCI *intelligent multimodal interaction environments* (pp. 589–598). Berlin, Germany: Springer.

Constantopoulos, P., Doerr, M., Theodoridou, M., & Tzobanakis, M. (2005). On information organization in annotation systems. In *Intuitive human interfaces for organizing and accessing intellectual assets* (pp. 189–200). Berlin, Germany: Springer.

Davis, J. K. & Frank, B. M. (1979). Learning and memory of field independent-dependent individuals. *Journal of Research in Personality*, *13*(4), 469–479.

Demick, J. (2014). *Group Embedded Figures Test: Manual*. Mind Garden, Inc. Menlo Park, CA.

Dillon, A. (2003). *Designing usable electronic text*. New York, NY: Taylor & Francis.

Drum, P. A., Calfee, R. C., & Cook, L. K. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, *16*(4), 486–514.

Foster, J. & Coles, P. (1977). An experimental study of typographic cueing in printed text. *Ergonomics*, *20*(1), 57–66.

Fowler, R. L. & Barker, A. S. (1974). Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology*, *59*(3), 358–364.

Gaddy, M. L. (1996). *Studying from highlighted material: The role of color in memory for text* (Master's thesis, San Jose State University, San Jose, CA).

Goodenough, D. R. (1976). The role of individual differences in field dependence as a factor in learning and memory. *Psychological Bulletin*, *83*(4), 675–694.

Hartley, J., Bartlett, S., & Branthwaite, A. (1980). Underlining can make a difference–sometimes. *The Journal of Educational Research*, *73*(4), 218–224.

Haslhofer, B., Simon, R., Sanderson, R., & Van de Sompel, H. (2011). The Open Annotation Collaboration (OAC) model. In *Proceedings of the 2011 work-*

shop on multimedia on the web (pp. 5–9). Washington, DC: IEEE Computer Society.

Heller, R. (2015). Better than earth. *Scientific American*, *312*(1), 32–39.

Jackson, H. J. (2001). *Marginalia: readers writing in books*. New Haven, CT: Yale University Press.

Kent-Davis, J. & Cochran, K. F. (1989). An information processing view of field dependence-independence. *Early Child Development and Care*, *51*(1), 31–47.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Kintsch, W. & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394.

Klare, G. R., Mabry, J. E., & Gustafson, L. M. (1955). The relationship of patterning (underlining) to immediate retention and to acceptability of technical material. *Journal of Applied Psychology*, *39*(1), 40–42.

Leighton, J. & Gierl, M. (2011). *The learning sciences in educational assessment: The role of cognitive models*. Cambridge, England: Cambridge University Press.

Liu, M. & Reed, W. M. (1994). The relationship between the learning strategies and learning styles in a hypermedia environment. *Computers in Human Behavior*, *10*(4), 419–434.

Lorch, R. F., Jr. (1989). Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review*, *1*(3), 209–234.

Lorch, R. F., Jr, Pugzles Lorch, E., & Klusewitz, M. A. (1995). Effects of typographical cues on reading and recall of text. *Contemporary Educational Psychology*, *20*(1), 51–64.

MacMullen, W. J. (2005). *Annotation as process, thing, and knowledge: Multi-domain studies of structured data annotation*. School of Information and Library Science, University of North Carolina at Chapel Hill. Chapel Hill, NC.

Marshall, C. C. (1997). Annotation: From paper books to the digital library. In *Proceedings of the second ACM international conference on digital libraries* (pp. 131–140). New York, NY: ACM.

Marshall, C. C. (1998). Toward an ecology of hypertext annotation. In *Proceedings of the ninth ACM conference on hypertext and hypermedia* (pp. 40–49). New York, NY: ACM.

Marshall, C. C. (2000). The future of annotation in a digital (paper) world. In *Successes & failures of digital libraries (papers presented at the 35th annual clinic on library applications of data processing, March 22–24, 1998)* (pp. 97–117). Champaign, IL: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

Marshall, C. C. (2009). *Reading and writing the electronic book*. San Rafael, CA: Morgan & Claypool Publishers.

Marshall, C. C., Price, M. N., Golovchinsky, G., & Schilit, B. N. (1999). Introducing a digital library reading appliance into a reading group. In *Proceedings of the fourth ACM conference on digital libraries* (pp. 77–84). New York, NY: ACM.

Messick, S. (1976). *Individuality in learning*. San Francisco, CA: Jossey-Bass Publishers.

Nelson, L., Held, C., Pirolli, P., Hong, L., Schiano, D., & Chi, E. H. (2009). With a little help from my friends: Examining the impact of social annotations in sensemaking tasks. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1795–1798). New York, NY: ACM.

Nist, S. L. & Hogrebe, M. C. (1987). The role of underlining and annotating in remembering textual information. *Literacy Research and Instruction*, *27*(1), 12–25.

