ABSTRACT

Title of Dissertation:	READING IN PRINT AND DIGITALLY: PROFILING AND INTERVENING IN UNDERGRADUATES' MULTIMODAL TEXT PROCESSING, COMPREHENSION, AND CALIBRATION
	Lauren Melissa Singer Trakhman, Doctor of Philosophy, 2019
Dissertation directed by:	Professor Patricia A. Alexander Department of Human Development and Quantitative Methodology

As a consequence of today's rapid-paced society and ever-changing technologies, students are frequently called upon to process texts in print and digitally. Further, multimodal texts are standard in textbooks and foundational to learning. Nonetheless, little is understood about the effects of reading multimodal texts in print or digitally. In Study I, the students read weather and soil passages in print and digitally. These readings were taken from an introductory geology textbook that incorporated various graphic displays. While reading, novel data-gathering measures and procedures were used to capture real-time behaviors. As students read in print, their behaviors were recorded by a GoPro@ camera and tracked by the movement of a pen. When reading digitally, students' actions were recorded by Camtasia@ Screen Capture software and by the movement of the screen cursor used to indicate their position in the text. After reading, students answered comprehension questions that differ in specificity (i.e., main idea to key

concepts) that cover content from three sources: text only; visual only; and, both text and visual. Finally, after reading in each medium, undergraduates rated their performance on the comprehension measure on a scale of 0-100 for each passage. The accuracy of these ratings formed the basis of the calibration score. The processing data were analyzed using Latent Class Analysis. In Study II, an intervention aimed at improving students' comprehension and calibration when reading digitally were introduced to participants from Study I who returned to the lab about two weeks later. Next, the undergraduates repeated the procedure for digital reading outlined in Study I with a passage on volcanoes. In Study I, students performed better when reading in print and spent more time with the text but were better calibrated when reading digitally. Three clusters were identified for the print data, and three clusters were identified for the digital data. Cluster movement across mediums suggests that some participants treat digital texts differently than when reading in print. After the intervention in Study II, comprehension scores and duration increased but calibration accuracy scores worsened. The LCA revealed three clusters, each showing improvement in processing behaviors, comprehension, or reading duration.

READING IN PRINT AND DIGITALLY: PROFILING AND INTERVENING IN UNDERGRADUATES' MULTIMODAL TEXT PROCESSING, COMPREHENSION, AND CALIBRATION

by

Lauren Melissa Singer Trakhman

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:

Professor Patricia Alexander, Chair
Associate Professor Daniel Dinsmore
Professor Mariam Jean Dreher
Associate Professor Geetha Ramani
Professor Min Wang

Acknowledgments

Patricia, thank you for seeing a potential in me that I did not know I had.

Geetha, Dan, Min, and Jean, thank you for the gift of your time and expertise.

Gary, thank you for believing in me and having the patience to put up with the past few years.

Malia, my four legged co-author, who has been by my side for every word in this document.

Mom, thank you for your unwavering enthusiasm and support.

Dad, thank you for teaching me the value of working hard and finishing what I've started.

Drew, thank you for your unconditional support.

My DRLRL family, thank you for your support, insight, and generous gifts of time.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

In light of the networked and technologically-advanced world in which today's students live and learn, it should come as no surprise that undergraduates in a recent study overwhelmingly preferred to read an informational text digitally rather than on paper (Singer & Alexander, 2017a). Further, after answering questions about what they read, a significant majority of those undergraduates judged their performance as better when they read on computer rather than on paper. Yet, surprisingly, comprehension performance for most participants in that study was significantly better when they read printed rather than digital texts for all but gist level (i.e., main idea) questions. This comparison of students' predicted or self-judged performance to their actual is referred as *calibration* (Alexander, 2013; Fischhoff, Slovic, & Lichtenstein, 1977; Glenberg, Sanocki, Epstein, & Morris, 1987). The outcomes of this study proved informative in two ways. First, it suggested that there may be advantages for reading in print rather than digitally even for capable readers. Second, this outcome suggested that undergraduates were particularly poor at judging their comprehension performance when text was read on a computer.

When my co-author and I encountered these unexpected outcomes, I was compelled to undertake a second investigation of medium effects to ascertain whether those results were merely an anomaly. In that follow-up study (Singer Trakhman, Alexander, & Berkowitz, 2017), longer informational texts were used and reading times were tracked. These modifications were included on the basis of a systematic literature review in which we found that text length and processing time were influential factors in studies reporting comprehension differences by medium (Singer & Alexander, 2017b). Specifically, the longer the text and the more scrolling involved, the greater the likelihood that reading on paper was found to be more advantageous to

comprehension than reading on the screen (Dundary & Akcayiri, 2012; Lenhard, Schroeders, & Lenhard, 2017). In addition, there were researchers who documented faster times for digital than for print reading and who hypothesized that this more rapid processing may negatively affect comprehension performance (Dyson & Haselgrove, 2000). Wickelgren (1977) referred to this hypothesized effect as the speed-accuracy tradeoff.

In that follow-up study (Singer Trakhman et al., 2017), the same seemingly paradoxical patterns emerged. Again, the clear preferences students displayed for digital reading and their performance judgments on a just-completed comprehension test conflicted with actual comprehension outcomes. In contrast to what the students judged after answering comprehension questions, performance for most was superior in print when they were called upon to list the key points from the readings or to recall any other relevant information. Only when the focus was on the main idea of the passages did the effect for medium fail to manifest. There was also a marked reading time difference for the two mediums. Reading digitally took significantly less time overall than reading in print. In accordance with the speed-accuracy tradeoff hypothesis (Wickelgren, 1977), we surmised that digital processing speed may well have been a contributory factor in the lower comprehension scores for that medium.

In yet another follow-up investigation (Singer Trakhman, Alexander, & Silverman, 2018a), we delved deeper into the effect of medium on text comprehension by not only including longer texts and recording reading times, as in the prior study, but also by examining students' use of more shallow and deeper processing behaviors while reading. In order to capture these processing behaviors, we had students underline text as they read in print while wearing Go-Pro© cameras. When reading digitally, students followed along with their cursor, and their screen movements were recorded using Camtasia Screen Capture© software. As in

prior studies, students answered questions immediately after reading that targeted the main idea, key points in the text, and other relevant information. The undergraduates again were asked to judge their comprehension performance. In contrast to the previous studies, however, we conducted person-centered analyses in an attempt to unearth processing profiles that may help to explain how students' behaviors were contributing to the consistent outcomes that had emerged for print and digital reading. Those profiles were created on the basis of the frequency of deeper-level or surface-level processing behaviors the undergraduates manifested, as well as their processing times. The same four profiles emerged for print and digital reading (i.e., Regulators, Plodders, Gliders, and Samplers).

The validity and significance of the resulting profiles were then confirmed through interrater reliability checks; comparisons to profiles documented in the literature on levels of strategic processing and approaches to learning (e.g., Dinsmore & Alexander, 2016; Vermunt & Donche, 2017); and, discriminant factor analyses. Once the profiles were verified for print and digital mediums, comprehension and calibration data were analyzed by medium and by profile. For those profiles, those who exhibited a reliance on deeper processing behaviors (Regulators) and those who moved through the text fluently, although linearly (Gliders), registered the highest comprehension scores and were better calibrated than those who moved linearly but slowly through the text (Plodders), and those who only selectively exhibited deeper processing behaviors (Samplers).

While the association among reading behaviors, comprehension performance, and calibration scores proved consistent across print and digital mediums, those populating the four profiles varied to some degree. That is to say, a number of participants changed profiles depending on the medium in which they were reading. Perhaps most intriguing in their

characteristics were the Gliders, who performed fluently when reading in print and scored well on the comprehension questions. However, a good portion of those falling into the Gliders profile for print took on the attributes of the Regulators when reading moved to the computer. The performance of this particular contingent of students, who adjusted their reading to be slower and more deliberate when the text was digitally displayed, ran counter to the majority of readers in this investigation and prior studies whose processing speed was faster when reading digitally (Singer & Alexander, 2017a; Singer Trakhman et al., 2017). This intriguing outcome suggests that more insights might be gained from additional person-centered analyses of reading in print and digitally.

