# EXPLORING THE RELATIONSHIP BETWEEN DIFFERENT SMARTPHONE READING ANNOTATION STRATEGIES AND STUDENTS' COMPREHENSION

by

Brandon J. Comstock

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Education in Educational Technology

Boise State University

May 2021

BOISE STATE UNIVERSITY GRADUATE COLLEGE

**DEFENSE COMMITTEE AND FINAL READING APPROVALS**

of the dissertation submitted by

Brandon J. Comstock

Dissertation Title: Exploring the Relationship Between Different Smartphone Reading Annotation Strategies and Students' Comprehension

Date of Final Oral Examination: 01 March 2021

The following individuals read and discussed the dissertation submitted by Brandon J Comstock, and they evaluated the student's presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

| | |
|---|---|
| Jui-Long Hung, Ed.D. | Chair, Supervisory Committee |
| Yu-Chang Hsu, Ph.D. | Member, Supervisory Committee |
| Patrick Lowenthal, Ph.D. | Member, Supervisory Committee |

The final reading approval of the dissertation was granted by Jui-Long Hung, Ed.D., Chair of the Supervisory Committee. The dissertation was approved by the Graduate College.

## DEDICATION

To my wonderful children, Coleton, Hayden, Brooklyn, Kenzie, and Madison. Showing you the importance of education was my greatest motivator to complete this dissertation. May you always do things you aren't sure you can do.

ACKNOWLEDGEMENTS

Several people made the completion of this dissertation possible. Thank you to my committee, Dr. Yu-Chang Hsu, Dr. Patrick Lowenthal, and especially Dr. Jui-Long (Andy) Hung, for the endless amounts of feedback and counsel provided throughout this process. Most importantly, thank you to my wife Carleen for supporting me and picking up all of the slack during the many hours I was locked in my office writing.

ABSTRACT

This study explores the relationship between different smartphone reading annotation strategies and students' comprehension. Subjects in the study are 139 teenage students enrolled in a religion class in the Southwestern United States. Each of the participants utilized a digital reading app on their personal smartphone to read an 842-word religious text. Subjects were encouraged to look for, highlight, or tag passages in the text that they felt were important to understanding the meaning of the text. After completing the reading, participants completed a multiple-choice quiz with both factual and inferential questions and wrote a short essay on how they felt the text could be used to resolve an issue in their personal life. The researcher analyzed the data by comparing the frequency of tags and highlights each subject created with their assessment scores. Results showed that higher highlighting frequency was related to higher factual comprehension scores but not higher inferential comprehension scores. In contrast, higher tagging frequency was related to higher inferential comprehension scores but not higher factual comprehension scores. In each case, the higher annotation frequency was only related to higher assessment scores when the subject created an above-average number of tags or highlights. The study suggests that different annotation methods are related to different comprehension outcomes.

*Keywords:* digital reading, smartphones, digital annotation

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER ONE: INTRODUCTION AND OVERVIEW

In 1440 Johannes Guttenberg, a German printer, invented the Gutenberg printing press. The creation of the printing press led to an explosion in printing and the spread of knowledge. In just 60 years, the number of printing presses in Europe grew to over 1,000, and by the year 1600, experts estimate that the number of printed books in Europe increased from just 30,000 to over 200 million (Andrews, 2012, 2015). As the availability of books increased, the ability of everyday people to learn and study from books like the Bible became commonplace; and, in many ways, changed the world.

Since the launch of the first smartphone in 2007, mobile device ownership has skyrocketed. Silver (2019) reports that global ownership of mobile devices is now higher than five billion. While the percentage of subjects who own smartphones varies by country, smartphone ownership in the United States has grown to 81% amongst adults and 95% amongst 18 to 35 year-olds. Given the remarkable growth of smartphone ownership and the unprecedented access to the content they provide, the comparison of reading on smartphones to Guttenberg's printing press, while bold, seems appropriate (Rappleye & Halverson, 2019).

## Digital Reading

On the surface, digital reading is essentially the same thing as reading from print. For example, Nordquist (2017) defined digital reading as "the process of extracting meaning from a text that is in a digital format" (para. 1). When the emphasis of that definition is placed on the phrase "extracting meaning from a text," digital reading is

likely no different than reading in print. If, however, the emphasis is placed on "the process," the unique nature of digital reading begins to emerge.

In chapter two, this study describes the unique aspects of digital reading at length, but first, it is essential to understand that not all digital texts utilize the same format. McFadden (2012) uses the terms "enhanced print" and "native digital" to describe the two major types of digital reading. An enhanced print text is simply a digital copy of a printed text (McFadden, 2012). A native digital text, however, is a text that is designed to maximize the benefits of digitization (McFadden, 2012). These books often contain features like interactive graphs, embedded media, and hyperlinks to additional content. In another study, Rockinson-Szapkiw et al. (2013) differentiate between the two types of digital text with the terms "page fidelity e-textbook" and "reflowable e-textbook." They explain that: "Page fidelity e-textbooks are simply scanned pictures of the print version of the book," while "reflowable e-textbooks use a flexible format system that includes dynamic media and allow the user to modify both the layout and interactive features of the e-textbook to suit the display medium" (p. 260).

## Digital Annotation

While this study occasionally contrasts the different digital reading formats, it focuses primarily on the annotation options of reflowable or digitally native texts in smartphone apps. In this study, “annotation” will refer to "any action that deliberately interacts with a text to enhance the reader's understanding of, recall of, and reaction to the text” (Eastern Washington University Writers’ Center, 2018, para. 1). While the features within reading apps vary, the following are common digital annotation features:

- highlighting

- tagging
- search functions
- hyperlinks
- embedded dictionaries
- marginal notes

Each of these features is designed to enhance the reading experience, not by changing the goal of the reading, but by changing the process for the reader (Hillesund, 2010; Rockinsaw-Szapkiw, Holder, & Dunn, 2011).

For instance, highlighting passages of text has long been the most popular annotation choice amongst print readers (Bold & Wagstaff, 2017; Schugar, Schugar, & Penny, 2011; Qayyum, 2008). However, research has shown that readers primarily highlight texts to flag keywords or themes so that they can quickly return to them later (Qayyum, 2008). While most digital reading apps still allow readers to highlight text, many digital reading apps also offer readers an advanced method for flagging passages via tagging. With tagging, the reader uses a keyword or phrase to tag a passage, which the app uses to create an alphabetized index of topics, allowing the reader to quickly return to all passages on a given topic without having to return to the text to find them (Durham & Raymond, 2016; Lauterman & Ackerman, 2014; Reid, Morrison, & Bol, 2017).

Even when the reader does not create a tag, most digital reading apps include a search function. In this way, by typing in a specific word or phrase, a user can locate each reference containing that specific word or phrase within a text.

Similarly, while printed texts have long contained footnotes or other references to outside material, the inability of local libraries to offer access to every publication can make locating and obtaining a physical copy of a given text challenging and time-consuming (Hillesund, 2010). In contrast, the ability to access content stored digitally by selecting a hyperlink expands readers' accessibility to outside content significantly (Antonenko & Niederhauser, 2010; Azevedo & Cromley, 2004; Spiro, Feltovich, Jacobson, & Coulson, 1991).

Likewise, embedded dictionaries allow readers to define a word at the touch of a finger rather than having to obtain and search through a large printed version of a dictionary to do so. By reducing the difficulty of the process, these features increase the likelihood that readers will actively seek to deepen their understanding and form connections between the text they are reading and their existing knowledge base (Greenlee-Moore & Smith, 1996; Stoop, Kreutzer, & Kircz, 2013).

Even processes like writing notes in the margins of the text, which have long been associated with print annotation, have the potential to be enhanced through digital reading formats. While print marginalia is limited to the extra space on a page, digital marginalia typically allows the reader to embed much lengthier notes in the text. Research indicates that eliminating this restraint leads to readers being as much as three times more likely to create marginal notes (Bold & Wagstaff, 2017; Rockinson-Szapkiw, Courduff, Carter, & Bennett 2013; Schugar et al., 2011).

The features mentioned above demonstrate that the benefits of digital reading are not limited to increased accessibility. The unique annotation features available with

digital reading allow readers to interact with the text in ways that have never before been possible.

## Problems with Digital Reading

Despite the significant potential benefits of digital reading in an educational setting, there are reasons to be concerned. Amongst the greatest concerns for educators that seek to utilize smartphones as reading devices are the following challenges:

- multi-tasking on digital reading devices
- the impact of scrolling through text rather than turning pages
- difficulties that arise from operating systems

Research has shown that people are often distracted by their smartphones. This research becomes even more problematic in educational settings. Several studies demonstrate that smartphone distractions in educational settings lead to lower academic grades and comprehension test scores (Baron, 2017; Froese et al., 2012; Rosen et al., 2013). Noyes and Garland (2005) explained that the problem youth have is that they "still view computers as 'toys' that allow them to play games, email their friends, and search the web, as opposed to being used for serious academic work" (p. 4). On the other hand, adults are often distracted by their phones because they serve as "a constant reminder of things undone" (Hillesund, 2010, p. 9). Regardless of whether someone sees a smartphone as a recreational device or a tool to accomplish work, research shows that the mere presence of a device that can be used for tasks other than reading is enough to impair a subject's ability to focus on academic tasks (Bowman, Levine, Waite, & Gendron, 2010; Przybylski & Weinstein, 2013; Thornton et al., 2014).

Even if multi-tasking distractions were not a concern, studies show that readers have lower text comprehension when reading digitally than they do when reading from print (Kong, Seo, & Zhai, 2018). Several explanations for this phenomenon exist, but a common explanation is that scrolling through text does not support the mind's ability to place a passage within the context of other passages as effectively as a printed page (Jabr, 2013; Liu, 2005; Mangen et al., 2013).

Other research suggests that the decrease in reading comprehension on smartphones may come from features designed to deepen understanding (Hillesund, 2010). For example, research shows that as readers use hyperlinks to shift between texts more rapidly, the process of determining how each text is related requires significantly more effort than focusing on comprehending a single text (Ackerman & Leiser. 2014; Antoneko & Niederhauser, 2010; Shapiro, 1988). Likewise, some scholars have suggested that in contrast to printed texts, which are virtually free from distraction, the design of many reading apps might be a distraction to the reader:

> Most reading software is designed with toolbars, side panels, and icons, and the applications are often placed within the interface of a web browser or an operating system, with their own toolbars and icons. Thus, with numerous eye-catching elements all offering actions to the user, there is considerable potential for fluency disruptions intruding on the text immersion (Hillesund, 2010, p. 7).

Studies using eye-tracking software support this idea (Silk, 2020). In each case, the idea that embedded annotation features may actually hamper comprehension is the underlying conclusion.

## Gaps in the literature

A large portion of research on digital reading suffers because it adopts overly narrow perspectives on digital reading. Most notably, many studies seek to label either print or digital reading as the superior format (Lauterman & Ackerman, 2014; Nichols, 2016; Sidi et al., 2016). The problem with this is different digital reading formats offer significantly different approaches to interacting with the text. For example, several studies on digital reading comprehension have utilized digital reading formats like PDF documents, which, while viewed digitally, do not offer the reader the benefits of advanced digital annotation previously discussed (Ben-Yehuda & Eshet-Alkalai, 2014; Kong et al., 2018; Nichols, 2016). Despite the significant differences, such studies are often used to categorize all digital reading.

A second challenge comes from participants' lack of familiarity with digital reading apps or platforms. When researchers have investigated things like digital reading comprehension while using an application that enables digital annotation, participants often received little to no training on how to utilize it (Ben-Yehudah & Eshet-Alkalai, 2014; Dobler, 2015; Van Horne et al., 2016). Thus, while some research has concluded that digital annotation features are ineffective at enhancing readers' experience and increasing comprehension, other research has shown that digital annotation is effective at these tasks when thorough training is received (Azevedo & Cromley, 2004; Chen & Chen, 2014; Dobler, 2015; Johnson et al., 2010; Lauterman & Ackerman, 2014). These studies support the idea that differences may have more to do with training than format.

Closely linked with proper training is the role of familiarity with print and digital reading formats and platforms. Several studies suggest that simply providing a one-time

training on how to digitally annotate rarely results in readers adopting and benefiting from the various features. Instead, subjects obtain familiarity when researchers provide training and modeling in multiple training sessions and provide opportunities for readers to practice what they have learned by using a specific annotation skill to annotate a text (Lauterman & Ackerman, 2014; Nichols, 2016; Nichols, 2018; Van Horne et al., 2016). When researchers use this approach, results have shown that digital annotation features typically provide a significant benefit to the subjects' reading experience and comprehension (Lauterman & Ackerman, 2014; Nichols, 2016; Nichols, 2018; Van Horne et al., 2016).

## Theoretical Framework

While many of the studies above utilize a perspective in which digital and print reading are essentially the same, Spiro (1988) proposed a theoretical perspective through which readers could implement the unique aspects of digital reading and approach digital reading with a different mindset. Cognitive flexibility theory encourages readers to break free from the traditional linear approach to reading, where the reader moves uninterruptedly from the first page of the text to the last and embraces a non-linear style. With non-linear reading, the focus shifts from trying to understand a text's perspective on a given topic to seeking to deepen understanding of a text by connecting it with other texts and the reader's existing knowledge (Lemke, 1991; Spiro, 2012).

According to Spiro, this process of continually forming connections is the key to readers obtaining knowledge they can flexibly apply to a variety of settings. Annotation features like tagging, search functions, embedded dictionaries, and hyperlinks encourage creating such connections. Thus, while using a digital reading format is not a prerequisite

for effective non-linear reading, the unique features of digital reading make the format particularly well suited to support effective non-linear reading (Spiro, 1988; Spiro, 2012; Spiro et al.,1991).

## Purpose of the Study

While cognitive flexibility theory represents a reasonable lens through which readers can properly utilize the unique annotation features of digital formats, the perspective also represents a significant shift in approach to reading. Given the lack of established best-practices for digital reading on smartphones, and the significant amount of scholarly research suggesting that smartphones in the classroom are detrimental to learning (Baron, 2017; Bowman et al. 2010; Junco & Cotton, 2012; Przybylski & Weinstein, 2013; Thornton et al., 2014), many educators may be hesitant to adopt the technology in their classrooms.

To help address those concerns, the primary research question of this study is, "What is the relationship between different smartphone reading annotation strategies and students' comprehension?" This study aims to answer that question by exploring the following sub-questions:

1. What difference is there between students who create an above-average number of highlights, a below-average number of highlights, and students who do not highlight on comprehension?
2. What difference is there between students who create an above-average number of tags, a below-average number of tags, and students who do not tag on comprehension?

3. Which text annotation feature (highlighting or digital tagging) is more beneficial to readers' comprehension?

In this study, annotation serves as the independent variable and consists of two methods, highlighting and digital tagging. Students' comprehension is the dependent variable and is measured using two different methods. First, the researcher asked participants to write a short essay, which was graded using a rubric based on the question-answer framework (QAR) (Raphael, 1982), a four-point scale designed to measure the depth of a student's understanding. Additionally, readers completed a short quiz with equal numbers of factual and inferential questions.

## Methodology

The sample for the study consisted of 139 fifteen to seventeen-year-olds. The participants were students attending a youth seminary program focusing on reading and understanding ancient religious texts, including the Old Testament, the New Testament, and the Book of Mormon (Seminary and Institute, 2017; Seminaries and Institutes, 2020). In addition to understanding the content within each text through the original author's eyes, students in the program are encouraged to understand how the principles taught apply to their own lives.

The setting was selected for at least two reasons. First, the Old English language and moral nature of the text for the course is complicated and difficult to understand. Spiro explained that such texts are ideal for testing the impact of annotation on subjects' cognitive flexibility (Fass & Schumacher, 1978; Spiro, 2012).

Second, the program provides each student with a free digital version of the text via the Gospel Library app and allows them to use it in class. The Gospel Library app has

more than three million active users and provides content in more than 120 languages (Olsen, 2018). In addition to scriptural text, the app provides users with access to a wide variety of articles from religious leaders and scholars and thousands of audio-visual resources to help deepen readers' understanding. In addition to receiving the app, as part of their class, students enrolled in the program receive training on how to use the app's annotation features. Thus, the study was able to overcome the challenge of participants having a lack of familiarity with the format that plagued other studies.

## Contribution to the Literature

This study contributes to the body of research in at least two significant ways. First, in contrast to studies that introduce participants to a reading platform then immediately test them using the platform without allowing them time to become familiar with its features, this study utilizes participants with substantial experience using the format (Gospel Library app). In so doing, this study provides practitioners with a reasonable idea of what they can expect when a smartphone reading program has matured rather than what they can expect when it is first introduced.

Second, in contrast to studies that assume that digital reading is essentially the same as traditional print reading, this study examines digital reading's unique annotation features. By quantifying each feature's unique relationship with comprehension, this study provides educational practitioners with a beginning point for best-practices that can be applied in classrooms implementing smartphone reading programs.

## Chapter Summary

As with Gutenberg's printing press, the ability to access and read texts through smartphones has the potential to change the world in many ways. In addition to

traditional annotation features like highlighting and writing notes in the margins, digital annotation offers opportunities to interact with texts in new ways like tagging, searching, and linking. In contrast, using smartphones as digital reading devices also has several unique problems, including challenges with cognitive load and distractions associated with smartphones. This study uses Spiro's (2012) cognitive flexibility theory as a lens through which the potential benefits of highlighting and tagging are examined. This study aims to serve as a guide for educators who are considering using smartphones as reading devices in the classroom.

# CHAPTER TWO: LITERATURE REVIEW

While chapter one introduced the literature regarding digital reading and annotation on smartphones, this chapter presents a more thorough literature review. The chapter begins by reviewing the definitions of keywords and phrases and presenting a brief history of how digital reading came about. The majority of the chapter is then devoted to examining the literature on the benefits and drawbacks of digital reading, specifically, digital annotation features. After exploring the literature on digital reading and annotation, this study presents a theoretical perspective through which digital reading's benefits can be magnified. Finally, after discussing the importance of training and familiarity, this chapter closes by discussing conclusions drawn from the research and deficiencies in the current body of literature

## Defining Key Terms

In chapter one, digital reading was defined as "the process of extracting meaning from a text that is in a digital format" (Nordquist, 2017, para. 1). Annotation, which was defined as "any action that deliberately interacts with a text to enhance the reader's understanding of, recall of, and reaction to the text" (Eastern Washington University Writers' Center, 2018, para. 1), is one way that digital reading encourages a different "process."

The format of most printed texts encourage readers to read linearly (Hillesund, 2010; Mangen et al., 2013; Spiro, 2012). The reader begins on the first page of the text then proceeds to the second and third, consuming the text in the order designated by the

author until they finish the final page. In contrast, digital formats make it possible to use hypertext, which Spiro (2012) defined as "computer-based texts that are not read in a linear fashion (even if they can be)" (p. 166). This element of digital reading uses annotation features like hyperlinks, tags, embedded dictionaries, search functions, and others, to shift the responsibility for the order, depth, and detail of information received from the author to the reader.

## Evolution of Reading Strategies

While the use of digital formats has been a relatively recent development in the history of reading, the evolution of the reading process to support a non-linear reading style has been hundreds of years in the making. Early forms of reading and writing utilized formats like cuneiform, papyrus scrolls, and metal plates as the means through which information was communicated (Gnanadesikan, 2011). Not surprisingly, the labor-intensive nature of creating texts through this approach led to the creation of a minimal number of texts. Thus, even if early readers wanted to move rapidly from text-to-text in a non-linear manner, the lack of readily available cross-references made the practice implausible. As such, the early days of reading forged a tradition of reading and re-reading texts in a linear manner (Hillesund, 2010).

While linear reading emerged as the most common approach in early reading, it is not necessarily the only or best approach. Since reading is a relatively new invention in humanity, the human brain does not have a designated system for completing the task. Instead, human brains adapt and utilize systems typically used for other tasks to process, internalize, and comprehend texts (Jabr, 2013; Tanner, 2014). With no designated way for the brain to read, the idea that any particular method is more "natural" than another is

un-based (Jabr, 2013; Tanner, 2014). However, this fact has not stopped printers and publishers from seeking to find the best way to present a text.

The quest for readability has led to a series of adjustments and improvements to printing designed to make non-linear reading possible. Hillesund (2010) explains that innovations in publishing like the turnable page (in contrast to scrolls), the addition of chapters, tables of contents, page numbers, and the addition of verse numbers in Bibles, were all designed to allow readers to easily navigate away from, and back to, specific places in a text. Likewise, standardized spacing and type fonts made it easier for readers to scan through a document looking for keywords or phrases quickly. As these features began to appear more frequently, a non-linear approach to reading became commonplace (Hillesund, 2012; Liu, 2012).

That, however, would change in the eighteenth and nineteenth centuries. Before that time, the majority of texts were non-fiction. However, as printing became less burdensome, the popularity of the novel surged. Given the importance of storyline in fiction books, the typical reader began again to adopt a linear approach to reading, beginning with page one and following the author's thought process until the text concluded (Hillesund, 2010; Liu, 2005, 2012).

Despite the change in style, the necessary elements for successful non-linear reading remained in place and, in certain realms, thrived. For example, the modern academic paper utilizes in-text citations designed to provide the reader with additional information on various claims. Not surprisingly, Hillesund (2010) found that the majority of highly expert academic readers used in-text citations to take a non-linear approach when reading, regardless of whether they read digitally or in print. His findings

demonstrate that, stylistically, the non-linear style encouraged by digital reading is not a new experience as much as the next step in the evolution of reading.

Regardless, many modern readers have struggled to make the leap to a digital reading format. Noyes and Garland (2005) studied students' preferences for digital and printed texts. They found that 52.8% of subjects preferred learning from print texts, while just 4.2% had a preference for learning from computers (the remainder having no preference). More recently, Baron (2017) surveyed 400 subjects between the ages of eighteen to twenty-four and found that 86% preferred to use print for schoolwork.

Despite many readers' preference to read from print texts, digital reading's prevalence continues to grow. Kurata et al. (2017) surveyed 1,755 readers and found that subjects conducted just 30.3% of their total reading in a print format, with 69.7% conducted digitally. Notably, the percentage of digital reading for subjects between the ages of 18-29 in the study was a whopping 80%, suggesting that the percentage of reading that takes place digitally is not likely to decrease in the future. When viewed together, this data creates a fascinating paradox. Despite consistently reporting that they do not prefer to read digitally, readers are conducting an increasingly large portion of their reading in a digital format.

Given the consistent message from readers that they prefer printed formats, it is not surprising that the early days of digital book publishing sought to align the digital reading experience as closely as possible to readers' experience when reading from print (Bold & Wagstaff, 2017). More recently, however, digital reading purveyors have sought to differentiate the format's unique features rather than mimic the print reading experience. The pivot in approach is exemplified by Keim (2014), who proposed that

"maybe it's time to start thinking of paper and screen another way: not as an old technology and its inevitable replacement, but as different and complementary interfaces, each stimulating particular modes of thinking" (p. 2).

## Annotation Features and Digital Reading

In addition to the differences that come from the screen-based nature of reading from a smartphone, the annotation features available in many reading apps also have the potential to alter the process through which readers read significantly. Reid et al. (2017) divided annotation into two main categories, namely, non-generative and generative.

Non-generative annotations emphasize things the author explicitly said. The following are common examples of non-generative annotation:

- highlighting, underlining, or boxing in the text
- using the search function to find similar passages
- adjusting the font and display controls for the text
- utilizing the embedded dictionary to define words

In contrast, generative annotations require readers to add their own unique content to the text. The following are examples of generative annotation:

- tagging
- marginal notes
- creating hyperlinks

This study will discuss each of the digital annotation features listed above within those two categories.

## Non-generative Annotation

### Highlighting

Of the various forms of digital annotation, highlighting is by far the most common. Qayyum (2008) used annotation tracking software to track how readers annotated a series of three texts that they read over the course of three weeks. The study found that 58% of all annotations came in the form of highlighting/underlining text, with an additional 32% coming from similar approaches like circling, creating asterisks, and drawing arrows.

The popularity of highlighting as a digital annotation approach should not come as a surprise since readers are likely more familiar with the concept of highlighting texts from their experiences reading from print. Still, several studies suggest that the highlighting skills readers develop when reading from print are not necessarily transferrable to their digital reading experience. Schugar et al. (2011) sought to understand the annotation habits of university students. Their study provided 30 university students with free access to an e-reader for a semester and then surveyed them on their annotation habits. Results showed that 50% of readers reported highlighting paper texts on a daily basis, while just 14.3% of digital readers reported doing the same. While that study used a relatively small number of subjects, Mizrachi (2015) conducted a similar survey on students' print and digital annotation habits but utilized a much larger subject pool. Her survey of 390 UCLA undergraduates found an even more significant gap, with 80% of students agreeing or strongly agreeing that they highlight print readings and just 33.6% claiming they do the same with digital readings. After reading subjects' open-ended responses about their digital highlighting habits, Mizrachi (2015) concluded

that "more would do so if they knew how, or if the particular format allowed them. Many PDF's, for example, do not enable such engagement" (p. 304). While most smartphone reading apps do allow highlighting, Mizrachi's comment highlights the importance of training readers to utilize digital annotation features.