O'Hara, K. & Sellen, A. (1997). A comparison of reading paper and on-line documents. In *Proceedings of the* ACM SIGCHI *conference on human factors in computing systems* (pp. 335–342). New York, NY: ACM.

Ovsiannikov, I. A., Arbib, M. A., & McNeill, T. H. (1999). Annotation technology. *International Journal of Human-Computer Studies*, *50*(4), 329–362.

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(3), 215–227.

Pearson, J., Buchanan, G., & Thimbleby, H. (2011). The reading desk: Applying physical interactions to digital documents. In *Proceedings of the* SIGCHI *conference on human factors in computing systems* (pp. 3199–3202). New York, NY: ACM.

Pearson, J., Buchanan, G., & Thimbleby, H. (2013). *Designing for digital reading*. San Rafael, CA: Morgan & Claypool Publishers.

Pearson, J., Buchanan, G., Thimbleby, H., & Jones, M. (2012). The digital reading desk: A lightweight approach to digital note-taking. *Interacting with Computers*, *24*(5), 327–338.

Peterson, S. E. (1991). The cognitive functions of underlining as a study technique. *Literacy Research and Instruction*, *31*(2), 49–56.

Price, M. N., Schilit, B. N., & Golovchinsky, G. (1998). XLibris: The active reading machine. In CHI *98 conference summary on human factors in computing systems* (pp. 22–23). New York, NY: ACM.

R Core Team. (2015). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ricard, M., Lutz, A., & Davidson, R. J. (2014). Mind of the meditator. *Scientific American*, *311*(5), 38–45.

Royer, J. M., Greene, B. A., & Sinatra, G. M. (1987). The Sentence Verification Technique: A practical procedure for testing comprehension. *Journal of Reading*, 414–422.

Royer, J. M., Hastings, C. N., & Hook, C. (1979). A Sentence Verification Technique for measuring reading comprehension. *Journal of Literacy Research*, *11*(4), 355–363.

Schilit, B. N., Golovchinsky, G., & Price, M. N. (1998). Beyond paper: Supporting active reading with free form digital ink annotations. In *Proceedings of the* SIGCHI *conference on human factors in computing systems* (pp. 249–256). New York, NY: ACM Press/Addison-Wesley Publishing Co.

Schilit, B. N., Price, M. N., Golovchinsky, G., Tanaka, K., & Marshall, C. C. (1999). As we may read: The reading appliance revolution. *Computer*, *32*(1), 65–73.

Sellen, A. & Harper, R. H. R. (2001). *The myth of the paperless office*. Cambridge, MA: MIT Press.

Shipman, F., Price, M., Marshall, C. C., & Golovchinsky, G. (2003). Identifying useful passages in documents based on annotation patterns. In *Research and advanced technology for digital libraries* (pp. 101–112). Berlin, Germany: Springer.

Shipman, S. & Shipman, V. C. (1985). Cognitive styles: Some conceptual, methodological, and applied issues. *Review of Research in Education*, 229–291.

Silvers, V. L. & Kreiner, D. S. (1997). The effects of pre-existing inappropriate highlighting on reading comprehension. *Literacy Research and Instruction*, *36*(3), 217–223.

Summa, K. C. & Turek, F. W. (2015). The clocks within us. *Scientific American*, *312*(2), 50–55.

Tinajero, C. & Páramo, M. F. (1997). Field dependence-independence and academic achievement: A re-examination of their relationship. *British Journal of Educational Psychology*, *67*(2), 199–212.

Tuinman, J. J. (1973). Determining the passage dependency of comprehension questions in 5 major tests. *Reading Research Quarterly*, *9*(2), 206–223.

von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung*, *18*(1), 299–342.

Ward, J. (2015). *The perfection of the paper clip: Curious tales of invention, accidental genius, and stationery obsession*. New York, NY: Touchstone.

Witkin, H. A., Dyk, R. B., Fattuson, H. F., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation: Studies of development*. New York, NY: Wiley.

Witkin, H. A. & Goodenough, D. R. (1977). Field dependence revisited. ETS *Research Bulletin Series*, *1977*(2), i–53.

Witkin, H. A. & Goodenough, D. R. (1981). Cognitive styles: Essence and origins Field dependence and field independence. *Psychological issues*, (51), 1–141.

Witkin, H. A., Goodenough, D. R., & Oltman, P. K. (1979). Psychological differentiation: Current status. *Journal of Personality and Social Psychology*, *37*(7), 1127–1145.

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1975). Field-dependent and field-independent cognitive styles and their educational implications. ETS *Research Bulletin Series*, *1975*(2), 1–64.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A manual for the embedded figures tests*. Palo Alto, CA: Consulting Psychologists Press.

Wolfe, J. (2002). Annotation technologies: A software and research review. *Computers and Composition*, *19*(4), 471–497.