In all prior studies, one feature of the research remained consistent—all experimental materials consisted only of text. Thus, in the most recent study in this program of research (Singer Trakhman, Alexander, Sun, & Silverman, 2018b), my colleagues and I sought to advance understanding of medium effects when texts were multimodal (i.e., consisting of text and visual content). For this study, each text included four visuals (two pictures, one diagram, and one map). We chose to use multimodal texts in this study for two reason. First, most of the textbook materials that undergraduates are required to read are multimodal in nature. Thus, we could test for medium effects in materials that were ecologically valid. Second, in our systematic review of the literature, there was little or no consideration of medium effects when text processing involved visual as well as textual content. Consequently, we could explore whether these non-linguistic elements had any appreciably effect on print or digital reading.

Important changes were also made to the measures in this study. For one, the two selections, which came from an introductory Geology textbook (Pipkin, Trent, Hazlett, & Bierman, 2014), differed according to their perceived level of familiarity. This determination

was made on the basis of a pilot study in which undergraduates rated their familiarity with 20 geology topics on a 0-100 scale. From those data, we selected the text segments from the textbook corresponding to the topic with the highest familiarity score and lowest standard deviation, *Weather*, and the topic with the lowest familiarity score and lowest standard deviation, *Soil*. Second, a dedicated measure of topic knowledge was created and administered to students. As in prior studies, the undergraduates also rated their familiarity with the topics of Weather and Soil on a 0-100 scale. Third, the comprehension measure was extended to include questions that asked about content conveyed only in the text, information conveyed in a visual only, and information that came from both the text and visuals.

To our knowledge, this study is one of the first investigations to be both multi-medium *and* multimodal. Based on the analyses of the resulting data, certain trends remained consistent. For one, reading in print was again found to result in significantly better comprehension than reading digitally and calibration. For another, medium had no discernible effect on students' ability to identify the main idea. However, important differences did emerge in this study that appear attributable to the multimodal nature of the texts. First, although print reading continued to translate into better comprehension performance for all but the gist level question, the magnitude of the difference, as indicated by the effect sizes, was greater for the question that relied on visual information only and the question where the answer could be found both in the visual and in the text. Second, this was the first time in the series of studies my colleagues have conducted (Singer & Alexander, 2017a; Singer Trakhman et al., 2017; Singer Trakhman et al., 2018a) where reading times were not significantly faster when texts were displayed on computers versus on paper. Third, it was also the first time that those reading digitally were significantly better calibrated than those reading in print.

Because processing behaviors were not recorded in this latest study, we cannot determine precisely what contributed to these novel outcomes. However, as it pertains to the anomalous findings from time and calibration, we can offer one plausible explanation based on our observations during the task administration. Specifically, given the length of the text and the various visual displays, students did not have access to an entire page of text when reading digitally as they did when reading directly from the textbook. Consequently, students were seen to scroll back and forth frequently when reading digitally on what constituted a single page of printed text. In fact, some students became visibly frustrated with having only a partial view of a page, and some chose to reduce the size of the document on screen so that they could visually access more of the content on the page. It is reasonable to assume that the scrolling behaviors we observed added processing time in the digital condition and caused the students to be more mindful of their level of performance than might otherwise be the case. Nonetheless, it would seem essential to gather processing behaviors in the future as a way to test such reasonable assumptions more systematically.

The Current Study

Purpose

Despite what has been learned about medium effects on comprehension, there is still much to be learned regarding the differences in comprehension and calibration when students read in print and digitally. Further, to my knowledge, there have been no attempts to examine the effects of an intervention that targets the speed, processing behaviors, comprehension, calibration, and group membership for individuals reading digitally. This dissertation undertakes such an intervention, but also introduces other relevant modifications to the

experimental measures, study design, and data-analyses suggested by the prior studies my colleagues and I have conducted.

In terms of experimental measures, this dissertation includes a dedicated topic knowledge measure, which consisted of three multiple-choice questions and three terms from the text passage to be defined. This topic knowledge measure is an advancement over earlier investigations that relied only on a self-report topic knowledge measure (Rolfhus & Ackerman, 1996). In fact, in the most recent investigation (Singer Trakhman et al., 2018b), topic knowledge was measured similarly and found to be a significant covariate. For this study, I again incorporated topic knowledge as a covariate, as indicated in Figure 1. The second improvement over previous studies pertains to the comprehension measure. In all prior investigations, the question about the main idea of the passage was found to not be informative with regard to medium effects. Therefore, it was decided that a similar question would be of little value in this study and would be replaced with several items that would rely on different information sources in the text passage. Specifically, the comprehension measure consisted of seven short constructed-response questions that draw on information from (a) the text only (2), (b) visuals only (2); and (c) text and visuals (3).

Further, there are two changes to the study design. For one, this dissertation improves the way in which the digital text is displayed. As we observed in the most recent study (Singer Trakhman et al., 2018b), when students read the digital texts they were unable to access a full page with visuals as easily as they could in the print condition. This led to more scrolling than we had witnessed previously and may have affected the findings regarding reading time and calibration for the digital condition. To resolve this issue, this dissertation presented the digital textbook pages in their e-book form directly from the publisher. These pages provided better

visual access to the text and pictures than in the most recent study. The second and largest change to the study design involves the introduction of an intervention (Study II) aimed at increasing students' self-regulation of their digital reading time and depth of processing. This intervention took place a few weeks after the initial study and involved the participants from Study I.

Finally, for this dissertation, as in Singer Trakhman et al. (2018a), I collected real-time processing behaviors using a GoPro© Camera and Camtasia Screen Record© Software to capture undergraduates' movements through the text passages. As in the prior study, the students signaled their position in the text by using either a pen or mouse curser to indicate the specific words they were reading. These processing behaviors, along with reading time and calibration accuracy, served as the database for Latent Class Analysis (LCA). LCA is a more sophisticated method of person-centered analyses than previously used.

Research Questions

As displayed in Figure 1, this dissertation examined the effects of medium (print or digital) on students' reading speed and processing behaviors, which then are presumed to affect comprehension and calibration, after controlling for prior topic knowledge. Comprehension was examined by means of a seven-item comprehension test consisting of short-answer questions and calibration accuracy was based on the difference between their comprehension test score and their judgment of performance. I conducted both variable-centered and person-centered analyses to address the following research questions across two studies:

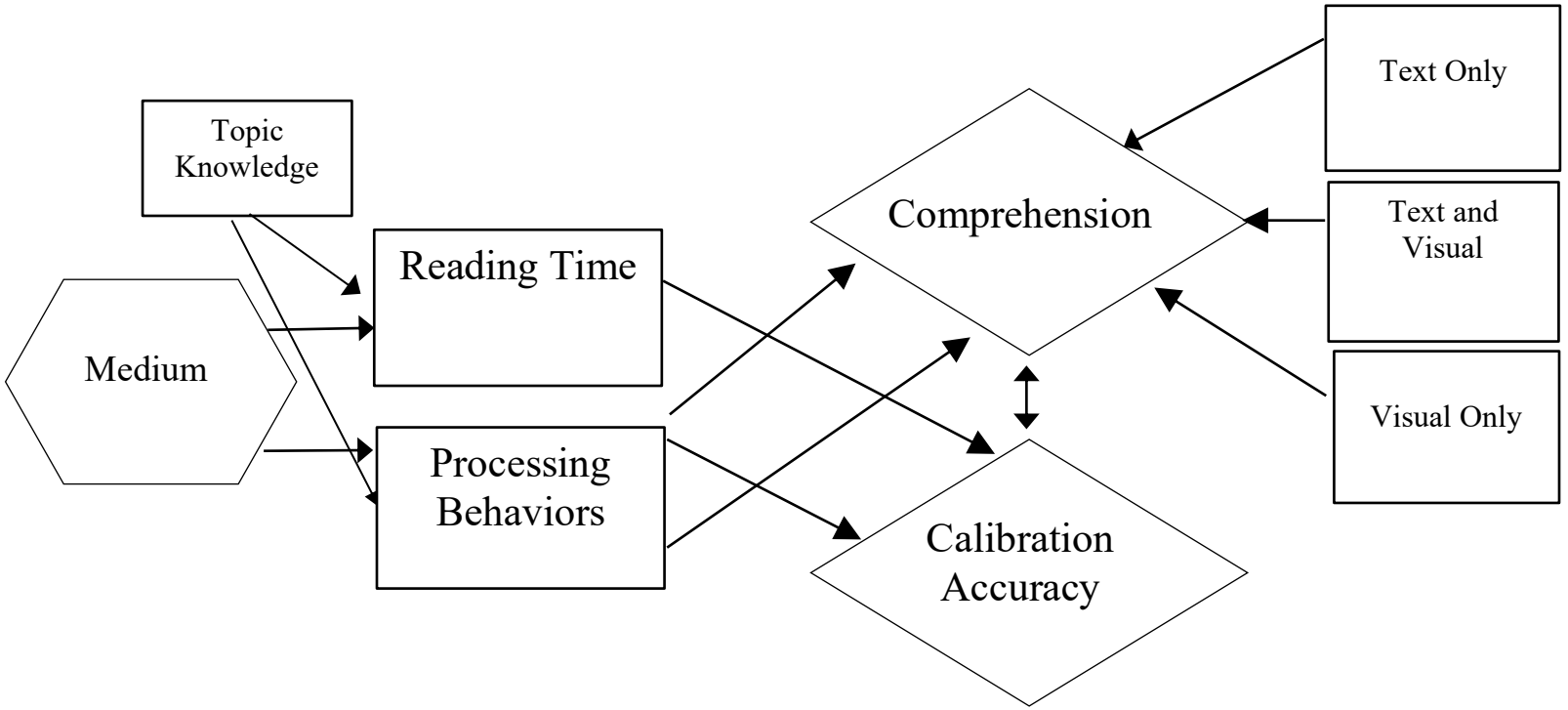

Figure 1. The above displays the conceptual model for this study. As indicated, the medium of delivery (print or digital) was predicted to affect participants' reading time and processing behaviors, which then would affect comprehension and calibration accuracy outcomes when controlling for the participants' topic knowledge.

Study I

1. How do students' reading times and real-time processing behaviors affect their comprehension performance and calibration accuracy when processing multimodal expository texts in print or digitally, after adjusting for demonstrated topic knowledge?

Based on prior studies (Singer & Alexander, 2017a; Singer Trakhman et al., 2017; Singer Trakhman et al., 2018a), it was hypothesized that reading time would be slower in print and overall comprehension would be better for those reading in print, even after controlling for topic knowledge. Although the most recent investigation found reading time to be slower when reading digitally (Singer Trakhman et al., 2018b), this dissertation was designed to reduce the navigational issues students encountered in the digital condition in that recent investigation. Further, I expected calibration accuracy to be relatively poor across mediums, as in previous studies (e.g., Singer Trakhman et al., 2018a; Singer Trakhman 2018b). However, I predicted calibration to be significantly more accurate in the print than digital condition.

2. What verifiable groups emerge from real-time processing behaviors, reading time, and calibration accuracy for undergraduates' reading of print and digital texts? Are there significant differences in comprehension performance for the resulting groups for print and digital text processing?

Similar to prior findings (e.g., Singer Trakhman et al., 2018a), I hypothesized that there will be verifiable groups that emerge for print and digital reading. Similar to findings from the previous study (Singer Trakhman et al., 2018a), I expected the groups identified in the print condition to parallel those in the digital condition. However, I hypothesized that there will be a number of participants whose group membership shifts when reading moved from print to digital.

Study II

Based on the findings from Study I, the following research questions were addressed in Study II:

3. What are the effects of an intervention targeting self-regulation of digital reading time and depth of processing on undergraduates' comprehension performance and calibration accuracy when reading digitally?

The intervention consisted of a pre-recorded training containing five modules. The decision to pre-record the modules was made to rule out any differences in delivery that might occur across participants. The training consisted of five modules in which general information pertinent to reading in print and digitally were summarized and illustrated (e.g., students' reading preferences, speed of reading, judgment of performance, and different reader profiles). After each section, students were asked to respond to a question that deals with the information covered that is followed by immediate feedback. The entire training took 30 minutes to complete. After completing the training, students completed a prior knowledge measure, read a passage, and complete the comprehension and judgment of performance measures digitally. It was hypothesized that after completing the intervention, participants reading speed in the digital condition would be significantly slower that their performance in Study I and the manifestations of processing behaviors that show more depth in comparison to the peripheral behaviors would significantly increase. Further, comprehension scores and calibration accuracy were hypothesized to show significant improvement overall, although it was expected that individuals who navigated through the text quickly using more surface-level behaviors in Study I would contribute most to that outcome.

4. What verifiable groups emerge from real-time processing data, reading time, and calibration accuracy for undergraduates' reading digital texts after an intervention? Are there significant differences for comprehension performance for the resulting groups from pre-intervention to post intervention?

I hypothesized that as a consequence of the intervention, there will be fewer profiles identified because there will be less variability in the manifest variables that made up the groups (e.g., processing behaviors, reading duration, and calibration accuracy).

Key Terms

Calibration Accuracy refers to the magnitude of the difference between an appraisal of one's performance and their actual performance (Bol, Hacker, O'Shea, & Allen, 2005).

Medium is defined as the manner of information delivery (Franklin & Zdonik, 1998), which in this dissertation means the presentation of informational texts in print or digitally.

Multimodal Texts are written texts that contain additional elements, such as visuals, hyperlinks, video clips, audio clips, or other modes of representation (Kress & van Leeuwen, 1996).

Judgment of Performance signifies an individual's self-evaluation of his/her level of accomplishment for a given task and is used in calculating calibration accuracy scores.

Text Comprehension is a dynamic cognitive process that allows students to understand written text, develop and interpret meaning, and use meaning as appropriate to the type of text, purpose, and situation (NAEP, 2009). In this proposed dissertation, "comprehension" was used with the same definition.

Topic Knowledge indicates the individuals' prior knowledge of the text subject (Alexander, Schallert, & Hare, 1991). Within this dissertation, topic knowledge was measured on the three text topics being covered in the reading comprehension tasks.

Visuals are non-linguistic representations that can take various forms, including photographs, illustrations, diagrams, maps, graphs, and tables (Schnotz, 2002).

CHAPTER 2

REVIEW OF THE LITERATURE

In this chapter, the theoretical background for this dissertation will be put forward, and relations among constructs displayed in the conceptual model that framed this investigation (Figure 1) will be explicated. Specifically, this review begins with an overview of relevant theory and empirical research on text comprehension, including the studies that address the effects that medium and modality play in the act of reading for understanding. Next, the literature pertaining to judgments of performance and calibration is summarized with consideration of how these variables are defined, measured, and influenced. Similarly, the empirical research on text processing behaviors is surveyed with the intention describing how this variable has been conceptualized and operationalized. As part of this survey, the research dealing with depth of processing and approaches to learning will be presented. Finally, this review will touch on interventions that have attempted to influence students' comprehension and strategic processing of text. The purpose of this brief foray into the intervention research is to identify features of effective comprehension and strategy interventions that were considered in the design of this investigation.