Van Horne et al. (2016) also highlighted the importance of training but demonstrated that a single training session, on its own, is often insufficient. Their study of 274 university students began by providing each student with a free digital textbook for the course. On the first day of class, a graduate student from the university provided students with training on how to utilize the basic features of the text, including highlighting, creating bookmarks, adding notes, and others. Researchers then analyzed data obtained from the publisher to track how long it took to begin using each feature and how regularly each feature was used. Results showed that students in the study took much longer than expected to create their first annotation, averaging 28 days from their first log-in to make their first highlight. The researchers reported that "only one tool, the highlighting tool, had a median time to first usage because it was the only tool that at least 50% of the subjects had used" (p. 420). Researchers also pointed to one particular class with an unusually fast adoption rate and suggested that the difference was likely a result of specific assignments given by the instructor throughout the semester that, while not requiring students to highlight the text, encouraged them to do so.

With that in mind, the data on digital reading's most commonly adopted and popular annotation feature sets an essential standard for digital annotation in general. Helping readers adopt digital annotation features typically requires more than one-time training and is greatly enhanced by providing readers with specific opportunities to use

each skill. While such training does not guarantee that a reader will utilize annotation features, several studies suggest that when training and practice opportunities are provided, readers are not only more likely to highlight, they report having a better reading experience and record higher levels of comprehension (Chen & Chen, 2014; Dobler, 2015; Lauterman & Ackerman, 2014; Rockinsaw-Szapkiw et al., 2011).

Search Functions

Beyond highlighting, digital reading offers several non-generative annotation features that are often taken for granted or overlooked. For example**,** the searchability of digital texts may not initially appear to be an annotation feature. In reality, however, the ability to search a digital text plays a significant role in how readers interact with, understand, recall, and react to a text.

The use of the search feature begins when a reader seeks to find a text. While the size, weight, appearance, and even smell of printed books are often a factor in readers' selection of a text, digital readers use a different approach. Digital readers use the search feature to significantly narrow the amount of content that they will further examine. Even with the list of potential texts reduced, the abundance of materials available through a digital search typically results in readers quickly exploring small portions of several texts rather than investing significant time into a single read. Nicholas et al. (2008) conducted a study of college students' reading habits and found that two-thirds of their digital article views lasted less than three minutes, and 40% lasted less than one minute. Likewise, Liu (2005) surveyed 113 people on how their current reading habits aligned with their reading habits ten years ago. In the study, 80% of subjects reported spending a higher percentage of time scanning and browsing while reading than they did ten years earlier.

While that process often begins when a reader is selecting a text, it seems to carry over into the actual study of a text as well. Baron (2017) suggests that the emergence of a search function is the root cause that readers often refer to "reading" print textbooks and "using" digital textbooks. She explains:

> We need to ask ourselves how the digital mindset is reshaping student's understanding of what it means to read. Since online technology is tailor-made for searching for information rather than analyzing complex ideas, will the meaning of “reading” become “finding information” rather than “contemplating and understanding” (p.19)?

Not surprisingly, this change in reading strategy includes both benefits and pitfalls. While the search function allows users to sort through an enormous amount of content quickly, it also removes a significant amount of the effort that was previously required to make sense of a text. Sanders (2017) likens the change to a traveler using either a printed map or a GPS. While the map requires significantly more effort to use, the constant requirement to determine how the map relates to their specific task leads to a deeper understanding of how the users’ surroundings are related. In contrast, the GPS gives access to far more content but encourages what Sanders (2017) refers to as a "response strategy." In this approach, the traveler no longer assumes the responsibility to position themselves within the information and simply responds to instructions to "turn-right" or "continue-straight." Similarly, utilizing a search function is tremendously convenient but has the potential to lead to the acquisition of knowledge that is decontextualized and thus less applicable to the reader.

Given the tradeoffs associated with using digital searches, educators should make great efforts to ensure that the feature's convenience does not lead to lower levels of comprehension. Fortunately, while readers use the search function to reduce the burden of finding a particular passage, many of the annotation features that are unique to digital reading are designed to help enhance a readers' ability to analyze and comprehend the meaning of a text (Reid et al., 2017; Sidi et al., 2016). Thus, when used together, the reader can enjoy the benefits of digital search functions while lessening or altogether eliminating the drawbacks.

### Adjustable Fonts and Displays

In addition to search functions, another unique aspect of digital reading is the reader's ability to customize how the content is displayed and presented. Customization options for the presentation of the text in digital formats vary but often include brightness control, adjustable font style and size, and options for audio narration.

While most scholars consider backlit screens to be a negative aspect of digital reading, research shows that increased screen brightness has both positive and negative impacts on digital reading. Benedetto et al. (2014) used a video-based infrared eye tracker to measure the differences in 50 subjects' eye movements as they read a text. The researchers found that when readers read with high screen brightness, they had increased eye-fatigue but had fewer struggles staying focused and read faster than readers in low brightness settings. In contrast, readers who used low screen brightness blinked more frequently, decreasing tear evaporation and resulting in lower levels of fatigue.

Likewise, research has shown the ability to adjust the font-size to be particularly helpful for readers with poor vision and learning disabilities like dyslexia (Chung, 2004;

DeLamater, 2010; Levi, 2008; O'Brien, Mansfield, & Legge, 2005). Schneps et al. (2013) explained that one issue faced by dyslexic readers is a limited visual attention (VA) span. Since readers with limited VA spans struggle when pages are crowded with too many words, researchers proposed that adjusting the font size to limit the number of words per line may be beneficial to dyslexic readers. The study utilized 103 subjects attending a high school for dyslexic students. Subjects in the experimental group read a text from an iPad with the font adjusted to size 42 so that the display showed just 3.5 words per line. Subjects in the control group read the same text from a printed page that used a smaller, more traditional size fourteen font. Results showed that those with limited VA spans had significant benefits in both reading speed and comprehension when the font size was adjusted to display fewer words per line.

Similarly, many digital texts now offer an audio narration feature, allowing readers to listen to the text rather than merely seeing it. Research has shown that this feature is advantageous to young readers who are in the process of learning to read (Grimshaw et al., 2007).

Embedded Dictionaries

Finally, many digital readers offer a built-in dictionary function that allows users to select any word and immediately view its pronunciation and definition. This feature is particularly useful in an academic setting, where texts can often be technical and contain unique words. Stoop et al. (2013) conducted a study of 173 students to determine the effectiveness of the built-in dictionary. Researchers utilized a paper group and a digital group. Researchers provided students who utilized the paper approach with a dictionary and practice questions at the back of the text.

In contrast, the digital text utilized a dictionary with a mouse-over function, and researchers placed the practice questions next to the location of the answer in the text. After completing the reading, students in both groups received a separate 24 question knowledge test. Researchers found that despite being provided the same materials, subjects in the digital reading group scored better or significantly better than the print group on eighteen of the questions. Researchers explained that the difference in scores was likely due to the embedded nature of the dictionary and study questions in the digital format, adding that digital readers were significantly more likely to use both features. In other words, the ability to study the text and supplementary material in a non-linear fashion was better supported by the digital format.

In a similar study, Greenlee-Moore and Smith (1996) examined the impact of embedded dictionaries on 31 nine and ten-year-old students. Results showed that subjects with an embedded dictionary performed particularly well when the text was long and difficult. The author proposed that while students in the print condition were permitted to raise their hand and ask for help with challenging words or phrases, they were often hesitant to do so. In contrast, using the embedded dictionary allowed them to enjoy "privacy of failure" when they did not understand the meaning of a word and increased the likelihood that they would take time to define challenging words.

It is important to note that research on embedded digital dictionaries is limited. However, the results from available studies are favorable, and there appears to be no drawbacks to their use.

### Summary of Non-generative Annotation Features

The non-generative annotation features discussed above demonstrate several important elements of digital reading. For example, in many cases, digital annotation skills appear to be essentially identical to print reading, but it should not be assumed that readers will naturally transfer annotation skills from print to digital reading (Mizrachi, 2015; Schugar et al., 2011; Van Horne et al., 2016). In contrast to highlighting, features like search functions are in many ways fundamentally transforming the way we read by encouraging readers to take a non-linear approach and focus on multiple short passages rather than a single lengthier passage (Baron, 2017; Liu, 2005; Nicholas, 2008). Finally, features like adjustable font sizes and brightness (Benedetto et al., 2014; Schneps et al., 2013), and embedded dictionaries (Greenlee-Moore & Smith, 1996; Stoops et al., 2013) offer users the opportunity to customize their reading experience to match their specific needs.

## Generative Annotation

### Tagging

While digital reading offers several intriguing non-generative annotation options, it is digital reading's generative annotation features that offer the most potential to enhance digital reading as a tool for learning. For example, in Qayyum's (2008) study on how students mark texts, students reported that they used highlighting and underlining for three main reasons: (1) to identify keywords, (2) to place flags in the text so that they could quickly return, and (3) to help spot themes. While readers can accomplish each of those things through highlighting, digital tagging offers a more practical approach.

Readers who utilize digital tagging look for themes while reading then create a tag by identifying those themes with keywords or phrases. While readers typically create tags while reading linearly, digital reading devices use each tag to create a digital filing system so that readers can later retrieve information topically, much like a customized index.

Durham and Raymond (2016) introduced a tagging system to a group of fourth-grade students. The study showed that students' attitudes about recreational reading moved from the 54th percentile to the 78th percentile, and academic reading moved from the 72nd percentile to the 87th percentile. A follow-up survey found that 95% of students felt that the process was helpful, with 75% feeling that it helped them remember more about the story, 65% claiming it helped them enjoy the book more, and 50% saying it helped motivate them to read and understand more. While Durham and Raymond (2016) simply asked students to create tags by writing in the margins of printed texts, digital results have shown similar promise.

Reid et al. (2017) hypothesized that studies suggesting different comprehension levels between print and digital reading were the result of subjects being less likely to take time to summarize what they read when reading digitally. They tested the theory on 80 college undergraduates. While comprehension did not improve for the summarizing groups, meta-comprehension (the subjects' ability to judge how well they would perform on the corresponding test) did significantly improve. In essence, students were more aware of what they knew or did not know when researchers asked them to create the summaries.

Lauterman and Ackerman (2014) were able to take their findings one step further than meta-comprehension. In their study of 76 undergraduates, the researchers proposed that readers could close the comprehension gap between digital and print reading if readers in both groups were using in-depth reading strategies. To test the theory, researchers asked both groups to read a text, then identify and write down four keywords after completing the reading. Results showed that when the method was applied, there was no significant difference in comprehension between the two groups.

The difference in comprehension when students annotate was explained by Sidi et al. (2016), who proposed that "while on paper in-depth text processing is the default, on-screen an external trigger is needed" (p. 6). Thus, while some elements of digital reading may encourage reading styles that lead to lower levels of focus and comprehension, when annotation features like tagging are introduced, the negative impacts are negated.

Importantly, each of the studies cited above focuses on the short-term comprehension of readers. A review of the research found no studies on the long-term comprehension benefits of tagging. However, given the significant impact tagging has on subjects' ability to organize, recall, and review content, it seems likely that the long-term ramifications on reader comprehension are likewise positive.

Marginal Notes

While digital tagging typically involves the reader summarizing a passage in a few words, when a longer summary is warranted, inserting notes into the margins provides digital readers with a second generative annotation option. Creating notes in the margins of a text is a digital annotation feature rooted in print reading. However, despite

its origin, several studies show that readers do not view writing notes in the margins of digital texts in the same way they do writing in a printed text.

Schugar et al. (2011) found that 28.6% of subjects in their study reported writing notes in printed books on a daily basis, while just 15.4% reported doing the same with digital books. Likewise, 64.3% of readers reported taking notes on a separate sheet of paper while reading from print, while just 21.4% reported doing the same while reading digitally, even when notepaper was provided for them.

Bold and Wagstaff's (2017) study of 510 active readers (84% of whom had read both digital and print books) also asked readers about their history taking notes in print and digital books. The researchers found that 74% of subjects had written a note in a print book, but just 35% of subjects had written a note in a digital book. While that finding seems ominous, researchers also reported that 53% of subjects reported a desire to create notes in digital texts, suggesting that confounding variables like training and familiarity may be to blame for the gap.

Rockinson-Szapkiw et al. (2013), however, found different results. When they allowed their 538 subjects to use their preferred format (digital or print) rather than being randomly assigned to a format, they found that digital readers were "nearly three times more likely to make notations directly into the text when compared to print text users" (p. 264). The contrast between the studies suggests that familiarity and user-preferences may play a significant role in the likelihood of a user creating notes while reading digitally.

While preferences and familiarity likely play a significant factor, it is worth noting that Mueller and Oppenheimer (2014) found that subjects actually approached digital notetaking differently than they did longhand notetaking. In their study,

researchers asked subjects to take notes on a series of TED Talks using either the traditional longhand approach or a laptop computer. Subsequent comprehension tests showed that subjects in the laptop condition scored similarly to the longhand group on factual questions but significantly lower on inferential questions. Researchers found that laptop users were far more likely to use lengthy verbatim quotes, while longhand note-takers were required to synthesize the information and put it in their own words. In other words, when taking notes digitally, subjects were less likely to take the time to process and internalize the content. While researchers have not yet examined the ramifications for notes taken within a text or on a smartphone, it may be the case that readers who wish to maximize digital marginalia may need to do so in tandem with an annotation method like tagging that requires the reader to synthesize the information concisely.

Hyperlinks

While each of the annotation features mentioned thus far contributes to the unique non-linear style of digital reading, none of them have as significant an impact as the use of hyperlinks. Because of their ability to quickly take a reader to a different portion of the text or a different text altogether, hyperlinks epitomize digital reading's non-linear nature.

Not surprisingly then, the use of hyperlinks is controversial. Antonenko and Niederhauser (2010) acknowledged the benefits of hyperlinks by explaining that:

> The unique characteristics of hypertexts allow hypertext authors to create connections to other related topics that are not easily accomplished in traditional print text presentations. Hyperlinks form a more intricate web of connected

information nodes than is permitted by the straightforward flow of a print text (p. 140).

Despite the benefits, however, the use of hyperlinks leads to a significantly higher burden on users' cognitive processing. Antonenko and Niederhauser (2010) demonstrated the increase by asking subjects to read a short hypertext while attached to an EEG brain monitor. Results showed that subjects had significantly higher brain activity levels in the 20 second period following the moment they clicked on a link.

Observers can view that increase in at least two ways. The first is that the increase in cognitive load will negatively impact readers' reading comprehension. The basic concept of cognitive load is that the mind has a fixed amount of working memory (Miller, 1956). When focused on a single simple task, the minds' cognitive load is relatively low. In contrast, when the mind is managing several pieces of complex information, cognitive load is dramatically increased. When the load is significant, fatigue quickly becomes a factor (Sweller, 1988). Thus, when a reader moves from processing a single text to processing the connections between multiple texts, they experience an increase in cognitive load and, in turn, fatigue.

In contrast, while many view this increase in cognitive load as a negative aspect of digital reading, others view it as a benefit. Shapiro (1988) proposed that the increased mental effort involved in making connections between various texts was a key advantage to helping digital readers obtain a deeper understanding of texts. To test the theory, Shapiro (1988) divided 48 Brown University students into two groups. The first group received a highly structured hypertext, where researchers provided the reader with the connection between each phrase and the text linked to it. The second group received the

same text but without an explanation of how each link was connected. After subjects completed their reading, researchers asked subjects to produce an essay that was graded by an expert in the field on three criteria: (1) depth of understanding, (2) clarity, and (3) overall quality. The researchers found that "on every measure of essay quality related to depth of content, the highly structured participants performed more poorly than their counterparts in the unstructured group" (p. 19).

In discussing their findings, the authors concluded that "structure mitigated the necessity of deeply processing information embedded in the links. Participants in the highly structured condition were able to move through the information less thoughtfully than those in the unstructured group" (p. 25). The authors' conclusion aligns nicely with the findings of Mannes and Kintsch (1987), who concluded that "refraining from providing readers with a suitable schema and thereby forcing them to create their own. . . might make learning from texts more efficient" (p. 93). In essence, while hypertext is requiring readers to work harder, it is also focusing their efforts on making the kind of connections that are necessary for meaningful learning.

While using hyperlinks can help readers develop connections between different concepts, it does come with cautions. Azevedo and Cromley (2004) found that training on how to utilize hypertext was critical to subjects' success. Likewise, Ackerman and Leiser (2014) found that when textbooks included hypertexts without apparent connections to the root-text, subjects had lower overall comprehension. Supporting that finding, Shapiro (1988) summarized the need for care when embedding hyperlinks by stating that "links must be used as more than vehicles for navigation or mere pointers to conceptual ties. A degree of thoughtfulness must be given to these relations" (p. 31). Still,

when properly used, it appears that hyperlinks have tremendous potential to deepen comprehension and understanding of texts.

It is worth noting that a significant shortcoming of the traditional hypertext model is that the author of the text retains complete control over which texts are linked to the original document. This unfortunate element of hyperlinks has the potential to limit the customization of schemas created by the user (Spiro et al., 1991). Despite significant effort, this researcher found no studies on the potential of allowing users to create custom hyperlinks between texts they feel are connected.

Summary of Generative Annotation Features

As with non-generative annotation, some forms of generative annotation are similar to methods used in print reading. However, research confirms that even with annotation methods like creating marginal notes, readers view the process differently when done digitally than when done in print (Bold & Wagstaff, 2017; Mueller & Openheimer, 2014; Schugar et al. 2011). As a result, training on such approaches is necessary. In contrast, annotation methods like digital tagging and hyperlinks offer a very different reading experience (Hillesund, 2010). While some scholars criticize such methods as being too cognitively demanding, others insist that the increased ability to organize material and connect it with other content enhances readers' comprehension rather than restricting it (Durham & Raymond, 2016; Lauterman & Ackerman, 2014; Reid et al., 2017; Sidi et al., 2016). The lack of scholarly consensus on the impact of such annotation methods is likely influenced by readers' format preference and demonstrates the need for a careful approach to annotation in educational settings rather than a casual one (Rockinson-Szapkiw et al., 2013; Sidi et al., 2016).

## Non-annotative Interactions in Digital Reading

This study defines annotation as: "any action that deliberately interacts with a text to enhance the reader's understanding of, recall of, and reaction to the text" (Eastern Washington University Writers' Center, 2018, para. 1). While each element of digital reading discussed so far meets that criteria, there are several aspects of digital reading that do not meet the standard of a deliberate interaction but nonetheless can have a significant impact on the readers' overall experience. Specifically, the following elements of digital reading present concerns:

- increased eye and body strain
- decreased emotional attachment with texts
- a lack of fixed navigational markers
- distractions caused by the multitasking capabilities of digital reading devices like smartphones

The remainder of this section will address those issues, examining the unique physical, emotional, and cognitive experience that a reader has when engaging with a digital text.

### Increased Eye and Body Strain

First, we will explore the physical impact of digital reading. Despite the significant amount of time that the average person spends in front of screens, a major concern for potential digital readers is the impact of backlit screens. One study found that as much as 50% of readers have concerns about the use of backlit screens impacting their health (eye strain, headaches, insomnia, etc.) (Two Sides, 2015). Tanner (2014) explains that the primary reason backlit screens cause eye-strain is that screen readers experience a

reduction in blink frequency. Since each blink lubricates the eye with a mixture of oils and mucous, a decrease in blink frequency leads to dry eyes and increased visual fatigue (Mauk, 2012).

Beyond eye-strain, however, traditional screen-based reading also results in an increased strain on the entire body. Taipale (2015) used an open-ended question to elicit insight from 30 subjects on "the gestures and postures [they] assume in reading and writing using paper and on a screen" (p. 770). Results showed that readers often reported utilizing a hunched forward posture when reading from screens. In contrast, readers typically reported using a leaned back position when reading from print. The study aligns with the findings of MacWilliam (2013), who also reported that digital readers experienced significantly more bodily strain than print readers.

While these concerns are both legitimate and significant, recent technological developments offer significant help. Emayr, Köpper, and Buchner (2017) studied the impact of pixel density on subjects reading comprehension, speed, and proofreading abilities. Their study utilized a screen with 132 pixels per inch (PPI), the density of Apple's first iPad, and a screen with 264 PPI, the density of Apple's 2012 iPad. While no significant differences were found in subjects' comprehension, reading speed, or proofreading abilities, researchers reported that "subjective ratings of physical discomfort revealed significantly more complaints about headache and musculoskeletal strain in the 132 PPI condition than in the 264 PPI condition" (p.41). Their findings are encouraging in a time when screen quality is rapidly advancing. Likewise, innovations like e-ink, which eliminate the use of backlit screens by using electromagnetic charges to push microscopic pixels against a screen to form text, are now making it possible to enjoy the

benefits of digital devices without the drawbacks of backlit screens (Hidalgo, 2019; Siegenthaler et al., 2011).

Regarding the increased musculoskeletal strain caused by fixed computer monitors, the rise of mobile devices like smartphones may be the solution. Hillesund (2010) explained that:

> Handheld devices, especially dedicated e-readers, seem to be capable of giving a fairly good approximation of the reading experience provided by printed books . . . Devices of this kind fit snugly into the hand and let users position the body for reading (p. 10).

Admittedly, despite significant innovation in the field, screen-readings negative impacts have yet to be eradicated. Still, developers have made significant strides towards that goal, and options are available to reduce the impact of screen-reading for genuinely concerned patrons.

Decreased Emotional Attachment

While measuring physical factors is easier than measuring emotional factors, the existing research does suggest that digital reading offers a different emotional experience. MacWilliam (2013) found that 73% of subjects either agreed or strongly agreed that digital devices do not offer emotional attachment in the same way as a paper book. While emotional attachment may seem to be of minimal significance, several studies suggest that it may significantly impact users' preferences for digital technology (Anton, Camarero, & Rodriguez, 2017; Jabr, 2013; MacWiliam, 2013; Read et al., 2011). Since research has shown that format preference has a significant impact on the effectiveness of

digital annotation features (Lauterman & Ackerman, 2014; Rockinson-Szapkiw et al., 2013), the emotional aspects of digital reading are worth consideration.

In many ways, readers' struggle to emotionally connect with a digital text is likely rooted in the framing of discussions about format being binary in nature. For example, Antón et al. (2017) conducted a study on the pleasure received during reading. Results showed a negative correlation between those who had a positive, pleasurable experience with traditional books and acceptance of e-book technology. Positive experiences with print, it appears, created a loyalty to the format that made readers hesitant to try a different format.

Read et al. (2011) found similar results and determined that many readers' passion for reading books extended beyond the text's content to include the physical book itself. The gap between the formats may result from digital books failing to produce a unique experience with each text. Jabr (2013) explains that the seemingly insignificant elements of printed books like weight, cover art, paper quality, and even smell combine to create a unique experience for the reader. In contrast, digital reading devices provide the reader with a sensory experience that, by design, makes one text nearly indistinguishable from another. While this approach has clear advantages, it limits the reader's ability to form a unique bond with each text.

While the ability to use digital annotation features to customize a text may have some impact on this phenomenon, studies suggest that the best way to overcome the challenge is to introduce a human element. Waheed et al. (2015) survey's 366 subjects to explore how the diffusion of innovation theory and readers reported self-efficacy with digital reading combined to shape their decision to accept or reject digital reading.

Results showed that cost and eye-strain were the most common concerns for potential users. In contrast, increased exposure to digital reading devices and receiving assistance from someone when learning to use the digital reading devices positively impacted subjects' adoption rate. Likewise, Stone and Baker-Eveleth (2013) found that from peers, family, and instructors played a significant role in students' willingness to utilize an e-textbook. Their findings support the myriad of studies suggesting that personalized training is critical to overcoming the drawbacks of digital reading (Bennett et al., 2008; Chen & Chen, 2014; Lauterman & Ackerman, 2014; Nichols, 2018).

Lack of Fixed Navigational Markers

While digital reading's physical and emotional elements raise certain concerns, the unique format of digital reading, which typically includes scrolling rather than page-turning, is also problematic. Liu (2005) explained that:

> Flipping and scanning (a reading pattern associated with printed documents) is not only a means for locating information in a document but also a means to get a sense of the whole text. Scrolling on a computer screen does not support this mode of reading and information processing. Readers tend to establish a visual memory for the location of items on a page and within a document. Scrolling weakens this relationship (p. 703).

Jabr (2013) explains that the same portions of the brain are used for reading that are used when navigating a landscape. Thus, just as the driver of a car might navigate when to turn based on a fixed location like a supermarket or large tree, the brain navigates texts based on fixed navigation markers as well. While printed texts offer an abundance of concrete navigational cues (two pages, four corners, multiple columns), the

adjustable nature of text-size in digital formats typically eliminates those markers. With fewer navigational markers, many readers find it more challenging to establish the connections between each new point.

Mangen et al. (2013) tested the idea by asking 72 fifteen-to-sixteen-year-old subjects to read a 1,400-1,600-word text. Researchers divided the subjects into two groups, with one group reading the text from a printed page and the other reading the text digitally via a PDF document. Post-reading assessments showed that subjects in the printed version had significantly higher comprehension than those in the PDF version. In explaining the discrepancy, the authors proposed that:

> Scrolling is known to hamper the process of reading by imposing a spatial instability which may negatively affect the reader's mental representation of the text and, by implication, comprehension. . . We know from empirical and theoretical research that having a good spatial mental representation of the physical layout of the text supports reading comprehension. . . The fixity of text printed on paper supports readers' construction of the spatial representation of the text by providing unequivocal and fixed spatial cues for text memory and recall (pp. 65-66).