World Wide Web Consortium. (2015). *Web annotation data model*. Retrieved from http://www.w3.org/TR/annotation-model/

Worley, G. M. (1999). *The effects of highlight color on immediate recall in subjects of different cognitive styles* (Doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA).

APPENDIX A

BETTER THAN EARTH QUIZ

Matching: Match the term with its description by placing the letter of the definition in the space preceding the term.

Terms	Definitions
1. _____ M Dwarfs and K Dwarfs	A. Jupiter's watery moon
2. _____ Habitable Zone	B. Potentially habitable planets orbiting stars other than our sun
3. _____ Exoplanets	C. A slender region where a star's light is neither too intense nor too weak
4. _____ Red Giant	D. A star that has exhausted its hydrogen fuel and has begun fusing more energetic helium in its core
5. _____ Ganymede	E. The biggest moon in our solar system
6. _____ Tidal Heating	F. Forces that cause crust to flex back and forth, creating friction
7. _____ Europa	G. Stars that are smaller, dimmer and much longer-lived than our sun
8. _____ Archipelago World	H. Flattened planetary landscape with shallow seas dotted with island chains
9. _____ Planet Kepler–186f	I. Larger in diameter than earth and probably rocky, orbiting in the habitable zone of its M dwarf star

True-false: Indicate whether the following statements are true to the meaning of the text. If the statement represent the meaning of the text, write a **T** for *true* in the space preceding the statement. If not, write an **F** for *false*.

10. _____ The identification of planets that could support life in solar systems other than our own has challenged the earth-centred perspective.

11. _____ Earth also has a life-friendly size: big enough to hold on to a substantial atmosphere with its gravitational field but small enough to ensure gravity does not pull a smothering, opaque shroud of gas over the planet.

12. _____ The geo-thermal forces within a planet produce volcanos and earthquakes, which restore the levels of CO_2 in the atmosphere.

13. _____ Our sun is 4.6 billion years old, approximately halfway through its estimated 10-billion-year lifetime.

14. _____ The first exoplanets found in the mid-1990s were all K-dwarfs similar in mass to Earth and orbiting far too close to their stars to harbor any life.

15. _____ The sun's habitable zone is not static but dynamic, so that over time, as it gradually shrinks, Earth will experience radical changes in climate.

16. _____ Earth today could be considered more habitable than it was a few short decades ago.

17. _____ One prominent example is the Great Oxygenation Event of about 2.4 billion years, that resulted in a dramatic rise in sea levels on Earth.

Summary: Write a summary of *Better than Earth* in the space provided.

APPENDIX B

QUESTIONNAIRE I

For each question, please choose the option that is most applicable.

1. What is your gender?

   ☐ Male

   ☐ Female

   ☐ Other

   ☐ Prefer not to answer

2. What is your age?

   ☐ 18–24

   ☐ 25–29

   ☐ 30–34

   ☐ 35–39

   ☐ 40–44

   ☐ 45–49

   ☐ 50–54

   ☐ Greater than 54

   ☐ Prefer not to answer

3. In which year are you currently enrolled (i.e. first, second, third, or fourth)?

   _____

4. In which degree program are you currently enrolled?

   _____

5. On average during a semester, how much time do you spend reading each day?

   A. Academic reading for courses and research:

      ☐ Less than thirty minutes

      ☐ Thirty minutes to one hour

      ☐ One to two hours

      ☐ Two to three hours

      ☐ More than three hours

   B. Personal reading for pleasure and interests:

      ☐ Less than thirty minutes

      ☐ Thirty minutes to one hour

      ☐ One to two hours

      ☐ Two to three hours

      ☐ More than three hours

   C. Other, please describe: _____

      ☐ Less than thirty minutes

      ☐ Thirty minutes to one hour

      ☐ One to two hours

      ☐ Two to three hours

      ☐ More than three hours

6. Of all the reading you do, what percentage is done on paper (books, photocopied or printed articles, etc.) as compared to on screen (e-books, online course materials, websites etc.)?

   A. Academic reading for courses and research:

   - I read about _____% of my total reading on paper.
   - I read about _____% of my total reading on screen.

   B. Personal reading for pleasure and interests:

   - I read about _____% of my total reading on paper.
   - I read about _____% of my total reading on screen.

7. Which, if any, devices do you use for reading? Please check all that apply.

   ☐ Computer (desktop or laptop)

   ☐ Tablet computer

   ☐ Mobile device

   ☐ E-book reader

   ☐ Other, please describe: _____

APPENDIX C

QUESTIONNAIRE II

1. What did you think of the study?

2. How did you read the articles? For example, did you read the articles from beginning to end or did you skim them?

3. What did you think of the highlights? Were some highlights more useful than others?

4. Please share any other comments you have about the study.