Reading: Learning from Text

In general, what I systematically investigated in this dissertation pertained to the processes and outcomes displayed by competent readers as they attempt to learn from expository texts. Consequently, while reading is a multidimensional and complex process that involves such fundamental skills as phonological awareness, fluency, and word knowledge (RAND Reading Study Group, 2012), such basic skills play far less of a role in the performance for older and more experienced (i.e., more competent) readers (Alexander & Fox, 2011; Cassidy &

Cassidy, 2007). This act of reading is commonly referred to as learning from text (Alexander & Jetton, 2000). As such, the ultimate goals of learning from text are to understand or comprehend what is read (NAEP, 2009; Singer et al., 2017b) and to remember and use the information acquired to accomplish a cognitive task (Anderson & Pearson, 1984).

In the context of this dissertation, undergraduate students were asked to process expository passages from the domain of geology in print and digitally for the purpose of answering questions about the information those passages conveyed both in text and in visuals. Thus, the theoretical and empirical foundations for this investigation pertain not only to text comprehension per se, but also comprehension under multi-medium and multimodal conditions.

Text Comprehension and Competent Readers

Regardless of whether one is a beginning or experienced reader, the basic conceptualization of text comprehension remains consistent. That is, *text comprehension* can be defined as the ability to understand written text, develop and interpret meaning, and use meaning as appropriate to the type of text, purpose, and situation (NAEP, 2009). The ability to comprehend text is of importance because this skill is foundation to learning—no matter the domain or level of expertise (Baker, Dreher, & Guthrie, 2000; Kintsch & Van Dijk, 1987). Within this dissertation, the decision to focus on competent readers as the population of interest was done to ensure that the examination of medium effects on comprehension and calibration would not be unduly influenced by more basic reading processes. Instead, these more competent readers are freer to focus on the meaning of the text (Alexander & Jetton, 2000).

The most widely-accepted theoretical model of reading comprehension is the construction-integration model of comprehension (Kintsch, 1988). Consistent with Kintsch's (1988) model, readers are expected form connections between their own prior knowledge and

ideas expressed in or inferred by the text per se. With this in mind, this dissertation was designed to use topic knowledge scores as a covariate when examining reading time and processing behaviors effects on comprehension. That decision is consistent with Kintsch's (1988) model in which prior knowledge of the text topic is projected to influence comprehension for even competent readers. Also, within Kintsch's construction-integration framework, there is also an acknowledgment that the type of text (i.e., narrative or expository) could well translate into differences in comprehension performance (Kintsch, 1988). As such, the dissertation aimws to introduce expository multimodal texts, which are frequently encountered by competent readers.

Text comprehension has been measured in a variety of ways, and the influx of standardized testing has only increased the frequency at which readers encounter these measures (RAND Reading Study Group, 2012). Frequently used methods include (a) the cloze technique, wherein one word is omitted from a sentence or passage and the student shows understanding by providing the missing word (Woodcock-Johnson Tests of Achievement-III; Woodcock, McGrew, & Mather, 2001); (b) multiple choice comprehension questions (Gray Oral Reading Test 3; Wiederholt & Bryant, 1992); (c), open-ended short-answer questions (Qualitative Reading Inventory 3; Leslie & Caldwell, 2001); (d) retelling of a passage, which involves a student having to retell a story to the proctor after reading it and is scored dependent on if the student understood the main idea of the passage and the frequency of detailed idea-units recalled (Qualitative Reading Inventory 3; Leslie & Caldwell, 2001); and (e) performance assessments, which ask the reader to display their understanding of a text through the creation of a product on the subject, such as a PowerPoint presentation (NAEP, National Center for Education Statistics, 2013).

The reading comprehension task used in my previous research has had participants respond to three short-construction questions immediately after reading a passage (e.g., Singer Trakhman et al., 2018b). My colleagues and I chose to assess comprehension using short-construction questions in light of positive evidence regarding the merits of this format, such as its broader applicability and the richer information garnered from generated versus selected response formats like multiple-choice items (Pearson & Hamm, 2005). Further, short-construction questions eliminate random guessing as a response option (Burton, 2001). In addition, well-known national assessments such as the National Assessment for Education Progress (NAEP; National Center for Education Statistics, 2013) and the Stanford Achievement Tests (SAT-9; Harcourt Educational Measurement, 1996), as well as a number of statewide reading tests, have long used short-construction questions to delve deeper into competent readers' comprehension (Sarroub & Pearson, 1998).

Reading in Print and Digital Mediums

Medium is a term that reflects the manner of information delivery through which a message is communicated (Franklin & Zdonik, 1998). This can encompass a physical transmission medium such as an optical fiber, or a presentation medium, such as an oral presentation. In this dissertation, medium has a more specialized meaning both in the mode of delivery and the content. Specifically, while the medium of text delivery changes from print to digital, the content that is present stays the same (Bodmann & Robinson, 2004). While reading on the computer, participants viewed the same texts and visuals in the same manner in which they were presented in print. This means there are no additional elements to the digital text, such as hyperlinks or videos.

Within the contemporary literature regarding reading, there are some who believe that reading digitally is different than reading in print (Coiro, Knoble, Lankshear, & Leu, 2008; Forzani & Maykel, 2013; Leu, Forzani, Rhoads, Maykel, Kennedy, & Timbrell, 2015). This "new literacy" perspective frames online reading comprehension as a process of problem-based inquiry involving skills, strategies, dispositions, and social practices (Leu et al., 2015). As a consequence, new skills and strategies are needed when reading online. Leu states that the most important difference between reading and online reading is that when reading online, there is a greater need for increased amounts of higher-level thinking (Leu et al., 2015). However, there are some researchers who disagree with those who define digital reading as a process that is different from reading (Kerr & Symons, 2006; Kurniawan & Zaphiris, 2001; Singer et al., 2017b). These researchers claim that the skills described by Leu et al. (2015) for digital reading are describing a construct that extends beyond reading text on a digital display. Leu and colleagues' (2015) discussion of the unique processing demands that come with processing in a digital environment seem to be describing digital reading, as opposed to reading digitally. In effect, they are describing the processing of Internet materials that include not only text but all other features common to online environments (e.g., hyperlinks, animations, and videos). This dissertation views the act of reading to be fundamentally equivalent across the two mediums because the text is a consistent factor.

Within a review of the literature (Singer & Alexander, 2017b), one of the research questions asked how those examining reading in print and digital mediums were measuring comprehension. This review determined that there was noticeable variability in the measures and approaches that researchers employed to test participants' comprehension. For example, of the 36 studies reviewed, 15 included multiple-choice reading comprehension questions (e.g.,

Noyes et al., 2004). Six studies included short construction questions (e.g., three short-response items; Singer & Alexander, 2017a). The remaining studies incorporated comprehension questions that were either unspecified, free recalls, or text summaries.

The literature review also noted that a majority (61.11%) of researched employed multiple comprehension measures within one study. For example, Kerr and Symons (2006) used free recall, cued recall, and multiple-choice questions to assess comprehension. Although this effort to include multiple indicators of reading comprehension is a positive characterization of the charted literature, other documented patterns raised critical questions. For instance, the psychometric properties of study measures were not always well documented. Moreover, the majority of the studies (63.89%) included researcher-developed measures of reading comprehension rather than established assessments, and in many of those instances (91.42%), the reliability and validity of data from those researcher-developed measures were not provided. For examples, Noyes et al. (2004) detailed their comprehension task as a measure consisting of "10 multiple-choice questions followed by the administration of the NASA-TLX" (p. 112). Although their measure of comprehension may be a sound choice, the lack of details regarding reliability and validity is concerning.

Despite these concerns, the literature base has shown differences across mediums in terms of reading comprehension (e.g., Kerr & Symons, 2006; Mangen, Walgermo, & Brønnick, 2013; Singer & Alexander, 2017a; Singer Trakhman et al., 2018b). Within our systematic literature review on this topic (Singer & Alexander, 2017b), the results suggested that medium plays an influential role under certain text or task conditions or for certain readers. Additionally, the researchers aimed to identify how comprehension has been defined and measured over the

past quarter century. Analysis showed that relatively few researchers defined either reading (n=9) or digital reading (n=5).