As previously discussed, readers can receive significant benefits from adjustable font sizes, and annotation features like tagging and user-generated hyperlinks have the potential to help users develop meaningful connections and deepen understanding. Still, digital readers should be aware that the malleability of digital texts does come at a cost.

Distractions Caused by Multi-tasking on Smartphones

Finally, while each of the issues discussed thus far offers unique challenges, from an educational standpoint, the impact of distraction might be most significant. Baron (2017) found that 85% of digital readers multitask when reading digitally compared to just 26% who reported doing the same when reading from print. Given the impact of multitasking on learning outcomes, that finding is concerning.

One group of researchers set out to determine the impact of a student receiving a text message during a lecture. Researchers asked subjects to take notes on a ten-minute power-point/audio presentation. Half of the students then received text messages from the researchers during the presentation, spending an average of 2.69 minutes reading and responding to the messages. Results of a post-lecture assessment showed that subjects in the testing group scored 27% lower than those in the control group (Froese et al., 2012). The considerable comprehension gap for students distracted by electronics is particularly alarming given a study by Rosen et al. (2013), which embedded researchers in a college class and found that 58% of students were using their devices for non-classroom-related tasks more than half the time.

However, while the impact of actual digital distractions on academic performance is alarming, the more concerning findings may have more to do with the mere presence of digital devices. Bowman et al. (2010) found that the very thought of being distracted was enough to impact performance during digital reading. In the study, researchers asked 89 subjects to read a 3,800-word text and complete a 25-question quiz. While researchers did not inform subjects which group they would be a part of, they did inform them that the experiment contained three groups. One group would receive messages before they

began reading. A second group would receive messages during their reading. The third would serve as a control group, receiving no messages at all.

While scores on the associated quiz yielded no significant differences, the time spent completing the assignment, excluding the time spent messaging, varied significantly. Not surprisingly, subjects who received messages throughout the process were the slowest, completing the assignment in an average of 45.57 minutes. Interestingly, subjects in the control group (37.44 minutes) took significantly longer to complete the assignment than subjects who were messaged before beginning the assignment (28.63 minutes). The authors explained this finding as follows:

> The instructions given to the three groups may have had an impact. Researchers informed each participant that they would receive IM's before reading, during reading, or not at all. Those who received the IMs at the beginning may have realized that they would not receive any more IMs and could attend to the passage they were reading without thinking they might be interrupted (p. 4).

In essence, even when distractions were not present, the simple thought that a distraction might occur was enough to impact digital reading performance.

While the findings in the Bowman et al. (2010) study were a surprise, Thornton et al. (2014) theorized from the beginning that the mere presence of a smartphone would be enough to impact academic performance. Their study included two tests. In the simple version, subjects were provided with a sheet filled with single-spaced one-digit numbers and asked to circle each occurrence of a given number. In the more complex task, subjects received a sheet with single-spaced one-digit numbers, but researchers asked

subjects to circle any combination of two digits that added up to a certain number (ex. 4 + 7 = 11). To test the impact of smartphone presence, researchers innocuously placed a stopwatch and a smartphone on the table of the experimental group and placed a stopwatch and smartphone-sized spiral notebook on the table of the control group. While the smartphone's presence had no impact on the simple task, subjects in the experimental groups scored lower on the task that required more complex thinking than subjects in the control group. Researchers concluded that the mere presence of a smartphone had impaired the participants' ability to focus entirely on the task. Surprisingly, the study is not the first to use the method to show that the mere presence of a smartphone is sufficient to impede subjects' ability to focus on meaningful tasks (Przybylski & Weinstein, 2013).

Results from these studies may, understandably, cause some educators to be hesitant about encouraging students to use smartphones as reading devices in the classroom. However, since the perception of smartphones being a distraction is a critical element in the discussion, teaching students to view their smartphones as more than just entertainment devices may be the key to overcoming the problem. One study found that a common theme among teenagers was that authority figures only taught them what not to do with their smartphones and never taught them what they should do with them (Blackwell, Gardiner, & Shoenebeck, 2016). Other studies have shown authoritative leadership, meaningful relationships, and increased communication about smartphones are all linked with lower rates of addictive smartphone use (Bae, 2015; Lee & Chae, 2012; Wisniewski, Xu, Rosson, & Carroll 2017). Thus, ironically, the solution to cutting

down on smartphone distractions in the classroom may be to properly use them more regularly.

In addition to the bigger-picture task of changing students' perceptions about smartphones, several practical suggestions can help to reduce the impact of distraction in the classroom. Tindell and Bohlander (2012) asked students what instructor characteristics made it easier to be off-task in class. The top result was the lack of a clear policy (32.9%), followed by the instructor always standing in the front of the room instead of circulating (16.8%), regularly turning their back to the class to write on board (19.9%), and not requiring student participation (10.6%). By utilizing this feedback, instructors can make several small adjustments that will allow them to maximize the benefits of smartphones as digital readers while minimizing distractions.

### Summary of Non-annotation Feature and Digital Reading

While digital reading offers a unique set of challenges, technological advances and scholarly research are continually offering new solutions. Musculoskeletal and eye-strain concerns are being addressed through the development of higher resolution screens, e-ink, and the prevalence of handheld reading devices (Emayr, 2017; Hidalgo, 2019; Hillesund, 2010; Siegenthaler et al., 2011). Likewise, research on decreased emotional attachment (Stone et al., 2013; Waheed et al., 2015) and distraction caused by electronic devices (Blackwell et al., 2016; Tindell & Bohlander, 2012) shows that actions as simple as increasing personal interaction with the reader can have a significant positive impact. While technological advancements and scholarly research have not eliminated such concerns, the progress being made is encouraging for the future of digital reading.

## A Strategy for Digital Reading

Theoretical Perspective

The literature above creates the opportunity for advocates of both digital and print reading to make persuasive arguments that one format is superior to the other. However, the literature is ultimately mixed. The literature as a whole suggests that digital reading is not inherently better or worse than any other format. Thus, the question becomes, "What can be done to help readers maximize their digital reading experience?"

In contrast to approaches that suggest that the growing generation inherently understands how to utilize technology (Prensky, 2001), several studies suggest that significant training is needed to help readers understand how to read digitally (Bennett, Maton, & Kervin, 2008; Chen & Chen, 2014; Lauterman & Ackerman, 2014; Nichols, 2018). Spiro's (1988) cognitive flexibility theory (CFT) provides a framework through which readers can successfully receive the benefits of digital annotation features. CFT suggests that while linear reading is sufficient for understanding simple concepts, it suffers from several shortcomings including an often rigid compartmentalization of knowledge components and an overreliance on a single basis for mental representation. Spiro suggests that this approach often leads to knowledge that readers cannot flexibly apply to different situations. In contrast, the non-linear approach taken in digital reading allows the reader to utilize several different perspectives to form their own custom schema of information, making it far easier to apply the information to a variety of situations flexibly.

Spiro offers two analogies to help understand the theory. First, Spiro (2012) likens the process of linear reading to a hiker following a trail and non-linear reading to an

explorer freely traversing a landscape. Since non-linear reading shifts the responsibility of constructing a schema from the author to the reader, CFT utilizes a constructivist world-view, but with a small twist. Spiro et al. (1991) explained that:

> It is clear that there are many variations on what is meant by "constructivist". . . Our constructivist position, as it applies to complex and ill-structured domains, rejects any view that says either that there is no objective reality, or that there is an objective reality that can be captured in any single absolute way. Rather, one of our principal tenets is that the phenomena of ill-structured domains are best thought of as evincing multiple truths: single perspectives are not false, they are inadequate (p. 15).

Thus, CFT encourages users to discover not only the knowledge (or truth) itself but the various ways that the knowledge can be rearranged and adapted to any circumstance.

Spiro's (1991) second analogy compares digital reading and annotation to the process of assembling an erector set. The author provides the reader with all of the right pieces, and the reader uses annotation as a method for determining the purpose of each piece and figuring out how it fits together with each of the other pieces. Spiro (2012) later explained that, like an erector set, CFT encourages readers to assemble and apply the information from a text in whatever way they see fit.

In short, CFT frames learning more in terms of the ability to apply information than the ability to recall and recite information (Spiro, 1988). For example, math teachers often ask students to memorize multiplication tables. Unfortunately, when they are later asked to apply their skill to a word problem, many of the students struggle to do so (Spiro

et al., 1991). While the initial memorization is not necessarily a bad thing, Spiro (1988) explains that, by itself, it is insufficient:

> In introductory learning, the goal is often mere exposure to content and the establishment of a general orientation to a field. . . [but] at some point in learning about a knowledge domain, the goal must change; at some point, students must "get it right". . . The learner must attain a deeper understanding of content material, reason with it, and apply it flexibly in diverse contexts" (p. 2).

Whether it be identifying what is most important and highlighting it, labeling concepts or ideas with a tag, or connecting new ideas to previous knowledge through a link, the process of annotating a text is designed to help readers encode texts in a way that facilitates this type of learning.

Closely related to that idea is the theory of situated cognition. In relation to reading, situated cognition proposes that, to be effective, readers must situate the information they receive into the scenario they plan to apply it. Lemke (1991) explained one way in which situated cognition occurs when he taught that:

> We interpret a text or a situation in part by connecting it to other texts and situations that our community or our individual history has made us see as relevant to the meaning of the present one. Our community and each of us creates networks of connections (and disconnections) among texts, situations, and activities. These networks of connections that we make, and that are made in the self-organizing activity of the larger systems to which we belong, extend backward in time as well [as] outwards into the social-material world (p. 50).

When Lemke (1991) made that comment, he was not referring specifically to digital reading. Regardless, digital annotation features like hyperlinks, embedded dictionaries, search functions, and tagging have been specifically designed to enable the kind of non-linear networking that was discussed by Lemke and Spiro.

Rethinking Comprehension

While both CFT and situated cognition represent a logical approach to reading, it is worth noting that neither theory is ideal in every setting. Spiro (2012) explains that the application of CFT is best suited for ill-structured domains, which he defines as an "area of advanced knowledge in which prepackaged facts are insufficient and deeper learning is required" (p. 165). One author explained that, in general, the non-linear nature of digital reading was better suited for what he called "reflective reading," where the reader engages in the arguments of the text and seeks to interpret its meaning and form a connection. In contrast, the use of digital annotation features is less useful in imaginary reading, where a reader is primarily concerned with storyline and characters (Hillesund, 2010).

Not surprisingly, research has shown that user preferences for digital reading follow this trend. For example, Nichols (2018) found that doctoral students, who engage in a significant amount of reflective reading, were significantly more likely to prefer digital reading (47.7%) than undergraduate students (18%). In contrast, MacWilliam (2013) found that just eleven-percent of those surveyed felt that reading fiction on an e-book was a better experience than reading it in print.

Since digital reading's annotation features are designed to deepen understanding of complex ideas, measurements of the effectiveness of the format should, likewise,

measure meaningful comprehension. Chen et al. (2014) divided comprehension into two main categories: literal comprehension and inferential comprehension. Literal comprehension is essentially the recall of content explicitly stated in the text. In contrast, inferential comprehension, which the authors refer to as deep comprehension, represents a highly coherent, richly integrated understanding of the text. This form of comprehension goes beyond what the author explicitly stated and is characterized by the reader creating their own unique understanding of the content.

From the perspective of Chen et al. (2014), inferential comprehension is clearly superior. However, the authors acknowledge that getting an accurate measurement is far more difficult for inferential comprehension than it is for literal comprehension. Literal comprehension researchers can quickly develop a standardized test that only asks questions about statements explicitly made within a text. In contrast, since the concept of inferential comprehension revolves around a user having a unique understanding of a topic, its measurement typically requires the use of open-ended questions.

Given the significant differences in both the style and comprehension objectives of digital reading, it should come as no surprise that the results of studies on digital reading comprehension have been less-than consistent. For example, Ben-Yehuda and Eshet-Alkalai (2014) sought to understand if digital annotation had the same benefits as print annotation. The study compared four groups, print-annotation, print-no-annotation, digital-annotation, and digital-no-annotation. After reading an 842-word text, subjects completed a ten question quiz, with five factual questions and five inferential questions. Researchers found that the print-annotation group performed significantly better than the print-non-annotation group, but the digital-annotation group did not out-perform the

digital-non-annotation group. After reviewing the findings, the researchers concluded that print annotation's benefits did not extend to digital annotation. However, a closer examination of the study reveals that the researchers offered no training on annotation to any of the subjects, and 63% of subjects reported that they "sometimes" or "very rarely" utilized digital annotation software.

Likewise, Kong et al. (2018) conducted a meta-analysis of reading comprehension studies and determined that print reading was better suited for comprehension. However, in explaining their findings, the authors proposed that "quite possibly the advantage of reading on paper could be accounted for by the readers' extensive experience of reading on paper, which shapes their preference for reading on paper and strengthened their use of reading on paper strategies" (p. 10).

Nichols (2016) conducted a review of literature on digital reading comprehension and concluded that not only is the learning gap between paper and digital reading shrinking, but it may not have ever existed in the first place. Instead, Nichols proposes that the issue is likely rooted in familiarity and training. He explains: "there is nothing inherently disadvantageous in on-screen reading except that readers tend to approach it differently," adding that "ultimately the solution lies in how learning designers leverage the on-screen experience to transcend what is possible in print" (p. 38, 39).

### The Role of Training and Familiarity on Digital Reading

Several studies suggest that when training and familiarity are sufficient, the outcomes for digital reading can be quite positive. Lauterman and Ackerman (2014) conducted a three-part study on comprehension to test the impact of familiarity on comprehension. In each round, subjects read two texts then completed an associated

comprehension test. The results of the first round of tests showed that digital readers scored significantly lower than print readers. In the second round, however, the comprehension rate for digital readers who were using their preferred format increased by 18%, making their scores statistically identical to the print readers. Researchers determined that familiarity was a significant factor in digital readers' comprehension scores.

Azevedo and Cromley (2004) sought to understand the impact of providing subjects with training on how to read hypertext. Both groups in the experiment completed a pretest and received 45 minutes to study a hypertext. The experimental group, however, received training on how to read hypertext. After completing the reading, subjects completed a post-test that consisted of matching, fill-in-the-blank, and open-ended questions. Results showed that subjects who received the training scored significantly higher and showed a substantially deeper understanding of the content than those in the control group.

Training that emphasizes the use of annotation features appears to be particularly impactful. Chen and Chen's (2014) study of fifth-grade students found that training on digital annotation features was sufficient to overcome the negative impact of digital reading on comprehension. Similarly, Johnson et al. (2010) tested the impact of modeling digital annotation methods on subjects' ability to annotate effectively. Results showed that annotation significantly impacted subjects' comprehension and metacognitive scores, but only when they received modeling from a peer or instructor. Researchers proposed that the subjects who examined other students' or their teacher's annotation approaches used them as a scaffolding for enhancing their own highlighting approach.

Beyond comprehension test scores, several studies suggest that subjects' reading experience is also enhanced when proper training is received. Dobler (2015) conducted a study in which 50% of subjects had utilized a digital textbook in the past, but just 22% claimed that digital reading was their preferred format. After providing subjects with training and modeling effective usage throughout a semester of instruction, 65% of the students felt that the digital text features enhanced their learning experience, and 50% reported that they now preferred to read from the digital text. Other studies have shown that the use of digital textbooks leads to significantly higher perceived affective learning scores as well (Rockinson-Szapkiw et al., 2011; Rockinsaw-Szapkiw et al., 2013).

Even when digital reading requires an increase in effort and cognitive load, the end result may be positive. Sidi et al. (2016) proposed that because of the inherent difficulties that readers have focusing on digital texts, readers may need an external trigger to encourage deeper thinking, a process they referred to as “desirable difficulties” (Bjork, 1994, 1999). Subjects in a qualitative study on digital reading voiced opinions that echoed that research. After the researcher required that subjects annotate their readings, one subject explained that "it is so much harder to fake read if you have to annotate like we have to do now. So now I actually read, because it's too hard to fake annotate" (Porter-O'Donnell, 2004, p. 87). Another subject opined that "I have learned how to be less distracted when I am reading. Annotating basically helps me comprehend and focus easier when I am reading" (Porter-O’Donnell, 2004, p. 87). In both cases, the annotation process served as a desirable difficulty by focusing the reader's attention on the text’s actual meaning.

Summary of a Strategy for Digital Reading

While digital reading faces serious challenges, the perspective offered through cognitive flexibility theory (Spiro, 2012) and situated cognition (Lemke, 1991) offers a viable lens through which the benefits of digital reading can be maximized. In many cases, this may require a slightly different perspective on comprehension, which focuses on a deeper, more personalized understanding of the content (Chen et al., 2014). Likewise, proper training is an essential element of maximizing the benefits of digital reading (Azevedo & Cromley, 2004, Johnson et al., 2010). Research shows that proper training should emphasize annotation (Chen & Chen, 2014), include modeling (Johnson et al., 2010), and be offered over an extended period rather than on a single occasion (Lauterman & Ackerman, 2014; Van Horne et al., 2016).

## Chapter Summary

While the literature on digital annotation and digital reading is far from unified, several conclusions can be obtained from examining it. Perhaps most importantly, digital reading is very different from print reading. Digital annotation features like tagging, hyperlinks, embedded dictionaries, search functions, and marginalia with unlimited space, encourage a non-linear reading style that is not easily supported in traditional print reading. Given the differences, the most prudent approach to examining the two formats is likely to focus less on which format is better and more on how the unique advantages of each format can be better used to achieve specific learning outcomes.

Importantly, an plethora of research has been conducted comparing digital and print reading that has failed to acknowledge the unique nature of digital reading. Likewise, an abundance of research has suffered from poor methodological design,

particularly regarding subjects' training on digital annotation features. In contrast, while some research has been conducted on non-generative annotations like highlighting, very little has been done on generative annotation approaches like tagging. In the coming chapter, this study will discuss a methodological approach in which this research gap can be properly addressed.

# CHAPTER THREE – METHODOLOGY

Chapter one introduced the purpose, research questions, and basic design of this study. Then, in chapter two, the existing literature behind digital reading and annotation was explored. This chapter combines elements of chapters one and two to provide the reader with a sound understanding of how this study was conducted and why each method was selected.

## Adaptations

Initially, this study intended to use a different approach to addressing the research questions. Subjects in the study were divided into three groups, with two groups receiving training on either tagging or highlighting and the third group serving as the control. The researcher then asked subjects to read a text on their phone. Subjects in the tagging group received instructions to tag passages that they felt were important to understanding the text. Subjects in the highlighting group received instructions to highlight passages they felt were important to understanding the text. The control group subjects received instructions to look for passages they felt were important to understanding the text. By giving a specific annotation invitation, the researcher anticipated that the data would clearly represent three distinct groups, highlighters, taggers, and non-annotators.

Unfortunately, when the researcher examined the number of tags and highlights each student reported making, it became clear that each group was not as exclusive as anticipated. For example, while subjects in the highlighting group created the most

highlights ($n = 47$, $m = 7.64$, $sd = 2.88$) subjects in the tagging group reported creating nearly as many highlights ($n = 45$, $m = 6.53$, $sd = 3.77$) and subjects in the control group also reported significant amounts of highlighting ($n = 1.81$, $sd = 2.95$). The overlap between the groups for tagging frequency was likewise concerning.

Rather than relying on data obtained from compromised experimental groups, the researcher determined that the study's data would be best examined by exploring the relationship between annotation frequency and comprehension test scores. The remainder of this chapter will discuss how that relationship was explored.

## Research Questions

This study uses a post-positivist worldview to answer the question, "What is the relationship between different smartphone reading annotation strategies and students' comprehension?" Annotation serves as the independent variable. Comprehension serves as the dependent variable and was measured by subjects' performance on the multiple-choice and essay assessments. Three sub-questions regarding two types of annotation (highlighting and tagging) are explored to resolve the main research question:

1. What difference is there between students who create an above-average number of highlights, a below-average number of highlights, and students who do not highlight on comprehension?
2. What difference is there between students who create an above-average number of tags, a below-average number of tags, and students who do not tag on comprehension?
3. Which text annotation feature (highlighting or digital tagging) is more beneficial to readers' comprehension?

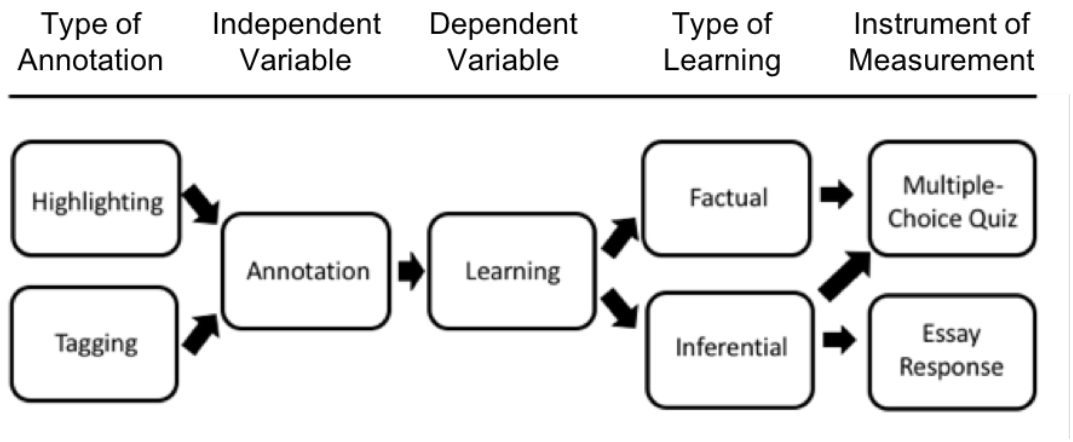


**Figure 3.1 Relationship of Variables, Methods, and Instruments**

In addition to examining the relationship between digital annotation features and readers' essay quality and quiz scores, these questions allow the researcher to compare tagging and highlighting relative to each other. Given the important role of training and modeling to the success of digital reading (Azevedo & Cromley, 2004; Bennett et al., 2008; Chen & Chen, 2014; Lauterman & Ackerman, 2014; Nichols, 2018; Rockinson-Szapkiw et al., 2013), this information is a valuable resource for educators seeking to understand which features to emphasize in their classrooms.

## Participants and Site

### Defining Released-time Seminary

The site for this study is a high school released-time seminary program in a mid-sized Southern Utah city. For context, released-time seminary is a common occurrence in much of the western United States, particularly in Utah, where nearly 100,000 high-school-aged students enroll in classes (Seminaries and Institutes of Religion, 2020). Since schools are not allowed to sponsor religious-based courses, The Church of Jesus Christ of Latter-d Saints (and some other religious organizations) hosts courses in privately-owned buildings that are typically adjacent to public schools. Students who wish to attend

courses are released from the high school for one period in their schedule and walk a short distance to the building to attend.

Appropriateness of Location

This site is a proper setting for testing the relationship of annotation and comprehension through the lens of cognitive flexibility theory for several reasons. First, in contrast to courses that use a text designed to convey specific facts, the nature of texts used in the program is designed to teach broader concepts that readers can apply in a variety of settings. Spiro (1988, 2012) described such texts as being complex and ill-structured and proposed that they provide the ideal setting for annotation and cognitive flexibility theory to thrive.

Second, the seminary program in-mention provides each student with a free digital version of the text via a mobile app (Gospel Library) and encourages teachers to train students on its use (Ashton, 2018; Olsen, 2018). The app utilized for the course is well-established and has more than three million active users. It contains a variety of content, including all of the church's scriptures, thousands of articles and speeches from the church's leaders, and significant amounts of audiovisual content (Olsen, 2018). In addition to the variety of content it provides, the app also allows users significant annotation options, including ten colors of highlighting and underlining, tagging, the ability to create hyperlinks, an embedded dictionary, a search function, a journal, the ability to create marginal notes, and others. Importantly, the app contains nearly identical operating interfaces for both IOS and Android operating systems.

The access and training that students in this setting received before the study provided an ideal environment to overcome the issues with familiarity that plagued

previous studies. To ensure that participants in the study received proper training and were familiar with the platform, the researcher utilized a convenience sample of 139 participants from six course sections taught by an instructor known for providing students with training on digital annotation. While this approach did not provide random selection at the time of the study, the program used a computer system to randomly assign students to classes at the beginning of each semester. Thus, the approach represents a reasonable balance between true random selection and the need to overcome the bias that occurs when students are not adequately trained and familiar with the experiment's reading platform.

## Training

Before beginning the study, the researcher provided participants with training on how to effectively tag and highlight. This training was provided in the eleven class sessions leading up to the day of the study. Appendix G provides QR codes and links to four training videos used to train students. In accordance with findings on the importance of modeling and providing practice opportunities, each training video displayed an instructor demonstrating how to annotate and concluded with participants completing a short practice exercise to ensure that they were comfortable with the process (Dobler, 2015; Johnson et al., 2010; Rockinson-Szapkiw et al. 2013). While not all training took place via pre-recorded videos, all of the training followed the model set out in the videos.

For example, one video training on how to highlight began by addressing technical aspects like "how to create a highlight" and "how to change the highlighting color." From there, the instructor trained students on how to identify the main ideas and important facts of a text. Finally, the instructor provided participants with a passage to

read and an assignment to practice the skill. Similar approaches were taken with tagging, with only slight modifications to the questions, so that the researcher asked students to "tag main ideas and important facts" rather than highlight them. Since participants in the study had prior experience with each feature, training videos were relatively short (five minutes or less) and designed to reinforce existing annotation skills rather than teaching students a new skill.

## Instructions

On the day of the experiment, the researcher directed subjects to read an 842-word scriptural text on their smartphone. The text, which came from the Book of Mormon, tells the story of a father and son discussing the importance of chastity (see appendix F). Before they began reading, the researcher informed subjects about the subsequent assessments, including the essay question topic. Subjects then received instructions to read the text and look for, highlight, or tag passages they felt were important to understanding the text's message. The researcher did not restrict participants to a single approach or implement any quotas for annotation but allowed participants to annotate as they pleased.

After the participants received their instructions, they began reading the assigned chapter. After studying the text, participants first completed the essay question. Once the essay was complete, subjects were asked to put away their smartphones and complete the multiple-choice assessment.

## Instrumentation

### Measuring Learning through the Lens of Cognitive Flexibility

Spiro's (1988) cognitive flexibility theory proposes that there are at least two types of learning. Surface-level learning is primarily concerned with memorizing content. However, deep learning focuses on the learners' ability to make inferences about the text and flexibly apply what they have learned to a multitude of settings (Spiro, 2012). Several authors propose that the purpose of digital annotation is to help readers organize information in a way that leads to this deeper level of learning (Niederhauser et al., 2000; Shapiro, 1998).

Unfortunately, measuring such learning has some inherent challenges. Since cognitive flexibility encourages learners to apply principles to different settings creatively, an appropriate assessment question can have any number of acceptable responses (Spiro, 2012). In contrast, most learning assessments use multiple-choice, matching, or fill-in-the-blank questions, which utilize a much more rigid approach. In these assessments, the participants have a limited number of options and are tasked with identifying a single correct response. Since these test's design makes it difficult to include questions that allow learners to flexibly apply their knowledge in any number of ways, such tests are (at least on their own) inadequate for measuring the type of deep learning proposed in CFT.

### Importance of Open-Ended Assessment Questions

Chen et al. (2014) explained that in order to overcome this setback, researchers should use open-ended, short answer, and essay questions. This study follows that recommendation by using a one-page essay to measure the depth of participants'

learning. Instructions for the essay asked participants to use the text they read as the basis for explaining a principle they could use to help resolve a problem in their life (see appendix B).

Two instructors then evaluated the essays using a rubric based on the question-answer relationship framework (Raphael, 1982). The question-answer relationship (QAR) framework was designed to classify the different levels of understanding demonstrated in writers' responses. In the QAR, the simplest level of understanding is a response citing a single factual reference in the text. In contrast, the most complete level of understanding combines several elements of the text with the writer's background knowledge to interpret and expand on the text. Table 3.1 (below) shows the four levels of comprehension described in the QAR. Schugar et al. (2011) utilized these definitions to separate the quality of thoughts in an essay into the four different categories outlined in the QAR. This study utilizes the same approach, using a grading rubric based on the QAR to assign participants' essays to one of four levels. Two instructors graded each of the essays, and the researcher generated a Pearson correlation coefficient to calculate an inter-rater reliability score to ensure the validity of the instrument. A copy of the rubric is provided in table 3.2 (below), and examples of each level of response can be seen in appendix D.

**Table 3.1 Question-Answer Relationship Framework**

| Answer Level | Expectation |
|---|---|
| Level 1 | The answer is pulled from information that is explicitly stated in one place in the text. |
| Level 2 | The answer is developed by making inferences from information found in multiple places throughout the text. |
| Level 3 | The answer is developed by combining participants' background knowledge with information found in the text. |
| Level 4 | The answer extends text using critical thinking and background knowledge. |

**Table 3.2 Grading Rubric for Essay Question**

| Answer Level | Expectation |
|---|---|
| Level 1 | The answer is pulled from a single reference in the chapter. If application is made, it is nearly identical to the context of the application provided in the text. |
| Level 2 | The answer is developed from combining information from multiple references in the chapter. If application is made, it is nearly identical to the context of the application provided in the text. |
| Level 3 | The answer is developed using information that is not contained in the original text. The application is still nearly identical to the context provided in the text. |
| Level 4 | The answer uses a principle from the text but provides an application of the principle that goes beyond the context that was presented in the text. |

Importance of Multiple-Choice Assessment

While the process above offers a legitimate measure of the learning cognitive flexibility theory seeks to support, the instrument has a significant downside. Schugar et al. (2011) utilized the method in their study on reading comprehension. They reported that "on the whole, students' responses were fairly low-leveled, and few students were

able to provide responses that looked at the material beyond what was explicitly and implicitly stated in the text" (p. 183). Thus, to ensure that valid data was obtained, in addition to the essay question, this study follows the lead of other successful studies and includes a multiple-choice quiz as well (Ben-Yehudah & Eshet-Alkalai, 2014; Lauterman & Ackerman, 2014; Mueller & Oppenheimer, 2014; Noyes & Garland, 2003). In addition to ensuring a more accurate representation of the data, using an instrument with clearly defined correct and incorrect responses simplifies the statistical analysis by removing any bias that may come from a subjective grader.

A team of three experienced instructors designed the quiz in consultation with an established curriculum to ensure high content validity. The designers began by independently studying the text used for the study and noting the major factual elements and principles taught in the text, both explicitly and implicitly. The designers then convened and compared each designer's initial thoughts with the course's established curriculum. Just as this study used the QAR framework to determine different levels of learning when assessing the essay question, quiz designers sought to measure two levels of learning by developing six questions that measured factual learning and six questions that measured inferential learning.

The factual questions focused on explicitly stated elements of the text. Given their nature, factual questions are best used to measure recall, or what some refer to as "surface-level learning" (Spiro, 1988). In contrast, the inferential questions focused on applying principles and lessons that were implied in the reading. These questions are more effective at measuring students' ability to flexibly apply what they have learned in a

variety of new settings (Spiro, 2012). For example, one factual question asked, “which of the following does Alma tell Corianton not to seek after?”

In contrast, one inferential question asked, “Which of the following four characters is most likely to benefit from reading Alma’s counsel to Corianton?” then provided four short character descriptions. By utilizing both types of questions, the researcher was able to gain a better understanding of the specific impact each annotation method had on subjects’ learning. A copy of the questions utilized in the multiple-choice assessment is available in appendix A.

### Using Multiple Perspectives to Understand the Bigger Picture

As mentioned previously, multiple-choice questions are, in many ways, an insufficient way to measure flexible learning. Still, the use of both factual and inferential questions in the assessment provided insight into how annotation impacts a student’s ability to understand the meaning of a text beyond what is explicitly stated. When used together with the essay question, this method provides an insightful perspective on participants’ total learning.

Finally, in addition to the data obtained from the designated instruments, participants answered a series of demographic questions on a variety of topics. Such questions included gender, grade-level, confidence in digital annotation skills, and preferred format. A copy of the specific questions used is available in appendix A.

## Data Entry and Analysis

### Measuring Validity and Reliability

After the researcher entered the numerical data for the essay, multiple-choice quiz, and demographic questions into SPSS, statistical analysis was conducted. Since the

quiz utilized two specific types of questions, principal component analysis (PCA) was used to ensure the instruments validity in measuring factual and inferential learning. Four assumptions must be met to conduct PCA: a 10:1 subject to component ratio, a correlation between the factors being considered, linearity between the variables being considered, and a lack of outliers (Abdi & Williams, 2010; Al-Sarmi & Al-Hemyari, 2014; Osborne & Costello, 2004; Statistics Solutions, N.D.). This study meets each of these criteria. The use of PCA resulted in the elimination of two questions from each of the constructs. Once the final questions for each construct were set, the researcher generated a score for how each participant performed on the factual comprehension questions and the inferential comprehension questions in the quiz.

Ruel et al. (2015) explained that while validity is the primary concern for closed-ended assessments, reliability is more often the concern for open-ended questions. Since this study utilizes an essay question, and grading essays has a certain level of subjectivity, it was important to establish a sufficient inter-rater reliability level (Ruel et al., 2015). In this study, the inter-rater reliability was calculated using the scores created from the two graders to generate a Pearson correlation coefficient score.

Creating and Comparing Groups

Once the researcher performed the preliminary tests for reliability and validity, the researcher compared differences in participants' multiple-choice quizzes and essay scores based on the number of tags and highlights each participant created. Non-highlighters/taggers were the first groups. After generating the mean number of highlights and tags created by the remaining students, the remaining students were divided into above-average and below-average groups based on the number of highlights

and tags they created. The researcher then compared each group's mean scores on the assessments to determine statistical differences.

When just two groups were compared, an independent sample t-test was used to compare means. Four assumptions must be met to perform an independent sample t-test. Each observation must exist in only one of the groups. There must be no significant outliers. The data must be normally distributed, and the variances between the groups should be equal. This sample meets those expectations.

When the researcher compared three groups, this study utilized the ANOVA statistic. ANOVA was selected because of its ability to determine if a statistically significant difference exists between three or more groups. ANOVA testing has assumptions that are very similar to independent sample t-tests. Specifically, the test requires samples that are independent of each other, normally distributed, and have equality of variance across each group (Fox et al., 2003). When the researcher identified statistical differences using ANOVA, Tukey's honestly significant difference (HSD) test was performed to confirm findings. Tukey's HSD is a post hoc analysis used to examine if a group's score is actually different from each of the other groups or if some of the groups are statistically the same.

Demographics and Confounding Variables

In some cases, the data from this study suggested that gender may be influencing test results. When this was the case, the researcher used ANCOVA testing to control for gender differences. In addition to the assumptions for ANOVA, ANCOVA requires that covariates be independent of each other. When this test was applied, that assumption was met.

Additionally, the researcher analyzed responses from the questionnaire regarding gender, grade-level, confidence with digital annotation, preferred format, and annotations made. In addition to allowing the researcher to identify potentially confounding variables, this data helped resolve research question three by identifying differences in the relationship between highlighting and tagging and different demographic groups. Understanding that relationship was useful in making suggestions to educators regarding when and with whom to utilize digital annotation. While significant results were found regarding the impact of each of these factors, annotation confidence was particularly important to the study.

Analysis of these factors used two approaches. When the researcher compared means, analysis utilized independent sample t-tests and ANOVA testing. However, the researcher also used Pearson's chi-square statistic to determine if the proportion of students in different groups was significantly different. Pearson's chi-square statistic has just two required assumptions: the data must be frequencies, not percentages, and the categories must be mutually exclusive. In each case that Pearson's chi-square test was used, these assumptions were met.

Table 3.3 (below) shows how the relationship between each of the metrics and tests used in the study and each of the research questions. When a checkmark is present, the metric or test listed to the left was used to help resolve the question listed at the top of the table.

**Table 3.3 Variables and Statistics for Each Research Question**

| | | Question #1 | Question #2 | Question #3 |
|---|---|---|---|---|
| Independent Variables | Highlighting Frequency | ✓ | | ✓ |
| | Tagging Frequency | | ✓ | ✓ |
| | Gender | ✓ | ✓ | ✓ |
| | Grade Level | | | ✓ |
| | Annotation Confidence | | | ✓ |
| | Highlighting Confidence | ✓ | | ✓ |
| | Tagging Confidence | | ✓ | ✓ |
| | Format Preferences | | | ✓ |
| Dependent Variabless | Multiple Choice Factual Score | ✓ | ✓ | ✓ |
| | Multiple Choice Inferrential Score | ✓ | ✓ | ✓ |
| | Multiple Choice Total Score | ✓ | ✓ | ✓ |
| | Essay Score | ✓ | ✓ | ✓ |
| Statistical Tests | ANOVA | ✓ | ✓ | ✓ |
| | Tukey's HSD | ✓ | ✓ | ✓ |
| | ANCOVA | ✓ | ✓ | ✓ |
| | T-Tests | ✓ | ✓ | ✓ |
| | Chi-Square | ✓ | | |

## Ethical Concerns & Limitations

### Minimizing Impact on the Research Location

The researcher designed this study in a way that presented minimal ethical concerns. For example, the loss of instruction time was a concern for both the host of the experiment and the participants. To resolve this concern, the researcher used a text (Alma 39) aligned with the instructors' projected teaching schedule for the week of the study. Similarly, since students in the course utilize their personal version of the text, not all formats were identical. While the operating interface of the app was nearly identical for both IOS and Android smartphones, some students in the course chose to use a paper version of the text rather than a digital version. To ensure that no student felt left out, the researcher provided an equivalent paper version of the activity for students who did not have access to a smartphone on the day of the study. While responses from the equivalent activity were not used in the study, the replacement activity ensured that no student lost learning opportunities because of the experiment. By taking these measures, the researcher eased the ethical concern of placing an undue burden on the experiment's hosts (Creswell, 2017).

### Obtaining Consent

More significantly, the utilization of participants who were minors necessitated that the researcher obtain consent from a guardian to use content generated from the study. As part of the course registration form, signed by each student’s guardian before they began taking classes, administrators informed parents that research studies and surveys might be conducted as part of the course and that participation in such studies is voluntary. Administrators then allowed parents to withdraw consent for their minor to

participate in any such studies. In addition to the consent provided in the previously mentioned form, the researcher provided parents and guardians with an additional consent form containing information regarding the study's purpose and methods. The researcher initially delivered this form to guardians via a printed form sent home with their students. The researcher then made additional requests to obtain permission via text, and phone call, depending on the parent's previously stated preference. Hard copies of both consent forms and participants' responses will be stored at the site of the study, and a digital copy was created and stored by the researcher. A copy of the consent form is available in appendix C.

While the study's risks are relatively minor, before beginning the study, the researcher received approval from the Boise State Institutional Review Board (IRB). Likewise, once the Boise State IRB granted approval, institutional approval was granted by Seminaries and Institutes of Religion.

## Chapter Summary

This chapter has outlined the methodological approach used in this study. The study's primary focus was the relationship between different smartphone reading annotation strategies and students' comprehension. Annotation served as the independent variable, with tagging and highlighting serving as different annotation options. Comprehension served as the dependent variable, with the multiple-choice quiz and essay question serving as the measurement instrument.

Participants for the study were 139 teenage students attending a released-time seminary course in Southern Utah. After participants completed training and practice

exercises on how to annotate, the researcher asked participants to read an 842-word text and highlight, tag, or look for passages that were important to understanding the text.

The analysis of learning consisted of two parts. Two instructors graded each essay by assigning it to one of the four categories of the question-answer framework. A short multiple-choice quiz with factual and inferential questions was also graded. Since participants were minors, the researcher obtained a guardian's permission for each minor participant. Participants who were unable to participate in the study were provided with an equivalent exercise to complete during class.

# CHAPTER FOUR: RESULTS

## Demographic Breakdown

The results of this study were obtained using the practices outlined in chapter three. On the day of the study, 143 students attended class. Four students opted not to participate in the study, leaving a sample of 139 high-school-aged students ($N = 139$). Figure 4.1 (below) shows the distribution of the sample by gender and grade level. Males ($n = 79$) represent 56.8% of the sample, while females ($n = 60$) represent 43.2%. Subjects in tenth grade (sophomores) represent 45.3% of the sample ($n = 63$), while eleventh grade students (juniors) represent 32.4% ($n = 45$), and twelfth grade students (seniors) represent 22.3% of the sample ($n = 31$). These numbers align with the demographics of the associated public school, which is in its second year of operation.

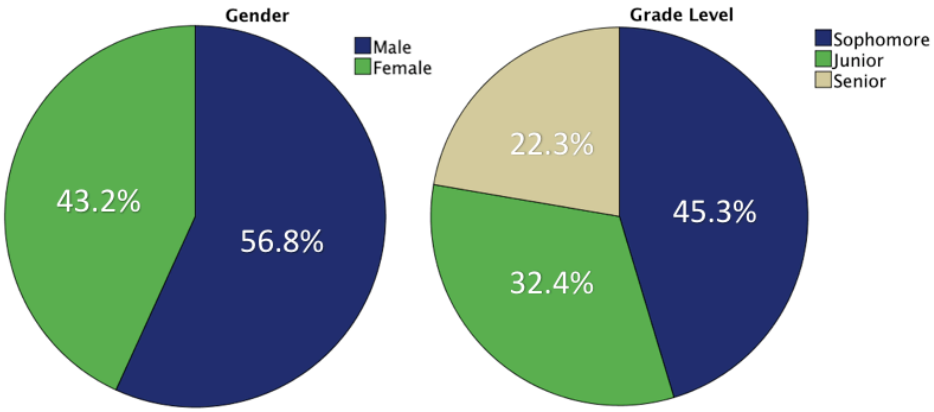


**Figure 4.1 Gender and Grade Level Distribution**

Basic Instructions

On the day of the study, the researcher asked participants to read an 842-word scriptural text on their smartphone. Participants also received instructions to look for, highlight, or tag passages they felt were important to understanding the text's message. The researcher did not restrict participants to a single approach or implement any quotas for annotation. After completing the reading, subjects wrote a short essay explaining how they could use the text's teachings to resolve a problem in their lives. When subjects completed the essay, the researcher asked them to close the text and provided them with the multiple-choice assessment and demographic questions.

## Validation and Reliability Testing

Principle Component Analysis

Three experienced teachers designed the multiple-choice assessment. The initial assessment (see appendix A) contained a total of twelve questions. Questions one through six were the factual questions, while questions seven through twelve were the inferential questions. This allowed the researcher to utilize ANOVA and T-tests to compare scores for total comprehension, factual comprehension, and inferential comprehension based on the number of tags or highlights a participant created. Importantly, the study was designed to eliminate questions that did not generate consistent results. The initial analysis determined that question twelve was likely not well understood by participants since just 34.5% ($n$ = 48) of participants selected the correct answer. In comparison, participants selected two of the incorrect responses 36.7% ($n$ = 51) and 25.9% ($n$ = 36) of the time. As a result, the researcher decided to remove question twelve from the survey.

The researcher used the remaining eleven questions to conduct principal component analysis (PCA) with results set to extract two factors. The purpose of this analysis was to verify that the factual and inferential elements of the assessment were unique. The Kaiser-Meyer-Olkin measure of sampling adequacy demonstrated that the set of variables was adequately related for factor analysis $kmo(21) = .602$, $p < .05$. Eliminating questions five and six from the factual questions and questions eight and nine from the inferential questions led to the highest explained variance (43.5%). Per the results of PCA, those questions were removed from the study. Tables 4.1 and 4.2 highlight the results of principal component analysis.

**Table 4.1 Total Variance Explained**

| Component | Total | % of Variance | Cumulative % |
|---|---|---|---|
| 1 | 1.79 | 25.61 | 25.61 |
| 2 | 1.25 | 17.89 | 43.50 |
| 3 | .99 | 14.27 | 57.78 |
| 4 | .88 | 12.65 | 70.43 |
| 5 | .79 | 11.35 | 81.78 |
| 6 | .67 | 9.68 | 91.47 |
| 7 | .59 | 8.52 | 100.00 |

Extraction Method: Principal Component Analysis

**Table 4.2 Principal Component Matrix**

Rotated Component Matrix

| | Component | |
|---|---|---|
| | 1 | 2 |
| Question 2 | .702 | -.200 |
| Question 4 | .634 | .372 |
| Question 3 | .587 | -.025 |
| Question 1 | .443 | .313 |
| Question 10 | -.115 | .769 |
| Question 7 | .022 | .683 |
| Question 11 | .294 | .420 |

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Inter-rater Reliability

In the week following the study, two instructors utilized a custom rubric based on the question-answer framework to rate each student's essay's depth on a one-to-four scale (see appendix D). While the subjects' official scores were generated by averaging the two grader's scores, the graders showed a significant amount of inter-rater reliability. Before beginning, graders reviewed the different levels of learning outlined in the question-answer relationship framework. The researcher provided an example of each level and allowed for questions. Graders then independently graded each essay and assigned it to one of the four levels defined in the QAR framework. Once grading for the essay was completed, the graders compared their coding and discussed any decisions that were not in agreement. The two graders reached a consensus on 133 of the 139 essays, which yielded a Pearson correlation coefficient score of $r = .98$, $n = 139$, $p < .000$. An average

of the graders' scores was used for the remaining six essays when comparing means. When analysis required that each score fit into one of the four categories, the higher of the two scores was used.

**Table 4.3 Inter-Rater Reliability Correlation Score**

| | | Essay Score 1 | Essay Score 2 |
|---|---|---|---|
| Grader 1 | Pearson Correlation | 1 | .982 |
| | Sig (2-Tailed) | | .000 |
| | N | 139 | 139 |
| Grader 2 | Pearson Correlation | .982** | 1 |
| | Sig (2-Tailed) | .000 | |
| | N | 139 | 139 |

**. Correlation is significant at the 0.01 level (2-tailed)

Comparing Means

The study's primary analysis utilized the number of tags and highlights each participant made as the measure of the independent variable. Additionally, the subjects' confidence in their annotation skills was used in some cases. In each case, comprehension was the dependent variable and was measured by the multiple-choice and essay scores.

This study frequently reduced the data from ordinal and scale variable questions from nine or eleven groups to two or three groups to increase the readability of the analysis. This reduction facilitated the ability to compare means amongst groups with sufficient sample sizes. In this study, when the independent variable consists of just two groups, an independent sample t-test is utilized to compare the means. When the independent variable had three or more groups, ANOVA testing was performed; when

ANOVA results were significant at the $p < .05$ level, post hoc analysis using Tukey's HSD was performed to determine which groups differ.

While the multiple-choice data analysis focused entirely on means comparison, the essay used two distinct approaches; when essay scores were treated as a scale variable, ANOVA was utilized to compare participants' mean scores. In contrast, since each numerical grade is associated with a specific expectation, essay scores can also be viewed as an ordinal variable. When this is the case, chi-square analysis was performed to determine if the proportion of students who received each score (one, two, three, or four) is different based on the number of annotations made.

Chapter Outline

The remainder of this chapter will report the results of the study. Results will be presented in relation to the sub-research questions outlined in chapter one:

1. What difference is there between students who create an above-average number of highlights, a below-average number of highlights, and students who do not highlight on comprehension?
2. What difference is there between students who create an above-average number of tags, a below-average number of tags, and students who do not tag on comprehension?
3. Which text annotation feature (highlighting or digital tagging) is more beneficial to readers' comprehension?

While this chapter reports the statistical testing results, it is important to note that it generally avoids providing commentary or analysis on findings. Discussion and analysis of these results will take place primarily in chapter five of this study.

**Question #1 – What is the relationship between highlighting and comprehension?**

Multiple-Choice Scores by Annotations Created

To measure the relationship between highlighting and comprehension, highlighting frequency was divided into three groups. The first group ($n$ = 36) consisted of students who created no highlights. The researcher divided the remaining students into two groups based on whether they created more ($n$ = 49) or less ($n$ = 54) than the average number of highlights. Students who create more than the average number of highlights were labeled "high-frequency highlighters," while students who created less than the average number of highlights were labeled "low-frequency highlighters."

Tables 4.4 and 4.5 show the results of ANOVA testing, which revealed that high-frequency highlighters scored statistically higher in factual ($m$ = 3.12, $sd$ = .927) comprehension than low-frequency highlighters ($m$ = 2.35, $sd$ = 1.10) and non-highlighters ($m$ = 2.36, $sd$ = 1.17) $f(2, 136) = 8.25$, $p < .000$. ANOVA results were also significant for total comprehension $f(2, 136) = 7.81$, $p = .001$, but did not reach significance for inferential comprehension $f(2, 136) = 2.02$, $p = .136$. Importantly, post hoc analysis using Tukey's HSD determined that differences between low-frequency highlighters and non-highlighters were not significant.

**Table 4.4 Mean Scores by Number of Highlights Created**

| | | | | | 95% Confidence | |
|---|---|---|---|---|---|---|
| | | N | Mean | Std. Deviation | Lower Bound | Upper Bound |
| Factual | No Highlights | 36 | 2.36 | 1.17 | 1.96 | 2.76 |
| | Low-Frequency | 58 | 2.38 | 1.10 | 2.09 | 2.67 |
| | High-Frequency | 45 | 3.16 | .90 | 2.88 | 3.43 |
| | Total | 139 | 2.63 | 1.11 | 2.44 | 2.81 |
| Inferential | No Highlights | 36 | 1.47 | .971 | 1.14 | 1.80 |
| | Low-Frequency | 58 | 1.67 | 1.03 | 1.40 | 1.94 |
| | High-Frequency | 45 | 1.91 | .92 | 1.63 | 2.19 |
| | Total | 139 | 1.70 | .99 | 1.53 | 1.86 |
| Total | No Highlights | 36 | 3.83 | 1.55 | 3.31 | 4.36 |
| | Low-Frequency | 58 | 4.05 | 1.75 | 3.59 | 4.51 |
| | High-Frequency | 45 | 5.07 | 1.25 | 4.69 | 5.44 |
| | Total | 139 | 4.32 | 1.63 | 4.05 | 4.60 |

**Table 4.5 ANOVA Scores for Number of Highlights Created**

| | | Sum of Squares | DF | Mean Square | F | Sig |
|---|---|---|---|---|---|---|
| Factual | Between Groups | 18.66 | 2 | 9.33 | 8.24 | .000 |
| | Within Groups | 153.88 | 136 | 1.13 | | |
| | Total | 127.54 | 138 | | | |
| Inferential | Between Groups | 4.34 | 2 | 2.17 | 2.25 | .108 |
| | Within Groups | 130.96 | 136 | .96 | | |
| | Total | 135.30 | 138 | | | |
| Total | Between Groups | 39.51 | 2 | 19.75 | 8.21 | .000 |
| | Within Groups | 326.91 | 136 | 2.40 | | |
| | Total | 366.43 | 138 | | | |

**Table 4.6 Tukey Results for Highlighting on Multiple-Choice Comprehension**

| Dependent Variable | (I)Highlighting Frequency | (J)Highlighting Frequency | Mean Difference (I-J) | Std. Error | Sig |
|---|---|---|---|---|---|
| Factual | None | Low-Frequency | -.01 | .22 | .99 |
| | | High-Frequency | -.79* | .238 | .003 |
| | Low-Frequency | None | .01 | ,22 | ,99 |
| | | High-Frequency | -.77* | .21 | .00 |
| | High-Frequency | None | .79* | .23 | .00 |
| | | Low-Frequency | .77* | .21 | .00 |
| Inferential | None | Low-Frequency | -.20 | .20 | .60 |
| | | High-Frequency | -.43 | .22 | .11 |
| | Low-Frequency | None | .20 | .20 | .60 |
| | | High-Frequency | -.23 | .19 | .44 |
| | High-Frequency | None | .43 | .22 | .11 |
| | | Low-Frequency | .23 | .19 | .44 |
| Total | None | Low-Frequency | -.21 | .33 | .78 |
| | | High-Frequency | -1.23* | .34 | .00 |
| | Low-Frequency | None | .21 | .33 | .78 |
| | | High-Frequency | -1.01* | .30 | .00 |
| | High-Frequency | None | 1.23* | .348 | .00 |
| | | Low-Frequency | 1.01* | .30 | .00 |

*. The mean difference is significant at the .05 level.