Research finding better comprehension outcomes for print reading. For instance, in a study involving 72 tenth graders in Norway who read text digitally or in print, Mangen et al. (2013) found that students who read print versions had significantly higher scores in reading comprehension than those who read digitally. Similarly, Kerr and Symons tested the recall of 60 fifth-grade students in Canada who each read two passages, one digitally and one in print. What the resulting data indicated was that participants recalled more from the printed text than from the digital text. Further, a survey by Rideout, Foehr, and Roberts (2010) suggested that those who read in print reported that they were less likely to multitask than when reading digitally.

Research finding better comprehension outcomes for digital reading. Within their literature review, Singer and Alexander (2017b) identified multiple studies where comprehension outcomes were higher for those reading digitally. For example, Lenhard et al. (2017) had fifth-grade students read texts that were 311-372 words in print and digitally. After reading, the students were asked free recall, cued recall, and multiple-choice comprehension questions. Although it took longer for students to complete the comprehension measure for the digital medium, their performance on the digital multiple-choice comprehension questions was significantly higher when compared to their performance on the print comprehension measure. Verdi, Crooks, and White (2014) also found the digital medium to be more advantageous for comprehension of geographical facts related to maps.

Research finding no differences in comprehension by medium. Several of the studies outlined by Singer and Alexander (2017b) reported no significant differences in comprehension outcomes as a consequence of medium. For example, Akbar, Al-Hashemi, Taqi, and Sadeq

(2013) had 76 adults in their 20's read a short text and then were quizzed with recall questions. These authors found no difference in recall ability as a result of the medium in which the text was read. Similarly, Margolin, Driscoll, Toland, and Kegler (2013) found no significant differences in reading comprehension scores on a multiple-choice quiz given to 90 participants after reading a short one-page text either in print or digitally.

Possible explanations. Researchers have reported that repeated or intensive engagement in digital multitasking may contribute to a more superficial processing of the textbase (Levine, Waite, & Bowman, 2007; Ophir, Nass, & Wagner, 2009; Wallis, 2010). For instance, Liu, Aamodt, Wolf, Gelernter, and Mark (2009) reported that when working on a digital device, readers switched activities every three to 10 minutes. Therefore, these researchers argued that it seemed improbable for these readers to engage in deep thought when they are switching activities so often during the processing of texts. Even in those instances when readers are not switching activities, they may be more apt to scan digital texts than to read them deeply (Rowlands et al., 2008; Wallis, 2010). It is also conceivable that the very wealth of information that resides in the digital universe, combined with the increasing speed and ease of access, may overwhelm intentions to process such digital content critically or analytically (Ophir et al., 2009; Wallis, 2010).

Alongside these aforementioned differences in readers' processing of digital texts, there are visual ergonomic characteristics of digital media that must also be entered into the comprehension calculus. Specifically, visual legibility of digital texts, basic to word processing and comprehension, is influenced by several factors, including text size, screen resolution, backlighting, and luminance contrast (Dillon & Emurian, 1995; Lee, Ko, Shen, & Chao, 2011). And such factors can result in physical and perceptual differences for digital versus print reading.

For example, LCD computer screens, the most common method of reading digitally, are found on most desktop, laptop, and handheld tablets such as an iPad. Researchers have found that these screens might contribute to visual fatigue as a result of the lighting source (e.g., Benedetto, Drai-Zerbib, Pedrotti, Tissier, & Baccino, 2013; Mangen et al., 2013). Further, Garland and Noyes (2004) determined that other features of a LCD screen, such as its refresh rate, contrast levels, and fluctuating light, can interfere with text recall.

Yet another perceptual factor that might play into processing differences between printed and digital text has to do with sequential versus continuous reading; that is whether a reader must scroll frequently between portions of text or whether reading progresses in a more uninterrupted fashion (Proaps & Bliss, 2014). From the available research, there is evidence that frequent scrolling increases the cognitive demands on readers and, thus, may negatively affect recall (Wästlund, 2007; Wästlund, Reinikka, Norlander, & Archer, 2005). The inclusion of hyperlinked materials can also increase cognitive demands when individuals are processing digital texts (DeStefano & LeFevre, 2007). Moreover, print texts allows readers to see and feel the spatial extension and physical dimensions of the text, and the material of the paper provides physical, tactile, and spatiotemporally fixed cues to the length of the passage (Mangen, 2010; Sellen & Harper, 2002).

The unexpected outcomes showing a comprehension advantage for print over digital reading led researchers to look more deeply at the effects of medium. Specifically, works examined the effect of reading time on comprehension across mediums. These modifications were included on the basis of a systematic literature review (Singer & Alexander, 2017b) that suggested that text length and processing time were influential factors in reported comprehension differences by medium (Lenhard et al., 2017). Specifically, the longer the text and the more

scrolling involved, the greater the likelihood that reading on the page was found to be more advantageous to comprehension than reading on the screen (Dundary & Akcayiri, 2012). In addition, there were researchers who documented faster times for digital reading than for print and who hypothesized that this rapid processing negatively affected comprehension performance (Dyson & Haselgrove, 2000).

In an earlier investigation, my colleagues and I discovered that students displayed clear preferences for digital reading and that their confident performance judgments for the digital medium on a just-completed comprehension test conflicted with actual results (Singer Trakhman et al., 2017). In contrast to what the students predicted, their comprehension was superior in print when called upon to list the key points or to recall any other relevant information from the readings. Only when the focus was on the main idea of the passages did the effect for medium fail to manifest. There was also a marked time difference for the two conditions. Reading digitally took significantly less time than reading in print. In accordance with the speed-accuracy tradeoff hypothesis (Wickelgren, 1977), my co-authors and I surmised that the speed of processing texts digitally may well have been a contributory factor in the lower comprehension scores in that medium, and this hypothesis was found to be correct in our follow up investigation (Singer Trakhman et al., 2018a). Once again, my co-authors and I found that when reading digitally, participants navigated the text much more quickly, while comprehension scores suffered the consequences.

Reading Multimodal Texts

Students in the 21st century are increasingly exposed to texts that are multimodal, which contain two or more of the following elements: text, visuals, hyperlinks, video clips, audio clips, and other modes of representation (Goldstone, 2004; Kress, 2003). Multimodal elements are

often included in texts in order to attempt to enhance a text in a way that written words cannot do alone (Kress, 2003). For example, science texts are often supplemented with visual elements in order to demonstrate representations that are hard to grasp from the written word alone (Alvermann & Wilson, 2011). There is abundant evidence that adding a visual element to a text can support and enhance learning for certain populations, such as the population in this dissertation (Ainsworth, 2008; Goldman, 2009). However, these multiple modalities need to be introduced in a manner that uses the visual elements in a way that supports text comprehension (Mayer, 2002). Specifically, an effective multimodal text allows for computational offloading, which is when an external representation, such as a visual, serves to reduce the amount of cognitive effort required to comprehend a text or solve a problem (Larking & Simon, 1987). When done correctly, learning from two representations (i.e., text and visuals) can be better than text alone (Goldman, 2003). However, researchers have frequently suggested that students often do not apply key the self-regulatory processes and mechanisms such as effective cognitive strategies and metacognitive monitoring necessary when learning from multimodal texts (e.g., Azevedo & Cromley, 2004; Azevedo, Guthrie, & Seibert, 2004). For this dissertation, I used multimodal texts from a geology text that contained written texts and visual elements because that is the kind of reading materials that undergraduates frequently encounter in their academic studies (Alvermann & Wilson, 2011).