## Highlighting Frequency and Essay Scores

In addition to the significant relationship between highlighting frequency and multiple-choice scores, there was also a significant relationship between highlighting frequency and subjects' essay scores. ANOVA testing showed that the difference in

scores between the three highlighting frequency groups was statically significant $f(2, 136) = 3.38$, $p = .037$. Post hoc analysis revealed that high-frequency highlighters ($m = 2.41$, $sd = 1.09$) scored significantly higher than non-highlighters ($m = 1.81$, $sd = .87$) but low-frequency highlighters ($m = 2.03$, $sd = 1.10$) were not statistically different than the other two groups.

**Table 4.7 The Relationship of Highlighting Frequency and Essay Scores**

| | | | | | 95% Confidence Interval for mean | |
|---|---|---|---|---|---|---|
| | N | Mean | Std. Dev. | Std. Error | Lower Bound | Upper Bound |
| No Highlights | 36 | 1.81 | .87 | .14 | 1.52 | 2.11 |
| Low-Frequency | 58 | 2.03 | 1.10 | .13 | 1.73 | 2.32 |
| High-Frequency | 45 | 2.41 | 1.09 | .16 | 2.08 | 2.74 |
| Total | 139 | 2.10 | 1.06 | .09 | 1.92 | 2.27 |

**Table 4.8 ANOVA for Essay Scores by Highlighting Frequency**

| | Sum of Squares | DF | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 7.43 | 2 | 3.71 | 3.38 | .03 |
| Within Groups | 149.65 | 136 | 1.10 | | |
| Total | 157.09 | 138 | | | |

**Table 4.9 Tukey's HSD for Essay Scores by Highlighting Frequency**

| | | | | | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| (I)Highlighting Frequency | (J)Highlighting Frequency | Mean Difference | Std. Error | Sig. | Lower Bound | Upper Bound |
| None | Low-Frequency | -.21 | .22 | .59 | -.74 | .31 |
| | High-Frequency | -.59* | .23 | .03 | -1.14 | -.-3 |
| Low-Frequency | No Highlights | .21 | .22 | .59 | -.31 | .74 |
| | High-Frequency | -.37 | .20 | .17 | -.87 | .11 |
| High-Frequency | No Highlights | .59* | .23 | .03 | .03 | 1.14 |
| | Low-Frequency | .37 | .20 | .17 | -.11 | .87 |

*. The mean difference is significant at the 0.05 level

T-test analysis of assessment scores by gender revealed no significant differences between females and males on the multiple-choice portion of the assessment. However, females ($n$ =60, $m$ = 2.40, $sd$ = 1.35) scored significantly higher than males ($n$ = 79, $m$ = 1.86, $sd$ = 1.01) on the essay portion of the assessment $t(137) = -.52$, $p < .00$. Since gender differences for the essay portion were significant, the researcher utilized ANCOVA to control for gender. Results indicated that the relationship between highlighting and essay performance was still significant.

**Table 4.10 Assessment Scores by Gender**

| | Gender | N | Mean | Std. Dev | Std. Error |
|---|---|---|---|---|---|
| Factual | Male | 79 | 2.58 | 1.22 | .13 |
| | Female | 60 | 2.68 | .96 | .12 |
| Inferential | Male | 79 | 1.61 | .98 | .11 |
| | Female | 60 | 1.82 | 1.00 | .12 |
| MC Total | Male | 79 | 4.19 | 1.80 | .20 |
| | Female | 60 | 4.50 | 1.35 | .17 |
| Essay | Male | 79 | 1.86 | 1.01 | .11 |
| | Female | 60 | 2.40 | 1.06 | .13 |

**Table 4.11 T-Test Results for Assessments by Gender**

| | | | | | 95% Confidence | |
|---|---|---|---|---|---|---|
| | Equal Variances? | t | df | Sig (2-tailed) | Lower | Upper |
| Factual | Assumed | -.52 | 137 | .59 | -.48 | .27 |
| | Not Assumed | -.54 | 136.80 | .58 | -.46 | .26 |
| Inferential | Assumed | -1.23 | 137 | .21 | -.54 | .12 |
| | Not Assumed | -1.23 | 125.80 | .22 | -.54 | .12 |
| MC Total | Assumed | -1.11 | 137 | .26 | -.86 | .24 |
| | Not Assumed | -1.15 | 136.99 | .25 | -.84 | .22 |
| Essay Score | Assumed | -3.05 | 137 | .00 | -.89 | -.19 |
| | Not Assumed | -3.03 | 123.97 | .00 | -.89 | -.18 |

**Table 4.12 ANCOVA Results for Essay Scores by Gender**

| Source | Type III Sum of Squares | df | Mean Square | F | Sig |
|---|---|---|---|---|---|
| Corrected Model | 17.87* | 2 | 8.93 | 8.73 | .00 |
| Intercept | 14.74 | 1 | 14.74 | 14.40 | .00 |
| Gender | 8.81 | 1 | 8.81 | 8.61 | .00 |
| Highlights | 7.88 | 1 | 7.88 | 7.70 | .00 |
| Error | 139.21 | 136 | 1.02 | | |
| Total | 770.50 | 139 | | | |
| Corrected Total | 157.09 | | | | |

*. R Squared = .11 (Adjusted R Squared = .10)

In addition to a comparison of means, analysis of the proportion of subjects who reached the highest comprehension level yielded several significant results. Table 4.14 shows that 90.4% ($n$ = 19) of the participants who scored a four on their essay made at least one highlight, and 57.1% ($n$ = 12) created nine or more highlights. In contrast, just 5.6% ($n$ = 2) of participants who did not create at least one highlight produced a level-four essay, while 24.5% ($n$ = 12) of participants who made at least nine highlights produced a level-four essay. Chi-square analysis found the results to be significant $\chi^2(2, 138) = 6.036$, $p = .049$.

**Table 4.13 Percentage of Students Who Scored a Four on the Essay**

| | | | Level Four? | | |
|---|---|---|---|---|---|
| | | | No | Yes | Total |
| Highlighting Frequency | None | Count | 34 | 2 | 36 |
| | | % Within Frequency | 94.4% | 5.6% | 100.0% |
| | | % Within Level 4 | 29.1% | 9.5% | 26.1% |
| | Low-Frequency | Count | 46 | 7 | 53 |
| | | % Within Frequency | 86.8% | 13.2% | 100.0% |
| | | % Within Level 4 | 39.3% | 33.3% | 38.4% |
| | High-Frequency | Count | 37 | 12 | 49 |
| | | % Within Frequency | 75.5% | 24.5% | 100.0% |
| | | % Within Level 4 | 31.6% | 57.1% | 35.5% |
| Total | | Count | 117 | 21 | 138 |
| | | % Within Frequency | 84.8% | 15.2% | 100.0% |
| | | % Within Level 4 | 100.0% | 100.0% | 100.0% |

Confidence in Highlighting

Since a significant premise for this study was the importance of subjects feeling confident in their ability to annotate, question nine of the survey asked students to report their confidence in their highlighting skills using a nine-point Likert scale. Participants generally reported very high confidence in their highlighting skills ($M = 7.99$, $SD$ 1.45), with more than half of students ($N = 71$, 51.1%) assigning themselves a nine, and just 12% ($N = 16$) reporting a confidence level lower than seven. This finding supports the premise that students in the study were comfortable highlighting.

To determine if the relationship between highlighting confidence and assessment scores was significant, the researcher divided subjects into two groups based on whether

their reported highlighting confidence was above or below average. High-confidence highlighters consisted of participants who reported confidence scores of eight or nine ($n$ = 105). Low-confidence highlighters consisted of participants who reported confidence scores of one through seven ($n$ = 34). Tables 4.14 and 4.15 show the results of highlighting confidence on the multiple-choice assessment. In each of the three categories, the groups' differences failed to reach significance at the $p < .05$ level.

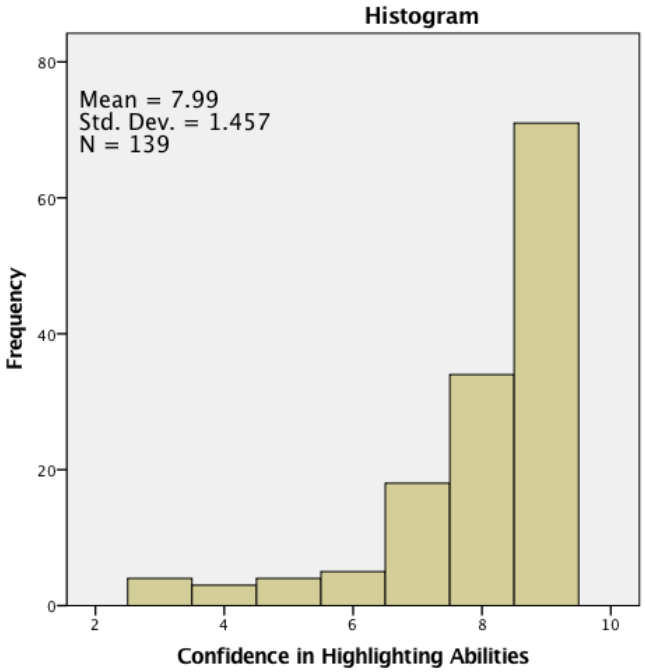


**Figure 4.2 Confidence in Highlighting Skills**

**Table 4.14 Mean Multiple-Choice Scores by Highlighting Confidence**

| | Confidence Highlighting | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Factual | Low-Confidence | 34 | 2.38 | 1.81 | .20 |
| | High-Confidence | 105 | 2.70 | 1.09 | .10 |
| Inferential | Low-Confidence | 34 | 1.65 | 1.04 | .17 |
| | High-Confidence | 105 | 1.71 | .97 | .09 |
| Total | Low-Confidence | 34 | 4.03 | 1.86 | .32 |
| | High-Confidence | 105 | 4.42 | 1.54 | .15 |

**Table 4.15 T-test for Multiple-Choice Scores by Highlighting Confidence**

| | | | | t-test for equality of means | | |
|---|---|---|---|---|---|---|
| | Equal Variance? | F | Sig | t | df | Sig (2-Tailed) |
| Factual | Assumed | .96 | .32 | -1.46 | 137 | .14 |
| | Not Assumed | | | -1.40 | 52.49 | .16 |
| Inferential | Assumed | .60 | .43 | -.34 | 137 | .73 |
| | Not Assumed | | | -.33 | 53.16 | .74 |
| Total | Assumed | 1.51 | .22 | -1.21 | 137 | .22 |
| | Not Assumed | | | -1.10 | 48.46 | .27 |

In contrast to the relationship between highlighting confidence and multiple-choice scores, the relationship between highlighting confidence and essay scores was significant. Subjects who reported high highlighting confidence ($n$ = 105) scored higher on their essays ($m$ = 2.20, $sd$ = 1.04) than subjects who reported low highlighting confidence ($n$ = 34, $m$ = 1.76, $sd$ = 1.06). T-test analysis confirmed that the differences were significant $t(137) = -2.14$, $p = .034$.

**Table 4.16 Mean Essay Scores for High and Low Confidence Highlighters**

| | Confidence Highlighting | N | Mean | Std. Deviation | Std. Error |
|---|---|---|---|---|---|
| Essay Score | Low-Confidence | 34 | 1.76 | 1.06 | .18 |
| | High-Confidence | 105 | 2.20 | 1.08 | .10 |

**Table 4.17 T-Test Results for Essay Scores Based on Highlighting Confidence**

| | | Levine's Test for Equality of Variance | | T-test for Equality of Means | | | | |
|---|---|---|---|---|---|---|---|---|
| | Equal Variances? | F | Sig. | t | df | Sig. (2-Tailed) | Mean Difference | Std. Error |
| Essay Score | Assumed | .33 | .56 | -2.14 | 137 | .034 | -.44 | .20 |
| | Not Assumed | | | -2.12 | 55.13 | .04 | -.44 | .20 |

Summary of Results on the Relationship of Highlighting and Comprehension

The results of this study regarding the relationship of highlighting and comprehension yielded several important findings. High-frequency highlighters scored significantly higher on factual and total comprehension on the multiple-choice assessment but did not score significantly higher or lower on inferential comprehension. High-frequency highlighters also scored higher than low-frequency and non-highlighters on the essay portion of the assessment. This finding held when controls were implemented for gender.

Finally, subjects across all groups reported high levels of confidence in their highlighting skills. However, high confidence was not related to higher scores on the multiple-choice portion of the assessment. In contrast, high confidence was related to significantly higher essay scores.

**Question #2 – What is the relationship between tagging and comprehension?**

To determine the relationship between tagging and comprehension, the researcher divided participants into three groups based on the number of tags they reported making. The first group consisted of subjects who created no tags ($n$ = 65) and received the label "non-taggers." The remaining students were divided based on whether they created more or less than the average number of tags. Subjects who created more than the average number of tags ($m$ = 5.5, $n$ = 34) were deemed "high-frequency taggers," while subjects who created less than the average number of tags ($n$ = 40) were deemed "low-frequency taggers."

Results of t-test analysis showed that high-frequency taggers ($n$ =34, $m$ = 2.00, $sd$ = 1.04) scored higher than no/low-frequency taggers ($n$ = 105, $m$ = 1.60, $sd$ = .95) on inferential comprehension $t(137) = 2.07$, $p$ = .04. High frequency taggers ($m$ = 4.91, $sd$ = 1.46) also scored higher than no/low frequency taggers ($m$ = 4.13, $sd$ = 1.64) on total comprehension $t(137) = 2.46$, $p$ = .01. In contrast to highlighting, creating tags was not related to significant differences in subjects' factual comprehension $t(137) = 1.72$, $p$ = .086.

**Table 4.18 Multiple-Choice Scores by Tagging Frequency**

| | Tagging Frequency | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Factual | High-Frequency | 34 | 2.91 | 1.08 | .18 |
| | No/Low-Frequency | 105 | 2.53 | 1.11 | .10 |
| Inferential | High-Frequency | 34 | 2.00 | 1.04 | .17 |
| | No/Low-Frequency | 105 | 1.60 | .95 | .09 |
| Total | High-Frequency | 34 | 4.91 | 1.46 | .25 |
| | No/Low-Frequency | 105 | 4.13 | 1.64 | .16 |

**Table 4.19 Multiple-Choice Scores T-test for Tagging Frequency**

| | | Leven's Test for Equality of Variance | | t-test for Equality of Means | | | | |
|---|---|---|---|---|---|---|---|---|
| | Equal Variance? | F | Sif. | T | df | Sig (2-Tailed) | Mean Difference | Std. Error |
| Factual | Assumed | 1.54 | .216 | 1.72 | 137 | .08 | .37 | .21 |
| | Not Assumed | | | 1.75 | 57.53 | .08 | .37 | .21 |
| Inferential | Assumed | .00 | .99 | 2.07 | 137 | .04 | .40 | .19 |
| | Not Assumed | | | 1.98 | 52.14 | .05 | .40 | .20 |
| Total | Assumed | .71 | .40 | 2.46 | 137 | .01 | .77 | .31 |
| | Not Assumed | | | 2.61 | 62.06 | .01 | .77 | .29 |

In addition to the multiple-choice scores, analysis was also performed to compare the essay question results. ANOVA testing revealed that high-frequency taggers ($n$ = 34, $m$ = 2.13, $sd$ = 1.10) did not score higher than low-frequency taggers ($n$ = 40, $m$ = 1.82, $sd$ = 1.07) or non-taggers ($n$ = 65, $m$ = 2.25, $sd$ = 1.01) $f(2, 136) = 2.05$, $p = .133$. Since the researcher identified gender as a potentially confounding variable for essay

performance, ANCOVA was also conducted to control for the influence of gender. Results from ANCOVA testing also showed that the relationship between tagging and essay scores was not significantly. Interestingly, subjects who created no tags at all scored higher than low-frequency taggers. T-test analysis comparing the two groups showed that this result was significant $t(103) = 2.04$, $p = .043$, suggesting that tagging may actually be detrimental to comprehension when performed at low-frequencies.

**Table 4.20 Essay Score by Tagging Frequency**

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Lower Bound | 95% Confidence Upper Bound |
|---|---|---|---|---|---|---|
| No Tagging | 65 | 2.25 | 1.01 | .12 | 2.00 | 2.50 |
| Low-Frequency Tagging | 40 | 1.82 | 1.07 | .17 | 1.48 | 2.16 |
| High-Frequency Tagging | 34 | 2.13 | 1.10 | .19 | 1.74 | 2.51 |
| Total | 139 | 2.10 | 1.06 | .09 | 1.92 | 2.27 |

**Table 4.21 ANCOVA for Essay Score Controlling for Gender**

| Source | Type III Sum of Squares | df | Mean Square | F | Sig |
|---|---|---|---|---|---|
| Corrected Model | 11.51* | 2 | 5.75 | 5.47 | .00 |
| Intercept | 24.87 | 1 | 24.87 | 23.23 | .00 |
| Gender | 10.65 | 1 | 10.65 | 9.95 | .00 |
| Tags | 1.52 | 1 | 1.52 | 1.42 | .23 |
| Error | 145.57 | 136 | 1.07 | | |
| Total | 770.50 | 139 | | | |
| Corrected | 157.09 | 138 | | | |

*. R Squared = .07 (Adjusted R Squared = .06)

**Table 4.22 Essay Score T-Test for Low-Frequency and Non-Taggers**

| | | Levene's Test for Equality of Variance | | t-test for Equality of Means | | | | |
|---|---|---|---|---|---|---|---|---|
| | Equal Variance | F | Sig | t | df | Sig (2-Tailed) | Mean Difference | Std. Error |
| Essay Score | Assumed | .02 | .87 | 2.04 | 103 | .43 | .42 | .20 |
| | Not Assumed | | | 2.02 | 79.19 | .04 | .42 | .21 |

Tagging Confidence

While results regarding the negative relationship between low-frequency tagging and essay scores are initially perplexing, differences in the confidence level of high and low-frequency taggers may offer an explanation. Question seven of the questionnaire asked students to report their confidence in their annotation abilities on a nine-point Likert scale. Since subjects' mean annotation confidence was $m = 7.01$, $sd = 1.95$, this study labels subjects who reported an eight or a nine ($n = 73$) as their tagging confidence as "high confidence taggers" and subjects who reported a seven or lower ($n = 66$) as "low confidence taggers."

High confidence taggers created more than twice as many tags ($m$ =3.99, $sd$ 4.15) as low-confidence taggers ($m = 1.76$, $sd = 2.48$) $t(137) = -3.78$, $p < .000$. Likewise, table 4.24 Shows that high confidence taggers scored higher than low-confidence taggers on factual $t(137) = -2.85$, inferential $t(137) = -2.63$, and total comprehension $t(137) = -3.628$, $p < .00$. Essay scores were likewise significantly higher for high confidence taggers ($m$ = 2.35, $sd = 1.06$) than low-confidence taggers ($m = 1.81$, $sd = .99$) $t(137) = -3.05$, $p < .00$. Notably, tagging confidence was the only annotation related metric that effected the scores of all three areas of the multiple-choice assessment and the essay question.

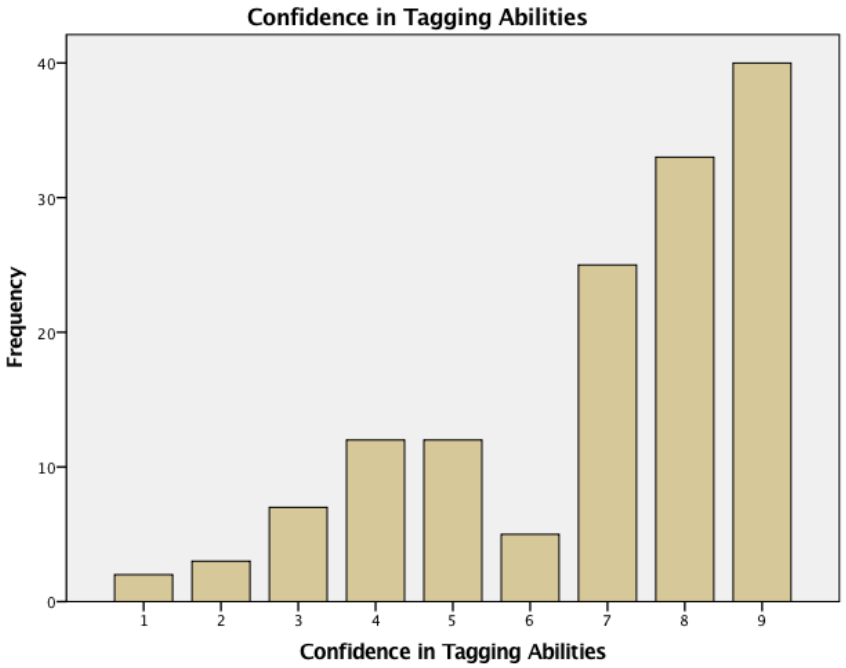


**Figure 4.3 Reported Confidence in Tagging Abilities**

**Table 4.23 Relationship of Tagging Confidence and Assessment Scores**

| | Tagging Confidence | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Factual | Low Confidence | 66 | 2.35 | 1.14 | .14 |
| | High-Confidence | 73 | 2.88 | 1.04 | .12 |
| Inferential | Low Confidence | 66 | 1.47 | .99 | .12 |
| | High-Confidence | 73 | 1.90 | .94 | .11 |
| Total | Low Confidence | 66 | 3.82 | 1.61 | .19 |
| | High-Confidence | 73 | 4.78 | 1.51 | .17 |
| Essay Score | Low Confidence | 66 | 1.81 | .99 | .12 |
| | High-Confidence | 73 | 2.35 | 1.06 | .12 |

**Table 4.24 Relationship of Tagging Confidence and Assessment Scores**

| | | Leven's Test for Equality of Variances | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Equal Variance? | F | Sig. | t | df | Sig (2-Tailed) | Mean Dif. | Std. Error Dif. |
| Factual | Assumed | 2.68 | .10 | -2.85 | 137 | .00 | -.52 | .18 |
| | Not Assumed | | | -2.83 | 131.95 | .00 | -.52 | .18 |
| Inferential | Assumed | 1.39 | .24 | -2.63 | 137 | .00 | .16 | .16 |
| | Not Assumed | | | -2.63 | 133.85 | .01 | .16 | .16 |
| Total | Assumed | .30 | .58 | -3.62 | 137 | .00 | .26 | .26 |
| | Not Assumed | | | -3.61 | 133.22 | .00 | .26 | .26 |
| Essay Score | Assumed | 1.63 | .20 | -3.05 | 137 | .00 | .17 | .17 |
| | Not Assumed | | | -3.06 | 136.84 | .00 | .17 | .17 |

Summary of Results for Question Two: The Relationship of Tagging and Comprehension

Several results regarding the frequency of tags created were significant. High-frequency taggers scored higher than no/low-frequency taggers on inferential and total comprehension portions of the multiple-choice assessment. In a complete flip of the highlighting results, tagging was not related to scores on the factual portion of the multiple-choice assessment.

For the essay question, neither high nor low-frequency tagging was related to higher participant scores. In contrast, low-frequency taggers actually performed poorer than subjects who created no tags at all. Since low-frequency taggers reporter lower confidence than high-frequency taggers, it may be the case that creating tags was a

particularly difficult task for this group. High confidence taggers outperformed low confidence taggers in every achievement metric.

**Question #3 – Which Method is Better for Improving Comprehension?**

Comparing Tagging and Highlighting

Results for questions one and two have soundly demonstrated that tagging and highlighting are each related to unique benefits. Highlighting had a positive relationship with higher essay scores and factual comprehension but was not related to higher inferential comprehension on the multiple-choice assessment. In contrast, tagging was not related to differences in factual comprehension and was negatively related to essay scores when done at a low-frequency but was positively related to subjects' inferential learning scores in the multiple-choice test. In this way, the answer to this question is largely dependent on the specific type of learning the reader or educator is trying to achieve.