Researchers interested in multimodal texts have turned to more contemporary methodologies for measurement, such as eye-tracking. For example, Holsanova (2010) used eye-tracking technology to monitor how individuals access multimodal newspapers digitally. She found that certain types of visuals were looked at more than others. Specifically, diagrams attracted more attention than photographs. In fact, photographs within the newspaper were only

looked at early on and for the shortest duration of time in comparison to the other visuals included. This finding illustrates the notion that readers do not process multimodal texts in a linear manner, as often is seen in passages with text as the singular modality (Mangen et al., 2013). To comprehend multimodal texts at a level that probes beyond the gist, it is necessary to direct one's attention to both the text and visuals in a non-linear manner (Holsanova, 2010).

The measurement of comprehension of multimodal texts is often assessed in the same manner as a text-only passage is (diSessa, 2004). Researchers tend to rely on traditional measures such as multiple-choice questions and short-answer questions that are often lacking the reliability and validity standards one would expect (Leu, Kinzer, Coiro, Castek, & Henry, 2017). However, there have been attempts to measure multimodal learning that extend beyond the pencil and paper test. For example, Hegarty, Kriz, and Cate (2003) examined the effect of a multimodal passage containing text and animations of a mechanical system by asking participants to verbally describe the dynamic processes they expected the systems to enact. Verbalization has been used by other researchers examining the effects of multimodal texts (e.g., Ainsworth & Loizou, 2003; Butcher, 2006; Moore & Scevak, 1997). Ainsworth and Loizou (2003) used this method in order to assess the effects of multimodal texts in comparison to a text-only passage. Despite the plethora of ways in which researchers have attempted to measure multimodal texts, Mayer notes that the technology for multimodal texts is developing and being introduced at a faster pace than a corresponding science of how we can measure learning in multimodal environments (Mayer, 2009).

Within this dissertation, I incorporated multimodal texts consisting of text and visuals from a science textbook. The comprehension of multimodal texts has been found to be of particular importance for learning from science texts (Otero, Léon, & Graesser, 2002). However,

there is substantial evidence that learners often fail to maximize the potential benefits of multimodal representations (Cromley, Snyder-Hogan, & Luciw-Dubas, 2010; Van Meter, Cameron, & Waters, 2017). Fortunately, attempts to modify the regulatory behaviors of readers as they navigate multimodal texts shows encouraging findings regarding improvement on this crucial skill (Bartholomé & Bromme, 2009; Mason, Pluchino, & Tornatora, 2015).

Judgments of Performance and Calibration

What They Are and Why They Matter

Within this dissertation, two related constructs dealing with students' perceptions of how well they have performed are of importance—judgments of performance and calibration. Simply described, *judgments of performance* or JOPs are the subjective determinations that individuals make about how well they have done on a given task or cognitive activity (Nelson, 1984). In the research on metacognition and self-regulation, these judgments play a crucial role because they influence the decisions that individuals make about whether more or different effort is necessary during task performance (Dinsmore & Parkinson, 2013). In most instances such judgments are initiated by individuals without any external prompting. In the current investigation, students are explicitly directed to numerically rate how well they have done on the reading tasks they were presented. These judgments also play a role in the calculation of calibration accuracy scores that are a key variable in this investigation.

Calibration refers to the degree of association between an individual's predicted and actual performance (Alexander, 2013; Fischhoff et al., 1977; Glenberg et al., 1987), and is a major component of several metacognitive and self-regulatory models of learning (e.g., Winne & Hadwin, 1998). In effect, the notion of calibration puts students' judgments of their performance up against the actual quality of their performance. When there is a marked discrepancy between

student's predicted and actual performance, either reflecting overconfidence or low confidence, the learning process is negatively affected. As such, the ability to accurately judge the quality of reading performance plays a critical role in an individual's ability to successfully regulate their learning (Dinsmore & Parkinson, 2013).

The literature on calibration accuracy suggests that judgment of performance and calibration abilities for children, youth, and adults has often been found wanting (Chen & Zimmerman, 2007; Hacker, Bol, & Bahbahani, 2008; Pajares & Miller, 1994). Students are found to be overly confident in their reading performance, for example, which leads them to engage at a more surface-level than deeper-level with the information (Alexander & Jetton, 2000; Goldman, 2009). In fact, studies conducted in a variety of subject-matters (e.g., computer science, Murphy & Tenenberg, 2005; math, Pajares & Kranzel, 1995; reading, Lin & Zabrucky, 1998) and with a range of academic tasks (e.g., addition problems, Bryan & Bryan, 1991; essay writing, Graham & Harris, 1989; multiple choice exam, Bol & Hacker, 2001) have consistently reported that participants demonstrate poor calibration abilities (Boekaerts & Rozendaal, 2010; Klassen, 2002; Stahl, Pieschl, & Broome, 2006). In effect, students are not particularly astute at critically evaluating and accurately judging their understanding of and memory for what they read (Ackerman & Goldsmith, 2011; Winne, Jamieson-Noel, & Muis, 2002). As was mentioned in the discussion on medium, competent readers' judgments of their performance and calibration accuracy were found to be even more problematic when they read digitally (Singer et al., 2017a; Singer Trakhman et al., 2018a).

Measuring Judgments of Performance and Calibration

In the research on judgments of performance and calibration, participants' self-judgments of their performance are collected and analyzed in different ways. Those differences pertain to

when self-ratings are made, *what* they are rating, and *how* the self-ratings are analyzed (Barnett & Hixon, 1997; Nelson & Dunlosky, 1991; Rawson & Dunlosky, 2007). I will first explain the differences in when these self-ratings are made. Researchers have collected these data after participants studied for a specific academic task but before they actually begin the task (Dunlosky, Rawson & Middleton, 2005). Alternatively, researchers have waited until the participant has completed the measure to collect the JOP data (Bryan & Bryan, 1991).

JOPs can also differ in the grain size of the rating; that is, whether they are being asked to evaluate their performance on a single item, a section of a task, or the entire task. For example, in their study on calibration accuracy in early elementary-aged students, Lyons and Ghetti (2011) used a calibration measure that asked participants if they were confident or not confident about their answer being correct after answering a single item. Conversely, Bryan and Bryan (1991) collected JOPs from participants after they had completed an entire 50-item arithmetic measure. For this decision, researchers recommend considering the underlying assumptions about the monitoring process involved in the task when selecting the appropriate JOP measure. For example, whether monitoring of potential correct and incorrect answers happens through a single process or separately through distinct processes (Schraw, Kuch & Gutierrez, 2013).

Calibration measures must also consider the scale at which they will capture JOP ratings. These can range from a dichotomous nominal measure (i.e., how confident are you that the answer is correct? Confident or not confident) to a ratio level (i.e., how confident are you that the answer is correct? Rate by selecting a number 0-100, 0 being I have no confidence in my correctness, 100 being I am fully confident that my response is correct). For example, when measuring calibration accuracy in early elementary-aged students, Lyons and Ghetti (2011) used a dichotomous nominal measure of calibration to ask participants if they were confident or not

confident in their correctness. When selecting the appropriate scale, experts recommend contemplating practical considerations. For example, when working with young learners, using a dichotomous measure may prevent overwhelming their working memory (Rutherford, 2017).

Once the JOP measure is completed, calibration accuracy has been calculated in two ways, resulting in different outcome scores: (a) the correlation between JOP and actual score, a number ranging from -1 to 1 and, (b) the absolute value of the difference between JOP and actual score, which typically falls from 0 to 100. Those who measure calibration accuracy using correlation (e.g., Maki, Shields, Wheeler, & Zacchilli, 2005; Rhodes & Tauber, 2011) do so with a scale ranging from -1 to 1, with a positive correlation nearing 1 indicating more accurate monitoring of performance. The second way in which calibration accuracy has been monitored is to consider JOP's in relation to actual performance by calculating the difference between these two scores (e.g., Ackerman & Goldsmith, 2011; Butler, 1998; Graham & Harris, 1989; Lin, Moore, Zabrucky; 2001; Singer et al., 2017a). In my previous research, I have transformed JOPs and comprehension scores to 0-100 and taken the absolute value of the difference between the two variables to calculate the calibration accuracy score. A lower calibration score would then be indicative of higher calibration accuracy.