Gender and Preferences

In addition to the desired learning outcomes, students' unique backgrounds may also play a role in resolving this question. For example, question five asked students which format they would prefer to use in class if they had both available. Results showed that 82.7% ($n$ = 115) of students preferred to use the digital version in class, with 3.6% ($n$ = 5) preferring the printed version, and the remainder ($n$ = 19) preferring either both or neither. Table 4.25 shows the breakdown of in-class format preferences for gender. Chi-square analysis determined that the difference between the genders was not significant $x^2(3) = 2.09$, $p = .55$.

**Table 4.25 Format Preferences by Gender**

| | | | Preferred Format at Seminary | | | | |
|---|---|---|---|---|---|---|---|
| | | | Gospel Library | Printed Scriptures | Both Together | No Preference | Total |
| Gender | M | Count | 67 | 2 | 3 | 7 | 79 |
| | | % in Gender | 84.8% | 2.5% | 3.8% | 8.9% | 100.0% |
| | | % in Preferred Format | 58.3% | 40.0% | 37.5% | 63.6% | 56.8% |
| | F | Count | 48 | 3 | 5 | 4 | 60 |
| | | % in Gender | 80.0% | 5.0% | 8.3% | 6.7% | 100.0% |
| | | % in Preferred Format | 41.7% | 60.0% | 62.5% | 36.4% | 43.2% |
| Total | | Count | 115 | 5 | 8 | 11 | 139 |
| | | % in Gender | 82.7% | 3.6% | 5.8% | 7.9% | 100.0% |
| | | % in Preferred Format | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

While there was no relationship between gender and format preferences in class, when students responded about their preferred format at home, differences appeared. Results showed that just 30% of female students ($n$ = 18) preferred to use the digital app at home, while the percentage of male students who preferred the format at home was nearly twice as high ($n$ = 43, 54.4%) $x^2(3) = 15.02$, $p < .00$. Females ranked the digital format as their third-most preferred format, behind "printed" and "both together."

**Table 4.26 Preferred Format at Home by Gender**

| | | | Preferred Format at Home | | | | |
|---|---|---|---|---|---|---|---|
| | | | Gospel Library | Printed Scriptures | Both Together | No Preference | Total |
| Gender | M | Count | 43 | 14 | 13 | 9 | 79 |
| | | % in Gender | 54.4% | 17.7% | 16.5% | 11.4% | 100.0% |
| | | % in Preferred Format | 70.5% | 42.4% | 38.2% | 81.8% | 56.8% |
| | F | Count | 18 | 19 | 21 | 2 | 60 |
| | | % in Gender | 30.0% | 31.7% | 35.0% | 3.3% | 100.0% |
| | | % in Preferred Format | 29.5% | 57.6% | 61.8% | 18.2% | 43.2% |
| Total | | Count | 61 | 33 | 34 | 11 | 139 |
| | | % in Gender | 43.9% | 23.7% | 24.5% | 7.9% | 100.0% |
| | | % in Preferred Format | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

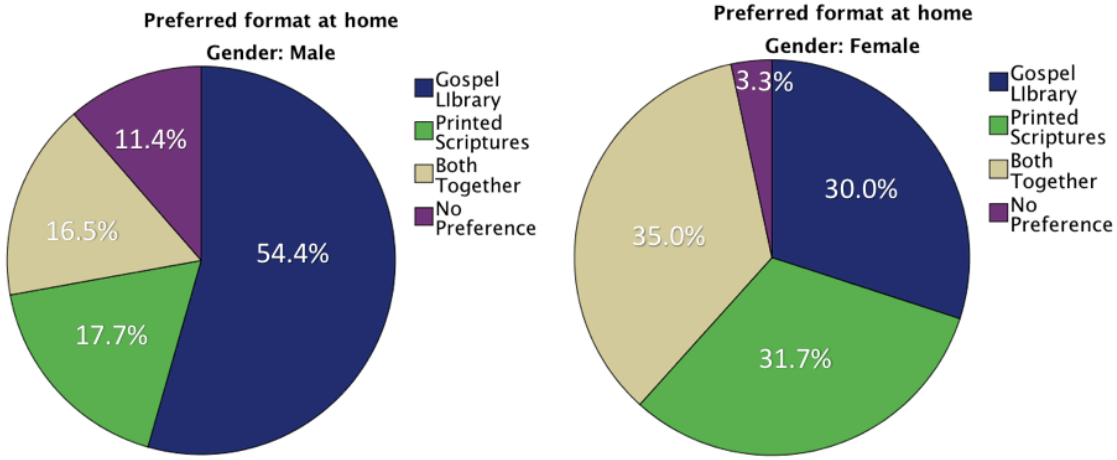


**Figure 4.4 Differences in Preferred Format at Home by Gender**

Similarly, males and females had significantly differing views on which format they felt their parents would prefer them to use. Male students were more than twice as likely (19.0%) as female students (8.3%) to report that their parents preferred that they use the digital version. Likewise, the proportion of female students who reported that their parents prefer that they use printed scriptures (31.7%) was 39% higher than the proportion of male students who reported the same thing (22.8%). Given the importance of readers using their preferred format to the success of digital annotation (Lauterman and Ackerman, 2014), the implications of the difference in preferences for males and females may be significant.

Perhaps most importantly, female students ($n = 60$) reported significantly higher annotation confidence ($m = 7.38$, $sd = 2.12$) than their male counterparts ($m = 6.72$, $sd = 2.11$) $t(137) = -1.99$, $p = .048$. Females also reported higher confidence scores for both tagging $t(137) = 2.05$, $p = .04$, and highlighting $t(137) = -3.47$, $p < .00$. That additional confidence is may be the reason that female students scored significantly higher ($m = 2.40$, $sd = 1.06$) than males students ($m = 1.86$, $sd = 1.01$) on the essay portion of the assessment $t(137) = -3.05$, $p < .00$.

**Table 4.27 Parental Preferences by Gender**

| | | | Preferred Format of Parents | | | | |
|---|---|---|---|---|---|---|---|
| | | | Gospel Library | Printed Scriptures | Both Together | No Preference | Total |
| Gender | M | Count | 15 | 18 | 14 | 32 | 79 |
| | | % in Gender | 19.0% | 22.8% | 17.7% | 40.5% | 100.0% |
| | | % in Preferred Format | 75.0% | 48.6% | 66.7% | 52.5% | 56.8% |
| | F | Count | 5 | 19 | 7 | 29 | 60 |
| | | % in Gender | 8.3% | 31.7% | 11.7% | 48.3% | 100.0% |
| | | % in Preferred Format | 25.0% | 51.4% | 33.3% | 47.5% | 43.2% |
| Total | | Count | 20 | 37 | 21 | 61 | 139 |
| | | % in Gender | 14.4% | 26.6% | 15.1% | 43.9% | 100.0% |
| | | % in Preferred Format | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

**Table 4.28 Annotation Confidence by Gender**

| | Gender | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Annotation Confidence | Male | 79 | 6.72 | 2.11 | .23 |
| | Female | 60 | 7.38 | 1.66 | .21 |
| Tagging Confidence | Male | 79 | 6.63 | 2.05 | .23 |
| | Female | 60 | 7.37 | 2.12 | .27 |
| Highlighting Confidence | Male | 79 | 7.63 | 1.65 | .18 |
| | Female | 60 | 8.47 | .98 | .12 |

**Table 4.29 Annotation Confidence by Gender T-test**

| | | Levene's Test for Equality of Variances | | t-test for equality of means | | | | |
|---|---|---|---|---|---|---|---|---|
| | Equal Variances? | F | Sig. | t | df | Sig. (2-Tailed) | Mean Dif. | Std. Error Dif. |
| Annotation Confidence | Assumed | 5.44 | .02 | -1.99 | 137 | .048 | -.66 | .33 |
| | Not Assumed | | | -2.06 | 136.7 | .041 | -.66 | .32 |
| Tagging Confidence | Assumed | .07 | .77 | -2.05 | 137 | .04 | -.73 | .35 |
| | Not Assumed | | | -2.04 | 124.86 | .04 | -.73 | .35 |
| Highlighting Confidence | Assumed | 16.81 | .00 | -3.43 | 137 | .00 | -.83 | .24 |
| | Not Assumed | | | -3.70 | 130.27 | .00 | -.83 | .22 |

The Relationship Between Grade Level and Comprehension

Grade level was also positively related to students' annotation confidence and comprehension scores. General annotation confidence scores were a full point higher for juniors ($m$ = 7.36, $sd$ = 1.70) than sophomores ($m$ = 6.35, $sd$ = 2.22), and a half point higher for seniors ($m$ = 7.84, $sd$ = 1.12) than juniors $f(2, 136) = 7.79$, $p < .00$. Significant differences also existed between grade levels for tagging and highlighting confidence. Predictably, older students also scored higher on both the multiple choice questions $f(2, 136) = 4.71$, $p = .01$ and essay question $f(2, 136) = 7.60$, $p < .00$.

**Table 4.30 Confidence in Annotation by Grade Level**

| | N | Mean | Std. Dev. | Std. Error | 95% Confidence Lower Bound | 95% Confidence Upper Bound | Min. | Max. |
|---|---|---|---|---|---|---|---|---|
| Sophomore | 63 | 6.35 | 2.22 | .28 | 5.79 | 6.91 | 2 | 9 |
| Junior | 45 | 7.36 | 1.70 | .25 | 6.84 | 7.87 | 1 | 9 |
| Senior | 31 | 7.84 | 1.12 | .20 | 7.42 | 8.25 | 5 | 9 |
| Total | 139 | 7.01 | 1.95 | .16 | 6.68 | 7.33 | 1 | 9 |

**Table 4.31 Multiple-Choice and Essay Scores by Grade Level**

| | N | Mean | Std. Dev. | Std. Error | 95% Confidence Lower Bound | 95% Confidence Upper Bound | Min. | Max. |
|---|---|---|---|---|---|---|---|---|
| Sophomore | 63 | 4.00 | 1.65 | .20 | 3.58 | 4.42 | 1 | 7 |
| Junior | 45 | 4.27 | 1.61 | .24 | 3.78 | 4.75 | 1 | 7 |
| Senior | 31 | 5.06 | 1.38 | .24 | 4.56 | 5.57 | 1 | 7 |
| Total | 139 | 4.32 | 1.63 | .13 | 4.05 | 4.60 | 1 | 7 |
| Sophomore | 63 | 1.75 | .94 | .11 | 1.51 | 1.99 | 0 | 4 |
| Junior | 45 | 2.25 | 1.05 | .15 | 1.93 | 2.57 | 1 | 4 |
| Senior | 31 | 2.58 | 1.11 | .19 | 2.17 | 2.98 | 1 | 4 |
| Total | 139 | 2.10 | 1.06 | .09 | 1.92 | 2.27 | 0 | 4 |

Summary of Results for Question #3 – Which Method is Best?

The results of the previous two questions have helped define the strengths and weaknesses of highlighting and tagging. Highlighting was positively related to higher scores on the factual portion of the multiple-choice assessment and the essay question.

Tagging was positively related to higher scores on the inferential portion of the multiple-choice assessment but was not related to higher scores for the essay question.

Beyond the features themselves, however, demographic factors, including gender, grade level, and training received, may play a significant role in how effective each tool will be. Chapter five will discuss how the relationship between confidence and assessmshould be considered when deciding which annotation features to focus on. For now, it will suffice to say that these factors appear to be related to readers' format preference and annotation confidence.

# CHAPTER FIVE – DISCUSSION

This chapter combines the findings from the literature discussed in chapter two with the results of this study to discuss the relationship between using smartphones to annotate texts and students' comprehension. The chapter is organized to address the main research question by addressing each of the three sub-questions introduced in chapter one:

1. What difference is there between students who create an above-average number of highlights, a below-average number of highlights, and students who do not highlight on comprehension?
2. What difference is there between students who create an above-average number of tags, a below-average number of tags, and students who do not tag on comprehension?
3. Which text annotation feature (highlighting or digital tagging) is more beneficial to readers' comprehension?

Each question is discussed by analyzing the results of this study within the context of the existing literature. The results from the multiple-choice assessments and essay questions in this study will be considered for each question. After the analysis, this study will outline several recommendations for educators seeking to utilize smartphones in their classrooms. The chapter will conclude with a brief discussion of the study's limitations and recommendations for further research.

## Question One – The Relationship Between Highlighting and Comprehension

Highlighting and Multiple-Choice Scores

Chapter two of this study reviewed the research of Ben-Yehuda and Eshet-Alkalai (2014) on the impact of digital and print annotation. In that study, the authors concluded that the benefits of annotating on paper were not transferable to digital annotation. This study demonstrates that such definitive statements may be oversimplified. Subjects who created ten or more highlights ($n$ = 49) scored 32% higher on factual comprehension ($m$ = 3.12, $sd$ = 2.927) than those who made one to nine highlights ($n$ = 54, $m$ = 2.35, $sd$ = 1.10) and those who made no highlights ($n$ = 36, $m$ = 2.36, $sd$ = 1.17). Importantly, the difference between those who made one to nine highlights and those who made no highlights was not significant. This finding suggests that the benefits of highlighting are only present for subjects who engage in relatively high-frequency highlighting. Also important, ANOVA testing to compare the means of high-frequency, low-frequency, and non-highlighting groups for inferential learning did not yield results that were significant $f(2,136) = 2.25$, $p = .108$.

These findings help connect what researchers have said about different types of learning with what researchers have said about different types of annotation. Chen et al. (2014) divided comprehension into two distinct categories: literal comprehension, which focuses on recalling stated facts, and inferential comprehension, which focuses on interpreting what the text means and how it is applicable. Likewise, Reid et al., (2017) proposed two types of annotation: non-generative annotation, which emphasizes what an author actually said, and generative annotation, which encourages the reader to insert their own interpretations of what a passage means into the text. The multiple-choice

portion of this study suggests that non-generative annotations like highlighting, while not beneficial for inferential comprehension, are effective for increasing literal comprehension.

This finding closely aligns with the body of research suggesting that highlighting is only effective in certain settings. Several studies have found that the primary benefit of highlighting text is that the brain prioritizes the text as important (Blanchard & Mikkelson, 1987; Dunlosky. Rawson, Marsh, Nathan & Willingham, 2013; Johnson, 1988). Unfortunately, this benefit comes at a cost. While readers benefit from an increased ability to remember highlighted passages, they are simultaneously less likely to recall passages that they choose not to highlight (Dunlosky et al., 2013; Fowler & Barker, 1974). Thus, for highlighting to be an effective method, the reader must have the ability to identify which parts of a text are more or less important. Likewise, since the primary benefit of highlighting is prioritizing important information, the method is best used when reading complex texts (Fass & Schumacher, 1978). This study involved readers that were trained to identify the most useful content in a text that was inherently difficult. In doing so, it supports the finding that highlighting is an effective tool for increasing comprehension when it is used in the proper settings.

Highlighting and Essay Scores

Despite findings from the multiple-choice assessment, the essay assessment results suggest that highlighting may be effective at generating inferential learning in certain settings. High-frequency highlighting was positively related to higher essay scores. Students who created ten or more highlights ($n$ = 49) scored 20.5% higher ($m$ = 2.48, $sd$ = 1.08) than those who created one to nine highlights ($n$ = 54, $m$ = 2.00, $sd$ =

1.11), and 32.4% higher than those who created no highlights ($n$ = 36, $m$ = 1.81, $sd$ = .879). Similarly, 24.5% ($n$ = 12) of subjects who created ten or more highlights received a level-four score, while just 5.6% ($n$ = 2) of those who made no highlights received the same score.

Since the essay question asked students to apply the text to a real-life problem, the positive relationship between highlighting and essay scores appears to contradict the multiple-choice assessment findings regarding the relationship between highlighting and inferential learning. One possible explanation for the discrepancy is that subjects received the essay question before reading the text but did not receive the multiple-choice questions until after completing the reading.

Lauterman and Ackerman (2014) found that assigning readers before they began reading to write down four keywords that summarized the text significantly increased their assessment scores. Likewise, Sidi et al. (2016) found that digital readers need an "external trigger" (p. 6) to help stimulate deeper comprehension of a text. In this study, the essay question served as the trigger, focusing readers' minds on extracting a single application, and highlighting contributed to the effectiveness of this process by allowing readers to emphasize and organize facts that were applicable to their essay. While this process was beneficial in deepening readers' responses to the question they received before they began reading, it did not increase comprehension for the questions readers received after they completed the reading. In essence, highlighting was ineffective at increasing readers' ability to flexibly apply what they read to a variety of settings. This suggests that highlighting is an effective way of deepening readers' comprehension but is not effective at broadening their ability to apply the text.

It may also be the case that creating highlights helped readers obtain a level of spatial fixity that is often missing in digital reading. Mangen et al. (2013) explained that the need to scroll while reading digital texts creates challenges that are not present when reading printed texts. Since printed texts have clear navigational markers like page and column breaks, readers can easily gain a clear spatial mental representation of the text (Jabr, 2013; Liu, 2005). While scrolling eliminates those navigational cues, it may be that highlighting serves as a suitable replacement. Since readers often use highlighting to emphasize major themes (Qayyum, 2008), highlighting may increase comprehension by creating clear navigational markers that facilitate a stronger spatial mental representation of the text.

Highlighting Confidence and Assessment Scores

While highlighting was beneficial to both factual multiple-choice scores and essay scores, readers' confidence in their highlighting abilities was an important factor. Participants in the study rated their confidence on a scale of one to nine. Since the mean annotation confidence score was $m = 7.01$, $sd = 1.95$, subjects who reported highlighting confidence of eight or higher ($n = 105$) were considered "high confidence," while those who rated themselves seven or lower ($n = 34$) were considered "low confidence." When scores for the essay portion of the assessment were analyzed, those who had high confidence in their highlighting abilities scored 26.7% higher ($m = 2.209$, $sd = 1.048$) than those who reported low confidence ($m = 1.764$, $sd = 1.067$). In contrast, participants' highlighting confidence was not related to higher or lower multiple-choice scores.

Results from a myriad of studies have emphasized the importance of students receiving quality training, modeling, and practice opportunities (Azevedo & Cromley,

2004; Chen & Chen, 2014; Dobler, 2015; Lauterman & Ackerman, 2014; Nichols, 2018; Rockinsaw-Szapkiw et al. 2011; Van Horne et al., 2016). This study supports those findings by demonstrating a positive relationship between readers' confidence in their annotation skills and their ability to identify meaningful applications of a text.

Summary of the Relationship Between Highlighting and Comprehension

While this study found no relationship between highlighting and creating cognitively flexible knowledge, results from this study demonstrate that highlighting is positively related to higher comprehension in some ways. Most significantly, the multiple-choice assessment results showed that subjects who made ten or more highlights scored 32% higher on factual comprehension than those who made less than ten or no highlights. This increase in factual comprehension is likely due to highlighting's non-generative nature, which emphasizes portions of what the author said rather than focusing on how to apply the content in different settings.

In addition to the relationship between highlighting and factual comprehension, highlighting may also have some benefits on inferential learning. While highlighting was not related to higher inferential learning scores on the multiple-choice assessment, it was related to higher-level essay scores. Students who created ten or more highlights scored 32% higher than those who created none. Since participants received the essay question before they began reading, it may be that the essay question elicited inferential thinking, while highlighting helped readers emphasize and organize passages that supported their idea. While this approach was beneficial when subjects received a prompt before reading, subjects did not perform better on questions they received after reading the text. This difference suggests that highlighting may be an effective method of deepening

understanding of a specific topic but an ineffective method of helping readers understand the content in a way that can be more broadly applied.

Finally, confidence in highlighting was positively related to higher essay scores. Subjects with high confidence in their highlighting abilities scored 27% higher than those with low confidence. This finding suggests that methods like instructor modeling and assigned practice that have been shown to increase readers' confidence are a prudent use of educators' time.

**Question Two - The Relationship Between Tagging and Comprehension**

Tagging Frequency and Multiple-Choice Scores

As with highlighting frequency, this study divided subjects into three groups based on the number of tags they created. Subjects who created no tags were labeled “non-taggers.” The number of tags created by the remaining subjects was used to generate a mean, and those who were above the mean ($n$ = 34) were deemed "high-frequency taggers." Those who created less than the mean number of tags ($n$ = 40) were deemed "low-frequency taggers," Results showed that high-frequency taggers scored 15.01% higher ($m$ = 2.91, $sd$ = 1.08) than low-frequency and non-taggers ($m$ = 2.53, $sd$ = 1.11) on the inferential portion of the multiple-choice test. Likewise, high-frequency taggers scored 18.88% higher ($m$ = 4.91, $sd$ = 1.46) than low frequency and non-taggers ($m$ = 4.13, $sd$ =. 1.64) on total comprehension multiple-choice scores. Interestingly, while the relationship between highlighting was significant for factual but not inferential learning, the relationship between factual and inferential comprehension for tagging was exactly the opposite. While tagging was related to higher inferential learning, tagging’s relationship with factual learning was not significant ($p$ = .08).

While high-frequency taggers scored significantly higher than low-frequency and non-taggers in some areas, the relationship did not begin until they created at least six tags (high-frequency). In contrast, subjects who created one to five tags ($n = 40$) scored lower than subjects who created no tags ($n = 65$) on all three measures of the multiple-choice assessment and the essay question (though the differences were only significant for the essay question).

This threshold presents an interesting situation. While high-frequency tagging appears to be beneficial to learning, creating a small number of tags may be detrimental. The most logical explanation for this phenomenon is the impact of tagging on readers' cognitive load (Miller, 1956; Sweller, 1988). While the creation of tags increases any users' cognitive load, creating tags requires more effort for readers who are less confident in their tagging abilities. In this way, those who are less confident are doubly cursed. Their lower confidence leads to less effective tagging than the high-frequency taggers and higher cognitive load than those who create no tags at all.

Data on users' self-reported tagging confidence supports this idea. As with highlighting, subjects who reported a score of eight or nine on the nine-point confidence scale ($n = 73$) were considered "high-confidence," while subjects who reported a one to seven confidence score ($n = 66$) were considered "low-confidence." Participants who reported low confidence created less than half as many tags ($m = 1.76$, $sd = 2.48$) as those who reported high confidence ($n = 73$, $m = 3.99$, $sd = 4.15$). Likewise, high confidence taggers scored higher on both the factual and inferential multiple-choice questions, as well as the essay portion of the assessment.

This finding suggests that low-confidence taggers likely experienced a disproportionately high cognitive load increase that left them at a disadvantage to both those who were comfortable tagging and those who chose not to tag. This finding should serve as a cautionary tale to teachers seeking to implement a standard annotation system for their students. While tagging has the potential to increase comprehension, if teachers are not careful to ensure that students are confident in their skills, asking them to tag may backfire.

In contrast, for high-frequency taggers, tagging may have served as what Bjork (1994; 1999) referred to as a "desirable difficulty" (Shapiro, 1988; Sidi et al., 2016). Since creating tags required readers to slow down and consciously identify more broadly applicable themes, readers appear to have processed and stored the information in a way that allowed them to obtain cognitive flexibility. This process likely led to an increase in their cognitive load, but high-frequency tagging appears to be the sweet spot where tagging's benefits outweigh its drawbacks. The advantage of creating a higher number of tags is likely twofold. First, with each tag a reader creates, they expand their ability to retrieve and apply the content they are reading in a future setting. Second, creating a higher number of tags helps the reader refine their skills. As the reader becomes more efficient at tagging, they benefit from the process, exerting a smaller burden on their cognitive load.

Regardless of which of the above benefits (decreased cognitive load or increased cognitive flexibility) is more significant, high-frequency tagging led to higher inferential comprehension. This finding, when paired with other studies, suggests that tagging effectively increases both comprehension and cognitive flexibility, but only when readers

are confident in their abilities and committed to regularly creating tags. For example, Lauterman and Ackerman (2014) found that digital reading's disadvantages could be overcome by increasing the subject's familiarity with the format. Likewise, Azevedo and Cromley (2004) found that providing digital readers with a 45-minute-training on digital reading strategies before asking them to read a text led to significantly higher scores on post-read comprehension tests. Chen and Chen (2014) also found that subjects overcame digital reading's disadvantages when they received training on how to annotate. As with those studies, this study suggests that training and familiarity are essential elements to overcoming the challenges associated with digital reading.

The Relationship Between Tagging and Essay Question Scores

Given the relationship between tagging and inferential learning during the multiple-choice assessment, it was anticipated that tagging would be related to higher essay scores as well. Surprisingly, this was not the case. Instead, low-frequency taggers ($n$ = 40) scored significantly lower ($m$ = 1.82, $sd$ = 1.07) than non-taggers ($n$ = 65, $m$ = 2.25, $sd$ = 1.01), while the difference between non-taggers and high-frequency taggers ($n$ = 34, $m$ = 2.13, $sd$ = 1.10) was not statistically significant.

Initially, this data appears to suggest that tagging does not positively impact inferential learning. However, given the multiple-choice assessment results, it is likely more accurate to say that tagging is only positively related to some aspects of inferential learning. Specifically, tagging strengthens readers' ability to apply a text in multiple situations but does not necessarily increase their understanding of how a text applies to a specific application.