Factors Influencing Judgments of Performance and Calibration

Researchers have examined a range of cognitive (e.g., knowledge, Winne & Hadwin, 2008) and motivational (Zimmerman, 2000) factors that may affect judgments of performance and calibration accuracy ability. In regard to the population of this dissertation, studies found that, generally speaking, college-aged students struggled with calibration accuracy despite their cognitive advantages when compared to younger populations (Barnett & Hixon, 1997; Lipko, Dunlosky, Hartwig, Rawson, Swan, & Cook, 2009). Research on calibration has involved a

variety of factors that influence accuracy and bias in judgments of performance, including personal characteristics and task parameters. For one, individuals bring certain characteristics to the learning setting (e.g., attitudes and beliefs), which are impacted by the influences embedded within the task itself (e.g., task format, item difficulty).

Personal characteristics include any traits that the individual brings to the setting (Schraw, 2009). Several personal characteristics explored in the literature include: the role of working memory in the accuracy of judgments, verbal ability, and sociocultural differences. Sprenger and Dougherty (2006) espoused that working memory capacity is often associated with probability judgments (i.e., comprehension monitoring). In regard to verbal ability, more proficient readers were found to be better equipped to predict and judge performance on a reading task than poor readers (Maki et al., 2005). Finally, research on cultural differences suggests that cultural influences, such as attitudes towards confidence, may affect judgments of performance (Wallsten & Gu, 2003).

The task itself may also have an influence on calibration accuracy. More specifically, item difficulty can significantly constrain calibration accuracy abilities, no matter the individual's ability level (Burson, Larrick, & Klayman, 2006; Schraw & Roedel, 1994). Further, understanding the importance or goal of the task itself has been found to effect calibration accuracy (Dougherty & Franco-Watkins, 2003; Dougherty & Sprenger, 2006). In fact, the instructions given about the task itself have been found to also effect calibration accuracy outcomes (Nietfeld & Schraw, 2002). Despite all that is known about the aforementioned factors that may influence calibration, questions still remain regarding mediums influence on calibration accuracy.

Judgments of Performance and Calibration and Medium of Processing

Three prior studies have expressly examined individuals' calibration accuracy when engaged in reading in print and digital conditions (Ackerman & Goldsmith, 2011; Singer & Alexander, 2017a; Singer Trakhman et al., 2017). In those investigations, researchers found that students were more confident in their performance in the digital than print condition but actually performed as well or better when reading in print than digitally. For instance, Ackerman and Goldsmith (2011) found that although participants scored similarly in comprehension tasks when reading in print and digital forms, their judgments of performance were often overconfident in the digital medium. In our previous investigation (Singer & Alexander, 2017a), we also found that participants judged their performance more favorably in the digital than the print condition. However, in that earlier study, comprehension performance for key points and other relevant information was significantly higher for the print condition.

While acknowledging that the past decades of metacomprehension research have not painted the most flattering picture of calibration accuracy among learners (Dunlosky & Lipko, 2007) including college students (Winne & Jamieson-Noel, 2002; Winne et al., 2002), there remains the possibility that the students in the current investigation will be better attuned to their ability to comprehend under differing mediums than what has typically been reported. As with prior research (Ackerman & Goldsmith, 2011; Butler, 1998; Graham & Harris, 1989; Lin et al., 2010), our previous work has considered calibration in relation to undergraduates' post-task predictions as to the medium (i.e., digital versus print) in which they had performed best and then compared those predictions to their actual performance.

What might account for this calibration effect of reading digitally versus in print? One plausible explanation for such a pattern comes from the work of Koriat, Ma'ayan, Sheffer, and

Bjork (2006) on the association between level of effort exerted in task performance and the JOPs made. In essence, the less effort that is exerted in task performance, the higher individuals' JOPs are expected to be. In the conclusion of their study, Ackerman and Goldsmith (2011) voiced such an assumption when they argued that the ease and speed of reading digitally may contribute to the miscalibration they witnessed by negatively effecting students' regulatory behaviors.

Within the empirical work examining print and digital mediums, there is a paucity of studies that expressly investigated time as an explanatory factor in comprehension outcomes (Singer & Alexander, 2017b). In fact, in our systematic review of the literature, we were unable to address time as a variable because so few of the studies (i.e., less than 25%) included any measure of processing time (Singer & Alexander, 2017b). We were able to locate a few studies where reading time was documented (e.g., Ackerman & Goldsmith, 2011; Dundar, & Akcayri, 2012; Gill, Mao, Powell, & Sheidow, 2013; Noyes & Garland, 2003; Van de Vijver & Harsveldt, 1994), although in those cases processing time was controlled (Ackerman & Goldsmith, 2011) or was not precisely gauged (Gill et al., 2013; Noyes & Garland, 2003). In one investigation in which processing time was carefully measured in relation to comprehension, Van de Vijver and Harsveld (1994) found that applicants to the Dutch Royal Military Academy completed the assessment battery faster on the computer than on paper, but that performance on the comprehension portion was higher for the print than for the digital condition.

Although research in print and digital reading comprehension often overlooks the potential effect of processing time on comprehension outcomes, there is literature within the perceptual-motor domain that suggests that speed of processing is related to quality of performance. Notably, the speed-accuracy tradeoff hypothesis (Wickelgren, 1977) contends that there is a tradeoff between how fast a task is performed and the quality or accuracy that results.

For example, Dyson and Haselgrove (2000) examined comprehension performance of undergraduates who were asked to read articles from *National Geographic* at both normal (244 words per minute) and fast (460 words per minute) rates. These authors determined that comprehension performance was significantly better at normal rates than at faster rates.

Text Processing Behaviors

In this dissertation, analysis of competent readers' performance involved the documentation of various behaviors that they manifested while engaged with print or digital texts. Over the course of prior studies investigating medium effects on undergraduates' comprehension and calibration, my co-authors and I found that these processing behaviors were key to understanding how and how well students learn from the assigned readings (e.g., Singer Trakhman et al., 2018a). Although I use the general label "processing behaviors" to identify the actions observed as participants navigate through the print and digital texts, the theoretical and empirical support for their inclusion and characterization comes from the literature dealing with strategic processing (Afflerbach, Pearson, & Paris, 2008; Alexander & Judy, 1988; Schunk, 2000) and depth or level of processing (Biggs, 1978; Dinsmore & Alexander, 2016; Marton & Säljö, 1976). My decision to use the more generic terminology was predicated on the fact that there would be no way, given the study design, to determine whether any behaviors observed and coded would reflect automatic or intentional acts (Block, 2009; Cano, 2007).

Strategic Processing

Strategies are conscious processes used to work toward a goal (Afflerbach et al., 2008; Hacker, 1998). Strategies can be either general or domain specific (Alexander, 2006). General strategies, as their name implies, are those procedures that can be used in different contexts and with a range of problems and can be applied and adapted to a specific situation, such as rereading

or paraphrasing (Alexander, Grossnickle Peterson, Dumas, & Hattan, 2018), whereas domain-specific strategies are particular to a field of study (e.g., titrating or mathematical proof; Pressley, 2000). Another way in which strategies have been classified is into cognitive and metacognitive or regulatory strategies (Garner, 1987). As it pertains to reading, cognitive strategies are the mental processes and procedures that individuals' use for support as they develop internal procedures that enable one to perform a complex task (Garner, 1987; Pressley, 2000). Metacognitive or self-regulatory strategies, by comparison, are those processes or procedures that allow individuals to oversee or monitor their mental processing as well as the choice and effectiveness of the cognitive strategies that they implement (Flavell, 1979).