As with the results for highlighting, tagging's relationship with essay scores was likely heavily influenced by readers receiving the essay question before they began reading. This order of procedures allowed the readers to hyper-focus on identifying a single application of the text. In some ways, this replaced the need for readers to use tagging to help identify unique applications, but it did come with at least one significant drawback. While the approach was effective at deepening readers' understanding of the text in a single context, it likely did so at the expense of seeing how the text applied more broadly. As a result, having the essay question led to higher essay scores but failed to produce knowledge that readers could flexibly apply during the multiple-choice assessment. In this regard, this study shows the unique role of tagging in creating cognitive flexibility. While annotation methods like highlighting and marginalia help focus a reader on how a passage applies in a specific setting, tagging helps readers obtain cognitive flexibility by encouraging them to step back and examine how the text applies more broadly to a variety of settings.

Similarly, having the essay question beforehand explains why those who created no tags outperformed low-frequency taggers. Though both groups knew to look for a specific application of the text, low-frequency taggers divided their focus between two very different methods. The essay question required readers to focus on a single specific application, while tagging required readers to focus on identifying multiple contexts in which they could apply the text. High-frequency taggers also divided their attention between two tasks; however, they reported significantly higher confidence in their tagging abilities ($m$ = 8.24, $sd$ = 1.55) than low-frequency taggers ($m$ = 6.33, $sd$ = 1.92). This increased confidence likely allowed the high-frequency taggers to create tags

without significantly increasing their cognitive load, while the same task significantly burdened the less confident low-frequency group. Regardless, the results support the large number of studies that suggest that increasing readers' confidence in their skills through training, modeling, and practice is critical to recognizing the benefits of tagging (Azevedo & Cromley, 2004; Chen & Chen, 2014; Dobler, 2015; Lauterman & Ackerman, 2014; Nichols, 2018; Rockinsaw-Szapkiw et al. 2011; Van Horne et al., 2016).

Summary of the Relationship Between Tagging and Comprehension

Combining the results of the multiple-choice and essay assessments provides a clearer vision of the usefulness of tagging. High-frequency taggers scored 15% higher on the inferential portion of the multiple-choice assessment but did not score higher on that assessment's factual portion. This finding suggests that tagging's primary benefit is helping users create knowledge they can flexibly apply in the future.

In contrast to the multiple-choice assessment, tagging was not positively related to subjects' essay scores. Providing subjects with the essay question before they began reading may have significantly impacted this outcome by focusing readers' attention on identifying a single application. Since readers already had a specific question to focus on, tagging may have distracted readers by encouraging them to identify multiple applications of the text. This difference in focus explains why high-frequency taggers saw significant benefits when asked questions after completing their reading but no benefit on the essay question they received before they began reading. In essence, tagging helps users broaden their vision of how a text is applicable in a variety of settings but is not effective at deepening their understanding of how it applies to a specific setting.

Finally, it is important to acknowledge that there was no relationship between low-frequency tagging and higher multiple-choice scores, and low-frequency tagging was related to lower essay scores. Since low-frequency taggers reported lower confidence in their tagging skills than high-frequency taggers, creating tags was likely particularly burdensome on this group's cognitive load. This finding should serve as a significant warning for educators who encourage students to create tags as they read. While tagging's benefits can be significant when subjects are comfortable and well trained, if those stipulations are not met, tagging may do more harm to readers' learning than good.

**Question Three – Which Annotation Feature is Most Beneficial to Comprehension?**

Understanding the differences in Annotation Methods

Question three is likely the most important of the three research sub-questions, but it is also the most difficult to answer. In reality, comparing different annotation tools is a lot like comparing a hammer and a wrench. Labeling a single tool as "best" is inappropriate since each tool is designed to perform a different task and excels in entirely different areas. With annotation, as with carpentry, it is better to understand which tool is best for a given situation.

One key to understanding when to use each annotation tool is recognizing the differences between non-generative and generative annotation (Reid et al., 2017). Highlighting is a non-generative annotation. When a reader highlights, they create no new content. Instead, the reader emphasizes the facts that the author presents in the text. Given the emphasis on facts, it should come as no surprise that, in this study, readers who highlighted scored higher on the factual portion of the multiple-choice test (Spiro 2012,

p. 51). In this regard, highlighting appears to be the superior tool for identifying factual elements of a text.

Similarly, highlighting seems particularly helpful when the reader has a clearly defined outcome. In this study, readers received the essay question before they began reading. As a result, they were better able to use highlighting as an effective way of identifying and organizing facts that supported their answer.

In contrast, generative annotation goes beyond simply emphasizing what the author said. With generative annotation, the reader processes the information, identifies unstated principles or applications, and inserts them into the text (Reid et al., 2017). Since the entire process focuses on identifying what the author did not say in the text, it should come as no surprise that those who were high-frequency taggers scored higher on the inferential portion of the multiple-choice assessment.

Significantly, highlighting was not related to subjects' inferential comprehension scores, and tagging was not related to their factual comprehension scores. This finding suggests that readers should deliberately use each method to accomplish a specific outcome rather than as a catch-all for improving learning. One practical tip is to label each potential read as either reflective or imaginary (Hillesund, 2010). Reflective reading focuses on interpreting texts by forming connections between what is read and what the reader already knows. The inferential benefits of tagging make it a natural fit for reflective reading.

While reflective reading encourages the reader to form connections outside of the primary text, imaginary reading is typically more self-contained. This reading style often emphasizes a story-line and encourages readers to recognize the facts laid out by the

author in the text. Since highlighting is designed to emphasize things that are explicitly stated in the text, highlighting is likely a better fit for this style of reading.

By identifying the style of reading they seek to utilize, educators can encourage their students to utilize an annotation skill that aligns with their goal. For example, a math teacher might encourage students to highlight key formulas to identify an object's area. Since highlighting helps students recall factual content, this would be an effective way of helping them learn the formulas. However, suppose the teacher was more interested in emphasizing how to use the formulas. In that case, they might encourage students to tag each formula with three real-life objects they could measure using the formula. Since tagging is designed to emphasize inferential learning, the student may not be able to recall the details of the formula but will be more likely to be able to apply it in future settings. In this case, both learning goals are appropriate, but the assigned annotation method will likely lead to very different outcomes.

### Understanding Individual Teachers and Students

Beyond knowing when to utilize each annotation method, each student's and teacher's desires should be considered. Several studies suggest that students' format preferences and confidence with digital reading and annotation significantly impact learning outcomes (Kong et al., 2018; Lauterman & Ackerman, 2014; Nichols, 2016). This study supports those findings.

For example, students' grade-level was related to both their annotation confidence and their multiple-choice and essay scores. In each case, students' confidence and assessment scores were significantly higher for older students. There are at least two interpretations for this finding. Since twelfth-grade students had attended two more years

of school, it may be that their general comprehension capacity was simply higher than the eleventh and tenth-grade students. Alternatively, it may be that the period required to be comfortable with a digital format is significantly longer than traditionally assumed. Since most twelfth-grade students had already attended three years of seminary before the study, their experience with the reading app was significantly greater than tenth-grade students, who typically had just one year of experience. If this is the case, educators should be aware that, while digital annotation may have significant comprehension benefits, the process of training students is a long-term investment, and the full benefits may take years to achieve (Van Horne et al., 2016).

Regardless of the reason, younger students having lower confidence means that annotating leads to a heavier cognitive load for them than it does for older students. Perhaps because of the familiarity of highlighting printed texts (Kong et al., 2018), students reported significantly higher confidence in their highlighting abilities than their tagging abilities. Thus, to minimize the increase in students' cognitive load, it may be wise to focus on highlighting when introducing younger or less experienced students to digital annotation.

While some settings will likely benefit from the simplicity of highlighting, at least one group appears better suited for a robust annotation system. Female students ($n$ = 60) reported significantly higher annotation confidence ($m$ = 7.38, $sd$= 1.66) than male students ($n$ = 79, $m$ = 6.72, $sd$ = 2.11). Since the cognitive load burden associated with annotation is lower for students with high confidence, female students are likely in a better position to benefit from advanced annotation methods like tagging.

In each of the cases above, the determining factor for which annotation method is best has more to do with the individual student's or teacher's desires and abilities than the method's inherent nature. As such, educators should focus less on whether particular annotation methods are "good" and focus more on whether or not those methods are a good fit for their students.

Summary of Which Annotation is Most Beneficial to Comprehension?

This section has demonstrated that various factors must be considered when determining which annotation method is best. The generative and non-generative nature of tagging and highlighting contribute in very different ways to readers' learning. This study shows that highlighting excels when readers are hoping to retain explicitly stated facts from a text. Highlighting also excels when readers begin their reading with a specific question in mind. In contrast, utilizing tagging is best when readers seek to understand general principles they can flexibly apply to a variety of settings. Given the differences, when determining which approach to use while reading, teachers and readers should ask themselves whether the goal of their reading is factual retention or flexible application.

While each tool has inherent strengths and weaknesses, each reader's preferences and capabilities should also be considered when determining which annotation method to use. It may be wise to focus on highlighting when working with younger and less experienced groups. Since highlighting confidence is generally higher, this will allow them to begin annotating without experiencing an unreasonably high cognitive load increase. Similarly, since in-depth training significantly impacts readers' tagging confidence, teachers who only plan on using annotation sparingly might be better off to

forego training on tagging and focus instead on the easier to master skill of highlighting. In contrast, older and more experienced readers' high confidence makes them ideal candidates for digital tagging. Gender was also a predictor for confidence, with females reporting significantly higher annotation confidence than their male counterparts. Given the significant impact each user's unique background plays in determining each annotation method's effectiveness, teachers should avoid overgeneralizing any annotation skill's effectiveness.

## Recommendations

### Modeling, training, and practice

For educators seeking to implement a digital annotation program, this study, combined with the existing body of research, lends itself to several recommendations. First, it is critical to ensure that readers are confident in their annotation skills. In this study, the relationship between comprehension and annotation confidence was stronger than any other relationship, including annotations created.

Research on digital annotation suggests three essential elements of building students' confidence: Modeling effective usage, training, and providing practice opportunities (Azevedo & Cromley, 2004; Chen & Chen, 2014; Dobler, 2015; Lauterman & Ackerman, 2014; Nichols, 2018; Rockinsaw-Szapkiw et al., 2011; Van Horne et al. 2016). Teachers can model effective annotation while projecting their own digital text in front of the class or encouraging students to share their annotations with a nearby student. Johnson et al. (2010) found that students use modeling examples as scaffolding to annotate more effectively.

Training also helps build students' effectiveness. Importantly, however, training students on a single occasion is typically insufficient (VanHorne et al., 2016). Further, teachers should avoid the temptation to focus their training entirely on the technical aspects of how to create an annotation and ensure that students receive adequate training on identifying what to annotate (Azevedo & Cromley, 2004). For example, in this study, participants watched two training videos on tagging (see appendix G). The first training video focused on identifying and tagging principles they saw in the text. The second video focused on identifying potential applications of the text. While both videos also explained how to select the "tag" button and type in a keyword, the training focused more on knowing what to tag than how to tag. Several studies suggest that this process is critical to students' willingness to adopt and effectively use annotation methods (Azevedo & Cromley, 2010; Van Horne et al., 2016).

Finally, educators should provide students with specific opportunities to practice annotating. In this study, subjects received specific invitations to annotate in each of the eleven class periods leading up to the study. Several studies suggest that the increased familiarity that comes from practice is critical for building students' confidence (Chen & Chen, 2014; Lauterman & Ackerman, 2014; Nichols, 2018). As with the training, practice opportunities should assign students a specific task. Examples include "highlight the contextual elements, including the who, what, where, when, and why of the story," or "highlight what you think is the most important word in each passage." Providing students with specific assignments allows them to discover which methods work best for them while also reducing the cognitive load associated with determining when to create

an annotation (Lauterman & Ackerman, 2014; Rockinsaw-Szapkie et al., 2011; Van Horne et al. 2016).

Understanding the value of each tool

Each time a teacher utilizes annotation in their classroom, it is important that they begin with a specific outcome in mind. While, in general, annotation can be beneficial to overall learning, this study has demonstrated that the relationship between each annotation tool and students' comprehension is different. As such, teachers must understand what their desired outcome is and how each annotation method can help them achieve it (Reid et al., 2017).

For example, this study has shown that if a teacher seeks to increase their students' factual retention, they are best suited to encourage highlighting. Likewise, if a teacher wants students to understand how a text applies to a specific question or case study, highlighting would likely be the most effective approach. If, however, the teacher is looking to help students to be able to flexibly apply what they are learning to a variety of settings, tagging is likely the best option. While this study explored just two annotation options, it is likely that methods like creating links (Ackerman & Leiser, 2014; Anotneko & Niederhauser, 2010; Azevedo & Cromley, 2004; Shapiro, 1988), defining words (Greenlee-Moore & Smith, 1996; Stoop et al., 2013), creating notes in the margins (Bold & Wagstaff, 2017; Mueller & Openheimer, 2014; Rockinsaw-Szapkiw et al. 2013; Schugar et al. 2011), and many others, also have unique benefits to learning. Just as a carpenter uses different tools for different problems, teachers should learn to use different annotation methods to achieve different outcomes.

Understand Individual Students

Finally, teachers who are encouraging annotation must understand their students' skill levels and preferences. Lauterman and Ackerman (2014) found that digital annotation positively impacted students' comprehension, but only when students reported that their preferred reading format was digital. Similarly, this study suggests that age, gender, and experience may play a significant factor in students' annotation confidence, format preferences, and effectiveness when annotating. Given the importance of student skills and preferences, teachers should be careful to ensure that digital annotation, and digital reading in general, is the right fit for their students.

Fortunately, it appears that students' preferences can be influenced by effective training, modeling, and practice. Dobler (2010) provided these things to his students over the course of a semester and found that student preferences for digital reading increased from 22% to 50%. Similarly, subjects in this study received modeling, training, and practice opportunities for the eleven class sessions leading up to the study. When asked about their format preference, 82.7% of subjects reported that they preferred utilizing the digital reading app in class. However, that number was significantly higher than reported preferences in other studies (Baron, 2017; Dobler, 2015; Nichols, 2018; Noyes & Garland, 2005; Two Sides, 2015). As such, while it is possible to influence student preferences, teachers should carefully consider their students' desires and preferences to determine if encouraging digital annotation is right for their classrooms.

## Limitations and Further Research Suggestions

While the findings of this study are transferable to a variety of settings, it is worth recognizing the specific settings in which the study was conducted. Subjects in the study

were very familiar with the digital reading app and had received significant training, modeling, and practice annotating digital texts. Even with these advantages, second-year students performed far poorer than fourth-year students. This discrepancy suggests that obtaining the level of familiarity necessary to be proficient should be considered a long-term undertaking. Educators should not expect to see similar results immediately when introducing annotation to their students.

While this study did not ask students about their prior experience with digital reading, of the 139 participants in this study, just five students reported that they preferred to use the printed text in class. While this overwhelming preference was beneficial for demonstrating the potential of digital annotation, preferences for the digital format were dramatically more common in this study than what has been seen in other studies (Baron, 2017; Dobler, 2015; Nichols, 2018; Noyes & Garland, 2005; Two Sides, 2015). Given the importance of preferred format on annotation effectiveness (Lauterman & Ackerman, 2014), educators should be careful to gauge their students' preferences before beginning.

In addition to the benefits of familiarity, the location for this study was carefully selected because of the text used in the course. Since the study took place at a youth seminary, the study utilized a religious text. Religious texts are typically both complicated and broadly applicable, making them particularly well suited to both highlighting and tagging. Simpler texts or texts designed to teach the application of a principle in a specific way (like an instruction manual) may not see the same results.

While the complex nature of the text was appropriate for the study, it should be noted that some participants had likely read the text prior to the study. Similarly, since

subjects were using their own personal text, some subjects may have created tags or highlights prior to the study beginning. While it is unknown what impact this would have on the data, it is important to note that some students may have been more familiar with the text than others.

One additional limitation comes from the researcher asking students to use their personal devices. Since students used their own devices, it was impossible to obtain a sample with each student utilizing the app on the same operating system. While there are no apparent differences between the IOS and Android versions of the app, since subjects were not asked about which operating system they used it is possible that differences in students devices may have had an impact.

More importantly, however, this study's scope touches only a small portion of the research necessary to understand digital annotation. For example, in this study, generative and non-generative annotations are represented entirely by tagging and highlighting. While these are excellent examples of each form of annotation, additional research is needed to explore the specific impact and role of methods like creating links, defining words, and creating marginal notes.

Likewise, this study focused entirely on the short-term benefits of each annotation feature. Subjects were tested on their understanding immediately after they completed the reading. Additional research is needed to determine what benefits digital annotation has on long-term comprehension and recall.

Finally, perhaps the most important question for educators is whether the benefits of using smartphones to annotate digital texts are worth the potential distractions. Several studies suggest that students inherently view their phones as distractions (Lin et al., 2013;

Noyes & Garland, 2005; Przybylski & Weinstein, 2013; Thornton et al., 2014). It remains to be seen if teaching students to use their smartphones as educational tools can impact that perception and reduce distractions in the classroom.

## Conclusion

This study set out to determine the relationship between using smartphones to annotate digital texts and students' comprehension. After reviewing the relevant literature and conducting original research, this study finds that there is a relationship between digital annotation and students' comprehension. Methods like highlighting offer similar benefits in digital reading as they do in print. Specifically, they excel at helping readers retain and organize information from a text. In contrast, digital formats offer readers several unique opportunities to interact with the text through annotation. This study suggests that digital tagging facilitated an entirely different relationship with comprehension than highlighting and may help readers to obtain knowledge that can be more flexibly applied to multiple settings.

Just as Guttenberg's innovation to the printing press opened the door to new ways of learning, digital reading is dramatically altering the way we access and read materials. With smartphone ownership becoming a standard part of modern life, learning to optimize the smartphone reading experience is becoming increasingly valuable. What remains to be seen is whether or not modern innovators can mainstream effective usage of digital texts. If they can, it may be the case that the advent of digital reading will one day be seen as an equally pivotal point in the quest to distribute knowledge as Guttenberg's press.

# REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433-459.

Ackerman, R., & Leiser, D. (2014). The effect of concrete supplements on metacognitive regulation during learning and open-book test taking. *British Journal of Educational Psychology*, *84*(2), 329-348.

Al-Sarmi, A. M., & Al-Hemyari, Z. A. (2014). Statistical characteristics of performance indicators. *International Journal of Quality and Innovation*, *2*(3-4), 385-309.

Andrews, E. (2012, Dec 18). 11 Innovations that changed history. History. https://www.history.com/news/11-innovations-that-changed-history

Andrews, E. (2015, Feb 23). 7 Things you may not know about the Gutenberg Bible. History. https://www.history.com/news/7-things-you-may-not-know-about-the-gutenberg-bible

Antón, C., Camarero, C., & Rodríguez, J. (2017). Pleasure in the use of new technologies: the case of e-book readers. *Online Information Review*, *41*(2), 219-234.

Antonenko, P. D., & Niederhauser, D. S. (2010). The influence of leads on cognitive load and learning in a hypertext environment. *Computers in Human Behavior*, *26*(2), 140-150.

Ashton, B. K. (2018). Teaching with tech: Engaging youth in a digital world. Church of Jesus Christ. https://www.churchofjesuschrist.org/study/ensign/2018/01/teaching-with-tech-engaging-youth-in-a-digital-world?lang=eng

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia?. *Journal of Educational psychology*, 96(3), 523.

Bae, S. M. (2015). The relationships between perceived parenting style, learning motivation, friendship satisfaction, and the addictive use of smartphones with elementary school students of South Korea: Using multivariate latent growth modeling. *School Psychology International, 36*(5), 513-531.

Baron, N. S. (2017). Reading in a digital age. *Phi Delta Kappan, 99*(2), 15-20.

Benedetto, S., Carbone, A., Drai-Zerbib, V., Pedrotti, M., & Baccino, T. (2014). Effects of luminance and illuminance on visual fatigue and arousal during digital reading. *Computers in Human Behavior, 41*, 112-119.

Ben-Yehudah, G., & Eshet-Alkalai, Y. (2014, February). The influence of text annotation tools on print and digital reading comprehension. In *Proceedings of the 9th Chais Conference for Innovation in Learning Technologies* (p. 28-35), n.p., n.d..

Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology*, *39*(5), 775-786.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowin*g (pp. 185–205). MIT Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application (pp. 435–459). MIT Press.

Blackwell, L., Gardiner, E., & Schoenebeck, S. (2016, February). Managing expectations: Technology tensions among parents and teens. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (pp. 1390-1401). ACM.

Blanchard, J., & Mikkelson, V. (1987). Underlining performance outcomes in expository text. *The Journal of Educational Research*, *80*(4), 197-201.

Bold, M. R., & Wagstaff, K. L. (2017). Marginalia in the digital age: Are digital reading devices meeting the needs of today's readers? *Library & Information Science Research*, *39*(1), 16-22.

Bowman, L. L., Levine, L. E., Waite, B. M., & Gendron, M. (2010). Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education*, *54*(4), 927-931.

Chen, C. M., & Chen, F. Y. (2014). Enhancing digital reading performance with a collaborative reading annotation system. *Computers & Education*, *77*, 67-81

Chen, G., Cheng, W., Chang, T. W., Zheng, X., & Huang, R. (2014). A comparison of reading comprehension across paper, computer screens, and tablets: Does tablet familiarity matter?. *Journal of Computers in Education*, *1*(2-3), 213-225.

Chung, S. T. (2004). Reading speed benefits from increased vertical word spacing in normal peripheral vision. *Optometry and Vision Science*, *81*(7), 525.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

DeLamater, W. E. (2010, April 29). How larger font size impacts reading and the implications for educational use of digital text readers. http://www.ereadia.com/research/How_Larger_Font_Size_Affects_Reading.pdf

Dobler, E. (2015). e-Textbooks. *Journal of Adolescent & Adult Literacy*, *58*(6), 482-491.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

Durham, P., & Raymond, R. D. (2016). Building cognitive reading fluency through "tagging" for metacognition. *Texas Journal of Literacy Education*, *4*(1), 46-56.

Eastern Washington University Writer's Center, (2018). What is annotating, and why do it?. Reading and study strategies. https://research.ewu.edu/c.php?g=82207

Emayr, S., Köpper, M., & Buchner, A. (2017). Effects of high pixel density on reading comprehension, proofreading performance, mood state, and physical discomfort. *Displays*, *48*, 41-49.

Fass, W., & Schumacher, G. M. (1978). Effects of motivation, subject activity, and readability on the retention of prose materials. *Journal of educational psychology*, *70*(5), 803.

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

Fox, J. A., Levin, J., & Forde, D. R. (2003). Elementary Statistics in Social Research.

Froese, A. D., Carpenter, C. N., Inman, D. A., Schooley, J. R., Barnes, R. B., Brecht, P. W., & Chacon, J. D. (2012). Effects of classroom cell phone use on expected and actual learning. *College Student Journal*, *46*(2), 323-333.

Gnanadesikan, A. E. (2011). *The writing revolution: Cuneiform to the internet* (Vol. 25). John Wiley & Sons.

Greenlee-Moore, M. E., & Smith, L. L. (1996). Interactive computer software: The effects on young children's reading achievement. *Reading Psychology: An International Quarterly*, *17*(1), 43-64.

Grimshaw, S., Dungworth, N., McKnight, C., & Morris, A. (2007). Electronic books: Children's reading and comprehension. *British Journal of Educational Technology*, *38*(4), 583-599.

Hidalgo, J. (2019, October 16). A Brief primer on E-Ink, what it is and how it works. https://www.lifewire.com/what-is-e-ink-2740879

Hillesund, T. (2010). Digital reading spaces: How expert readers handle books, the web and electronic paper. *First Monday,*15(4-5).

Jabr, F. (2013). Why the brain prefers paper. *Scientific American*, *309*(5), 48-53.

Johnson, L. L. (1988). Effects of underlining textbook sentences on passage and sentence retention. *Literacy Research and Instruction*, *28*(1), 18-32.

Johnson, T. E., Archibald, T. N., & Tenenbaum, G. (2010). Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills. *Computers in Human Behavior*, *26*(6), 1496-1507.

Junco, R., & Cotten, S. R. (2012). No A 4 U: The relationship between multitasking and academic performance. *Computers & Education*, *59*(2), 505-514.

Keim, B. (2014). Why the smart reading device of the future may be… paper. *Wired (May 2014)*. https://www.wired.com/2014/05/reading-on-screen-versus-paper/

Keppel, G., & Wickens, T. D. (2003). *Design and analysis: A researchers' handbook* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Kong, Y., Seo, Y. S., & Zhai, L. (2018). Comparison of reading performance on screen and on paper: A meta-analysis. *Computers & Education*, *123*, 138-149.

Kurata, K., Ishita, E., Miyata, Y., & Minami, Y. (2017). Print or digital? Reading behavior and preferences in Japan. *Journal of the Association for Information Science and Technology*, *68*(4), 884-894.

Lauterman, T., & Ackerman, R. (2014). Overcoming screen inferiority in learning and calibration. *Computers in Human Behavior*, *35*, 455-463.

Lee, S. J., & Chae, Y. G. (2012). Balancing participation and risks in children's internet use: The role of internet literacy and parental mediation. *Cyberpsychology, Behavior, and Social Networking*, *15*(5), 257-262.

Lemke, J. L. (1997). Cognition, context, and learning: A social semiotic perspective. *Situated cognition: Social, semiotic, and psychological perspectives*, 37-56.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635-654.

Lin, W. Y., Zhang, X., Jung, J. Y., & Kim, Y. C. (2013). From the wired to wireless generation? Investigating teens' Internet use through the mobile phone. *Telecommunications Policy*, *37*(8), 651-661.

Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation*, *61*(6), 700-712.

Liu, Z. (2012). Digital reading. *Chinese Journal of Library and Information Science (English edition)*, 85-94.

MacWilliam, A. (2013). The engaged reader. *Publishing Research Quarterly*, *29*(1), 1-11.

Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and instruction*, *4*(2), 91-115.

Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, *58*, 61-68.

Mauk, B. (2012). Why do we blink?. Live Science. https://www.livescience.com/32189-why-do-we-blink.html

McFadden, C. (2012). Are textbooks dead? Making sense of the digital transition. *Publishing Research Quarterly*, *28*(2), 93-99.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97.

Mizrachi, D. (2015). Undergraduates' academic reading format preferences and behaviors. *The Journal of Academic Librarianship*, *41*(3), 301-311.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard advantages of longhand over laptop note taking. *Psychological Science*, 25(6), 1159-1168.

Nichols, M. (2016). Reading and studying on the screen: An overview of literature towards good learning design practice. *Journal of Open, Flexible, and Distance Learning*, *20*(1), 33-43.

Nichols, M. (2018). Addendum: Reading and studying on the screen. *Journal of Open, Flexible, and Distance Learning*, *22*(2), 49-60.

Nicholas, D., Huntington, P., Jamali, H. R., Rowlands, I., Dobrowolski, T., & Tenopir, C. (2008, May). Viewing and reading behaviour in a virtual environment: The full-text download and what can be read into it. In *Aslib Proceedings* (Vol. 60, No. 3, pp. 185-198). Emerald Group Publishing Limited.

Niederhauser, D. S., Reynolds, R. E., Salmen, D. J., & Skolmoski, P. (2000). The influence of cognitive load on learning from hypertext. *Journal of educational computing research*, *23*(3), 237-255.

Nordqusit, R. (2017, Mar 20). Online Reading: Glossary of grammatical and rhetorical terms. ThoughtCo.https://www.thoughtco.com/what-is-online-reading-1691357

Noyes, J. M., & Garland, K. J. (2003). VDT versus paper-based text: Reply to Mayes, Sims and Koonce. *International Journal of Industrial Ergonomics*, *31*(6), 411-423.

Noyes, J., & Garland, K. (2005). Students' attitudes toward books and computers. *Computers in Human Behavior*, *21*(2), 233-241.

O'Brien, B. A., Mansfield, J. S., & Legge, G. E. (2000). The effect of contrast on reading speed in dyslexia. *Vision research*, *40*(14), 1921-1935.

Olsen, A. C. (2018, August). Worshipping in a digital age. Church of Jesus Christ. https://www.churchofjesuschrist.org/study/ensign/2018/08/worshipping-in-a-digital-age?lang=eng.

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research, and Evaluation*, *9*(1), 11.

Porter-O'Donnell, C. (2004). Beyond the yellow highlighter: Teaching annotation skills to improve reading comprehension. *English Journal*, 82-89.

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, *9*(5), 1-6.

Przybylski, A. K., & Weinstein, N. (2013). Can you connect with me now? How the presence of mobile communication technology influences face-to-face conversation quality. Journal of Social and Personal Relationships, 30(3), 237-246.

Qayyum, M. A. (2008). Capturing the online academic reading process. Information Processing *& Management*, *44*(2), 581-595.

Raphael, T. (1982). Teaching children question-answer strategies. *Reading Teacher, 36*(2), 186-191.

Rappleye, J., & Halverson, T. (2019). ScripturePlus: The future of scripture study, 2019 FairMormon Conference, Utah Valley Convention Center, Provo Utah, Aug 7, 2019.

Read, W., Robertson, N., & McQuilken, L. (2011). A novel romance: The technology acceptance model with emotional attachment. *Australasian Marketing Journal (AMJ)*, *19*(4), 223-229.

Reid, A. J., Morrison, G. R., & Bol, L. (2017). Knowing what you know: improving metacomprehension and calibration accuracy in digital text. *Educational Technology Research and Development*, *65*(1), 29-45.

Rockinsaw-Szapkiw, A., Holder, D., & Dunn, R. (2011, March). Motivating students to learn: Is there a difference between traditional books and e-books?. In *Global Learn* (pp. 235-239). Association for the Advancement of Computing in Education (AACE).

Rockinson-Szapkiw, A. J., Courduff, J., Carter, K., & Bennett, D. (2013). Electronic versus traditional print textbooks: A comparison study on the influence of university students' learning. *Computers & Education*, *63*, 259-266.

Rosen, L. D., Carrier, L. M., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948-958.

Ruel, E., Wagner III, W. E., & Gillespie, B. J. (2015). *The practice of survey research: Theory and applications*. Sage Publications.

Sanders, L. (2017). Digital Minds. *Science News*, 191(6), 18.

Schneps, M. H., Thomson, J. M., Chen, C., Sonnert, G., & Pomplun, M. (2013). E-readers are more effective than paper for some with dyslexia. *PloS One*, *8*(9), e75634.

Schugar, J. T., Schugar, H., & Penny, C. (2011). A Nook or a Book? Comparing College Students' Reading Comprehension Levels, Critical Reading, and Study Skills. *International Journal of Technology in Teaching & Learning*, *7*(2), 174-192.

Seminary and Institute at a glance (August, 2017). *Ensign Magazine.* https://www.churchofjesuschrist.org/study/ensign/2017/08/seminary-and-institute-at-a-glance?lang=eng

Seminaries and Institutes of Religion (2020). Annual Statistics Report 2020. The Church of Jesus Christ of Latter-day Saints, Salt Lake City, UT.

Shapiro, A. M. (1998). Promoting active learning: The role of system structure in learning from hypertext. *Human-Computer Interaction*, *13*(1), 1-35.

Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority-does the medium, screen versus paper, affect performance even with brief tasks?. *Metacognition and Learning*, *11*(1), 15-33.

Siegenthaler, E., Wurtz, P., Bergamin, P., & Groner, R. (2011). Comparing reading processes on e-ink displays and print. *Displays*, *32*(5), 268-273.

Silk, S. (2020), Electronic vs. Printed: A Mobile Eye Tracking Study of Attention During Reading. *The National High School Journal of Science*. https://nhsjs.com/2020/electronic-vs-printed-a-mobile-eye-tracking-study-of-attention-during-reading/

Silver, L. (2019). Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center, Washington DC. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/

Spiro, R. J. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. *Center for the Study of Reading Technical Report; no. 441, 1-13*.

Spiro, R. J. (2012). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In *Cognition, education, and multimedia* (pp. 177-220). Routledge.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Knowledge representation, content specification, and the development of skill in situation-specific knowledge assembly: Some constructivist issues as they relate to cognitive flexibility theory and hypertext. *Educational Technology*, *31*(9), 22-25

Statistics Solutions. (N.D.). *Principal componenet analysis.* Statistics Solutions. https://www.statisticssolutions.com/principal-component-analysis-pca/

Stone, R. W., & Baker-Eveleth, L. J. (2013). Students' intentions to purchase electronic textbooks. *Journal of Computing in Higher Education*, *25*(1), 27-47.

Stoop, J., Kreutzer, P., & Kircz, J. G. (2013). Reading and learning from screens versus print: a study in changing habits: Part 2-comparing different text structures on paper and on screen. *New Library World*, *114*(9-10), 371-383.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285.

Taipale, S. (2015). Bodily dimensions of reading and writing practices on paper and digitally. *Telematics and Informatics*, *32*(4), 766-775.

Tanner, M. J. (2014). Digital vs. print: Reading comprehension and the future of the book. *SLIS Student Research Journal*, *4*(2), 6-21.

Thornton, B., Faires, A., Robbins, M., & Rollins, E. (2014). The mere presence of a cell phone may be distracting. *Social Psychology*, 45(6) 479-488.

Tindell, D. R., & Bohlander, R. W. (2012). The use and abuse of cell phones and text messaging in the classroom: A survey of college students. *College Teaching*, *60*(1), 1-9.

Two sides, (2015), Reading from Paper or Reading from Screens. What do Consumers Prefer?. https://twosidesna.org/wp-content/uploads/sites/16/2018/05/TS-North-America-Reading-from-paper-or-screens.pdf

Van Horne, S., Russell, J. E., & Schuh, K. L. (2016). The adoption of mark-up tools in an interactive e-textbook reader. *Educational Technology Research and Development*, *64*(3), 407-433.

Waheed, M., Kaur, K., Ain, N., & Sanni, S. A. (2015). Emotional attachment and multidimensional self-efficacy: extension of innovation diffusion theory in the context of eBook reader. *Behaviour & Information Technology*, *34*(12), 1147-1159.

Wisniewski, P. J., Xu, H., Rosson, M. B., & Carroll, J. M. (2017, February). Parents Just Don't Understand: Why Teens Don't Talk to Parents about Their Online Risk Experiences. In *CSCW* (pp. 523-540).

# APPENDIX A

## Demographic and Multiple-Choice Questions

**Demographic Questions**

1. What is your gender?

**Male** **Female**

---

2. What grade are you in?

**Freshman** **Sophomore** **Junior** **Senior**

---

3. How many of each of the following annotation did you make while reading Alma 39 today?

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Highlights | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10 or more** |
| Tags | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10 or more** |

---

4. If you had both, which format would you prefer to use when studying the scriptures at seminary?

**Gospel Library** **Printed Scriptures** **Both Together** **No Preference**

---

5. If you had both, which format would you prefer to use when studying the scriptures at home?

**Gospel Library** **Printed Scriptures** **Both Together** **No Preference**

---

6. If you had both, which format do you think your parents would prefer that you use?

**Gospel Library  Printed Scriptures  Both Together  No Preference**

---

7. On a scale of 1-9, how confident are you in your ability to use the annotation features on the Gospel Library app?

**Not at all 1  2  3  4  5  6  7  8  9 Extremely Confident  Confident**

---

8. On a scale of 1-9, how confident are you in your ability to effectively tag verses in the Gospel Library app?

**Not at all 1  2  3  4  5  6  7  8  9 Extremely Confident  Confident**

---

9. On a scale of 1-9, how confident are you in your ability to effectively highlight verses in the Gospel Library app?

**Not at all 1  2  3  4  5  6  7  8  9 Extremely Confident  Confident**

---

10. On a scale of 1-9, how confident are you in your ability to create links to quotes from church leaders or other scriptures in the Gospel Library app?

**Not at all 1  2  3  4  5  6  7  8  9 Extremely Confident  Confident**

**Alma 39 Quiz**

1. According to Alma, what should Corianton have done differently?
   A. Followed his brother's example.
   B. Had more confidence in himself.
   C. Stayed in the land of Siron.
   D. Hidden what he did from the Zoramites.

2. Aside from his sexual sins, what else does Alma say that Corianton had done wrong?
   A. He didn't hearken to the words of his father.
   B. He was not spiritually prepared for his mission.
   C. He continually boasted in his strength.
   D. He had failed to always remember Christ.

3. Which of the following did Alma use to describe Corianton's sin?
   A. Minor.
   B. Abominable
   C. As bad as murder.
   D. Unpardonable.

4. What was it about Corianton's actions that made them particularly bad?
   A. Isabelle was a harlot, not a person he truly loved.
   B. As the son of a prophet, Corianton should have known better.
   C. Isabelle was not a believer.
   D. They kept the Zoramites from receiving the gospel message.

5. Which of the following does Alma tell Corianton not to seek after?

A. The honor and praise of men.

B. The riches and vain things of the world.

C. Happiness through iniquity

D. The desires of his heart

6. Which of the following did Alma tell Corianton?

A. You can't hide your sins from God.

B. Sexual relations are reserved for marriage.

C. Focusing on Christ can help us avoid falling to temptation.

D. Repenting of serious sins is a long process.

7. Which of the following four characters is most likely to benefit from reading Alma's counsel to Corianton?

A. Jake is a leader in his quorum, but the things he says on Sunday often don't line up with the things he does during the rest of the week.

B. Jenna is a young woman who struggles with her self-image and often thinks that she is not good enough.

C. Cade wants to believe the church is true, but struggles with some of the church's policies and history.

D. Natalie is pretty popular at the school but constantly feels the need to put others down to make herself look better.

8. From what you read about Alma's conversation with Corianton, if Alma were a modern parent, which of the following phrases would he be most likely to use?
   A. In our family we don't do things like that.
   B. I wouldn't say these things if I didn't love and care about you.
   C. Don't worry about what you've done, I still love you.
   D. It doesn't matter what you think, In this house, I make the rules

9. We don't hear anything from Corianton in this chapter, but based on Alma's response, which one of the following would you expect to have heard Corianton say to his father about the situation?
   A. "Helaman and Shiblon did stuff like this too and you never got mad at them!"
   B. "C'mon dad, it's just not like it was when you grew up anymore."
   C. "Ok I messed up, but it's not that big of a deal, relax!"
   D. "You've got the facts wrong, this isn't what it looks like."

10. Which of the following principles is the best summary of Alma's teachings to Corianton?
    A. Because of Jesus Christ we can be forgiven of our sins.
    B. Even when we love someone, sexual relations should only take place within the bounds of marriage.
    C. Parents have a responsibility to correct their children when they make mistakes.
    D. The people we surround ourselves with have a serious impact on our ability to live the gospel.

11. Based off of Alma's description of what happened to Corianton, which of the following characters best aligns with Corianton's personality:

A. Spencer loves studying the scriptures with his family and always asks questions to make sure he really understands. While he sometimes makes mistakes that he regrets, he knows how to repent and seeks to do so regularly.
B. Ashley is pretty good about reading her scriptures, but often feels like she doesn't really understand them. As a result, she often feels like her efforts to follow Christ are inadequate.
C. Nick is well-liked and has a lot of confidence in himself. While he always attends church and seminary, he isn't super willing to listen to what his teacher and classmates have to say about the gospel.
D. Amanda loves her parents, but doesn't believe in a lot of the things that they do. She keeps the commandments out of respect for her parents, but has always been clear that she intends to live a different lifestyle when she moves out of the house.

12. Which of the following quotes from modern church leaders fits best with what Alma said to Corianton?
    A. "If you will study the scriptures diligently, your power to avoid temptation and to receive direction of the Holy Ghost in all you do will be increased." Thomas S. Monson
    B. "Discord or jealousy inhibits the ability of the Holy Ghost to teach us and inhibits our ability to receive light and truth." Henry B Eyring
    C. "What a glorious blessing! For when we want to speak to God, we pray. And when we want Him to speak to us, we search the scriptures." – Robert D. Hales
    D. "When you are willing to listen and learn, some of life's most meaningful teachings come from those who have gone before you." - M. Russell Ballard

APPENDIX B

**Essay Question**

Please use your understanding of Alma 39 as the basis for explaining a principle that you could use to resolve a problem in your life.

APPENDIX C

**Parental Consent Form**

Dear Parent or Guardian,

In an effort to better understand the impact of smartphones in the classroom, your student's seminary class has been selected to participate in a research study. The study will ask students with access to smartphones to read a passage of scripture using the Gospel Library app and complete a comprehension test and questionnaire on their learning experience. As part of the experience, students will receive training on how to utilize the various annotation features in the app and assignments to help them practice using them.

To ensure that no undue burden is placed on your student's class, the text for the assignment will align with what your student was scheduled to study on the day of the experiment. Likewise, students who do not have access to a smartphone or prefer to access the text in a different format will be permitted to complete any assignments in their preferred format.

Brother Brandon Comstock, a doctoral student at Boise State University who has taught seminary in southern Utah for the past twelve years, will oversee the experiment as part of his dissertation. While results of the study will be made available electronically, the names and information of participants will be kept anonymous.

Since participation in the study aligns with the content and instruction used in the course, it is not anticipated that there will be any risks associated with the study. As with

all studies of this nature, participation is voluntary but appreciated. Additional information about the background and design of the study is available upon request or can be accessed via the QR code provided below.

Thank you for your consideration,

Brandon Comstock

Instructor - Seminaries and Institutes of Religion

Doctoral Candidate - Boise State University

By signing below, you authorize your student to participate in this study.

Student's Name _____

Guardian's Name _____

Guardian's Signature _____

APPENDIX D

**Instructions and Rubric Provided to Graders**

Prior to writing their essays, students read the text from the Book of Mormon in Alma chapter 39. The chapter summary reads as follows, “Sexual sin is an abomination – Corianton’s sins kept the Zoramites from receiving the word – Christ’s redemption is retroactive in saving the faithful who preceded it. About 74 B.C.” However, regardless of what the students’ interpretation of the chapter is your purpose is to measure how deeply each student understands the principle they identified, not how well it aligns with traditional interpretations of the chapter.

To grade the essays you will assign each individual thought introduced to one of the four levels defined in the question answer-relationship (QAR) framework. The QAR was designed to measure the depth of answers to open-ended questions. The rubric below was designed for this study using the definitions of each level of learning in the QAR.

| Answer Level | Expectation |
|---|---|
| Level 1 | The answer is pulled from a single reference in the chapter. If application is made it is nearly identical to the context of the application provided in the text. |
| Level 2 | The answer is developed from combining information from multiple references in the chapter. If application is made it is nearly identical to the context of the application provided in the text. |
| Level 3 | The answer is developed using information that is not contained in the original text. Application is still nearly identical to the context provided in the text. |
| Level 4 | The answer uses a principle from the text, but provides an application of the principle that goes beyond the context that was presented in the text. |

The following page provides examples of potential responses for each of the four levels.

**LEVEL 1**

**Response:** **"One important lesson from the text comes from verse seven, when Alma says to Corianton 'I would not dwell upon your crimes to harrow up your soul if it were not for your good.' That could be helpful if you were getting frustrated with leaders for being too hard on you."**

**Explanation:** *Notice that the answer makes a specific reference to a single place in the text and frames the application in a way that is nearly identical to what is introduced in the text. This is a clear example of a level one answer.*

---

**LEVEL 2**

**Response:** **"While Alma was a bit hard on Corianton, you can tell that he does it out of love. He tells him that he wouldn't do it 'if it were not for his good.' He also explains that he's telling him this stuff so that Corianton can avoid being led away, and so that he will be able to inherit the Kingdom of God. I think it's important when we are corrected by our parents or leaders to remember that they love us and want what's best for us."**

**Explanation:** *Notice that the answer used elements from several parts of the chapter (even though a specific verse wasn't always referenced) to make an assumption about Alma's feelings for Corianton. Still, the application was nearly identical to the context of the chapter.*

---

**LEVEL 3**

**Response:** **"The word 'harrow' was really interesting to me. Just a few chapters ago Alma talked about how he felt harrowed, now he's telling Corianton 'I would not dwell upon your crimes to harrow up your soul, if it were not for your good.' That connection shows that even though it was hard for Corianton to hear, his father was doing this because he cared about him.**

**Explanation:** *Notice that the writer used information that was not explicitly in the text. The fact that there was no direct quote or citation is irrelevant. The learner was making connections beyond what they were provided.*

---

**LEVEL 4**

**Response:** **"Alma's willingness to correct his son Corianton, even when it was probably uncomfortable to do, is something we can follow as well. Even though I'm not a parent, I have friends that make stupid choices all the time. It can be pretty hard to tell a close friend that the choices they're making are going to ruin their life, but you have to ask yourself, if you aren't willing to say something that will keep them from harm, are you really their friend?"**

**Explanation:** *Notice that in this case the writer goes beyond the example that was used in the actual text. This example demonstrates that the writer has achieved a level of cognitive flexibility that makes it possible for them to transition the content to different contexts.*

APPENDIX E

**Student Assent Script**

As part of class today you will be reading from Alma 39 and answering twelve multiple choice questions and one essay question about what you read. While the activity should feel no different than a typical day of class, results from the quiz will be used in a study on the effectiveness of different digital annotation methods on learning. Specifically, the study will explore whether creating tags and highlights helps to improve students ability to understand what they read.

To participate in the study, you will need to use the Gospel Library app on a smartphone. If you do not have access to a smartphone with the app, or prefer to study using a different format, you can still participate in the activity but your results will not be included in the study.

While you will be asked to complete a short demographic questionnaire, all responses will be anonymous, and the study will focus on the results of groups, not individuals. Regardless, if you prefer that your information not be used in the study just let me know and I will make sure that your quiz is not included in the results.

I expect that the results from the study will be available by the end of December 2020. When they are available I will take a few minutes of class to let you know what was discovered. If you are interested in reading the more detailed version, the full report will be available by May of 2021.

If you have any questions or concerns about the study you can speak with me directly, or contact the office of research compliance at Boise State University. That department can be reached via phone at (208) 426-5401 or via email at HumanSubjects@BoiseState.edu.

## Instructions Script

Today's reading will be from Alma 39. Please read the chapter heading and each of the 19 verses in the chapter. As you are reading, please highlight, tag, or look for any passages that you feel are important to understanding the message of the chapter.

After you have finished reading the chapter you may begin answering the essay question on the paper you were provided. As you can see, the essay question asks you to use your understanding of Alma 39 as the basis for explaining a principle that you could use to resolve a problem in your life. If you need additional space, please raise your hand and I will provide you with more paper.

Once you have completed the essay, please hand it in to me and begin the multiple-choice quiz. Please note that while the essay is open-book, the multiple-choice quiz is closed-book. As such, you shouldn't have your phone out when you are taking the multiple choice portion of the quiz. Once you have finished and handed in the multiple-choice portion of the quiz you are welcome to rest quietly. When everyone is finished we will review the answers to the multiple choice quiz.

APPENDIX F

**Copy of the Text Used in the Experiment**

# CHAPTER 39

*Sexual sin is an abomination—Corianton's sins kept the Zoramites from receiving the word—Christ's redemption is retroactive in saving the faithful who preceded it. About 74 B.C.*

1 And now, my [a]son, I have somewhat more to say unto thee than what I said unto thy brother; for behold, have ye not observed the steadiness of thy brother, his faithfulness, and his diligence in keeping the commandments of God? Behold, has he not set a good [b]example for thee?

2 For thou didst not give so much heed unto my words as did thy brother, among the people of the [a]Zoramites. Now this is what I have against thee; thou didst go on unto boasting in thy strength and thy wisdom.

3 And this is not all, my son. Thou didst do that which was grievous unto me; for thou didst forsake the ministry, and did go over into the land of Siron among the borders of the Lamanites, after the [a]harlot Isabel.

4 Yea, she did [a]steal away the hearts of many; but this was no excuse for thee, my son. Thou shouldst have tended to the ministry wherewith thou wast entrusted.

5 Know ye not, my son, that these things are an abomination in the sight of the Lord; yea, most [a]abominable above all sins save it be the shedding of innocent [b]blood or denying the Holy Ghost?

6 For behold, if ye [a]deny the Holy Ghost when it once has had place in you, and ye know that ye deny it, behold, this is a sin which is [b]unpardonable; yea, and whosoever murdereth against the light and knowledge of God, it is not easy for him to obtain [c]forgiveness; yea, I say unto you, my son, that it is not easy for him to obtain a forgiveness.

7 And now, my son, I would to God that ye had not been [a]guilty of so great a crime. I would not dwell upon your crimes, to harrow up your soul, if it were not for your good.

8 But behold, ye cannot [a]hide your crimes from God; and except ye repent they will stand as a testimony against you at the last day.

9 Now my son, I would that ye should repent and forsake your sins, and go no more after the [a]lusts of your eyes, but [b]cross yourself in

all these things; for except ye do this ye can in nowise inherit the kingdom of God. Oh, remember, and take it upon you, and [c]cross yourself in these things.

10 And I command you to take it upon you to counsel with your elder brothers in your undertakings; for behold, thou art in thy youth, and ye stand in need to be nourished by your brothers. And give heed to their counsel.

11 Suffer not yourself to be led away by any vain or foolish thing; suffer not the devil to lead away your heart again after those wicked harlots. Behold, O my son, how great [a]iniquity ye brought upon the [b]Zoramites; for when they saw your [c]conduct they would not believe in my words.

12 And now the Spirit of the Lord doth say unto me: [a]Command thy children to do good, lest they [b]lead away the hearts of many people to destruction; therefore I command you, my son, in the fear of God, that ye [c]refrain from your iniquities;

13 That ye turn to the Lord with all your mind, might, and strength; that ye lead away the hearts of no more to do wickedly; but rather

return unto them, and [a]acknowledge your faults and that wrong which ye have done.

14 [a]Seek not after riches nor the vain things of this world; for behold, you cannot carry them with you.

15 And now, my son, I would say somewhat unto you concerning the [a]coming of Christ. Behold, I say unto you, that it is he that surely shall come to take away the sins of the world; yea, he cometh to declare glad tidings of salvation unto his people.

16 And now, my son, this was the ministry unto which ye were called, to declare these glad tidings unto this people, to prepare their minds; or rather that salvation might come unto them, that they may prepare the minds of their [a]children to hear the word at the time of his coming.

17 And now I will ease your mind somewhat on this subject. Behold, you marvel why these things should be known so long beforehand. Behold, I say unto you, is not a soul at this time as precious unto God as a soul will be at the time of his coming?

18 Is it not as necessary that the plan of redemption should be [a]made known unto this people as well as unto their children?

19 Is it not as easy at this time for the Lord to [a]send his angel to declare these glad tidings unto us as unto our children, or as after the time of his coming?

APPENDIX G

**Links to Training Examples**

Links to Training Examples

Training example #1 – Highlighting in Alma 36

Training example #2 – Tagging in Alma 36

Training example #3 – Highlighting in Alma 37

Training example #4 – Tagging in Alma 37