Strategies are of importance for learning because they contribute to the acquisition and utilization of information, including information contained in text (Carrell, 1985; Meyer & Ray, 2011). However, this important skill does not always develop independently. Strategies need to be learned and supported (Carrell, 1985). Pressley (1988) stressed that students should be taught strategies that help them monitor their own learning in order to enhance performance.

Strategic processing has been measured in a plethora of ways, including:

- think-aloud procedures in which participants narrate what they are thinking as they complete a task,
- retrospective interviews, which entail participants explaining their rationale for behaviors made after completing a task,
- eye-tracking studies that involve the use of technology to monitor individuals' eye movement as they complete a task, and
- self-report measures, where participants respond to questions about themselves without researcher interference

All the aforementioned methods of data gathering about readers' strategic processing come with certain advantages and disadvantages. For example, think-aloud procedures are thought to extend the amount of time it takes to complete a task (Jakobsen, 2003). Due to reading duration being a variable I was interested in measuring within this dissertation, a think-aloud would be an inappropriate method for capturing strategic processing data. Retrospective interviews raise validity threats because often participants do not remember the specific processes they used during task performance (Pöchhacker, 2010). While eye-tracking studies do offer real-time insights into participants' movements, it cannot be applied to the reading of print as it can be with digital materials, thus limiting its use in this dissertation.

Despite the presence of these multiple ways to study strategic processing, there remains an overreliance on self-report measures (Dinsmore, Alexander, & Loughlin, 2008; Levin, 1983; Pressley, 1990). The overreliance on self-report measures in the strategic processing literature is troubling because the literature on calibration has demonstrated how poorly individuals judge their own performance (Dinsmore & Parkinson, 2013; Winne & Hadwin, 2008, Zimmerman, 2000). As such, individuals may lack the ability to provide accurate answers to self-report questions. In order to try to negate the potential threats to validity encountered with popular methods of strategic processing measures, I gathered real-time processing data while students read both print and digital texts through GoPro© and Camtasia Screen Record© Software. This real-time method for capturing movement while reading a text allowed for in-the moment access to how students navigate a page, without the disadvantages that come with the previously described methods. Still, the shortcomings of this alternative way to study processing must be reiterated in that there is no way to know for certain if I am capturing behaviors indicative of strategic processing or habituated actions.

Levels of Processing

Within the psychological literature, perhaps the earliest theoretical model of deep and surface processing was articulated by Marton and Säljö (1976). In their program of research, these researchers attempted to investigate qualitative differences in the outcome and processes of learning. Through semi-structured interviews, the authors identified two different levels of learner engagement while participants read text, surface-level and deep-level processing. These levels spoke to the aspects of the learning material on which the learner focused. Surface-level processing suggested that students had a "reproductive conception of learning and direction their attention toward the features of the text itself," whereas deep-level processing indicated that students "directed their intentions toward comprehending what the author had to say, or what the features of the text signified" (Dinsmore & Alexander, 2012, p. 9). This conceptualization of deep and surface-level processing strategies was the start to an important line of research aimed at understanding how students navigate texts and what effect these strategies have on comprehension and informed my conceptualization of processing behaviors, which are defined as either peripheral (e.g., *Read, Skim)* or behaviors which show more depth relative to the peripheral group (e.g., *Re-Read, Re-Position).*

Deeper and surface-level strategies. Dinsmore and Alexander (2016) contend that surface-level strategies are the actions related to "the basic encoding of textual content" (p. 214), which means the processes needed just to make sense of the text or the task. In contrast, deeper processing entails the "use of strategic and monitoring behavior that involves a more extensive manipulation or transformation of a task or text" (Dinsmore & Alexander, p. 215), such as re-reading a complex sentence. As evidenced in the Model of Domain Learning (MDL), deeper- and surface-level strategies are present at all stages of domain expertise (Alexander, 1997, 2003).

The ability to know when and how to enact deeper-level strategies during appropriate times, such as when assessing author credibility or forming a mental representation, is a crucial skill necessary for learning. The MDL stresses the notion that both deeper and surface-level strategies have a place in text-based learning. Surface-level strategies can and should be used to form an understanding of the text. In order to successfully comprehend, competent learners must be able to discern the appropriate times and ways to enact both deeper and surface-level strategies.

Previously, eye tracking has been used to record students' movements through digital texts as a means for examining such processing behaviors (Ariasi, Hyönä, Kaakinen, & Mason, 2017; Ariasi & Mason, 2011). Eye-tracking has helped researchers understand reading processes over the past decades (Rayner, 2009). Jarodzka and Brand-Gruwel (2017) used eye-tracking tools to examine reading at three levels: (a) reading single words of sentences, (b) reading and comprehending a complex text, and (c) reading multiple text documents. No matter the reading level, researchers have used data from the tracking tools, such as second-pass (i.e., re-reading) durations as an indicator of deeper processing. Some researchers have also associated particular eye movements patterns, such as long gaze fixations on extraneous text, with more surface processing (Catrysse et al., 2018; Hyönä & Lorch, 2004).

Although the findings from these eye-tracking studies are informative, this technique was only employed for reading of digital texts and does not provide insights into the processing of print texts. This dissertation aimed to capture peripheral and behaviors showing more depth using a methodology from a recent investigation (Singer Trakhman et al., 2018a), wherein real-time processing behaviors were captured using a GoPro© and Camtasia Screen Record©

Software to record as the reader "follows along" in the text with a pen or their mouse cursor. These behaviors served as the basis for the formation of processing profiles.

Processing profiles. Catrysse et al. (2017) have used eye-tracking data from students in order to create strategic profiles. Their cluster analysis identified four learning profiles with differing reliance on surface or deep processing: (a) all-lows—low scores on both deep and surface processing; (b) surface—high scores on surfacing processing and low scores on deep processing; (c) deep—high scores on deep processing and low scores on surface processing; and, (d) all-high—high scores on both deep and surface processing. However, this is not the first example of using observations of strategic behavior to create processing profiles. For example, Lindblom-Ylänne and Lonka (1998) profiled learners based on observations of deeper- and surface-level strategies that medical students used when studying for an exam. Their results suggest that observing strategic processing behaviors and using them to create groups is an effective way to understand factors influencing an individual's success (Lindblom-Ylänne & Lonka, 1998)

In an effort to investigate processing profiles, my colleagues and I used real-time processing data to construct reader profiles (Singer Trakhman et al., 2018a). In this investigation, we identified four profiles that distinguished readers based on their real-time processing behaviors. The four profiles—Regulators, Plodders, Gliders, and Samplers—were similar to the groups identified by Dinsmore and Alexander (2016). For this dissertation, I again examined what groups of readers emerge using processing behaviors as a manifest variable, along with reading duration and calibration accuracy.

Reading and Strategic Interventions

Previous reviews on interventions have found that the inclusion of strategy instruction yields significant improvements in comprehension performance (Biancarosa & Snow, 2004; Kleeck, 2008; McNamara, 2012; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007; Suggate, 2016). Although these interventions were targeting the *outcome* of reading, close inspection of the relevant literature suggests that effective interventions are often made up of various activities designed to help students *during* reading—such as using strategies to attend to particular aspects of the text (Kendeou & van den Broek, 2005). In addition to improvements in comprehension, there is also evidence that interventions that include instructions on how to monitor comprehension as one navigates text can increase other outcome variables, such calibration accuracy (Kamil, 2004).

The literature on effective interventions has stressed that no matter the desired outcome of the intervention, it is critical to incorporate strategy instruction. For example, Meyer, Young, and Bartlett (1989) had a study where they examined the specific effects of strategic instruction within an intervention. Specifically, the first group participated in an intervention that modeled how to connect ideas across sentences and paragraphs, while the second group received an intervention that simply drilled definitions and correct answers to similar comprehension measures, while the third group was the control group and received no intervention. Meyer et al.'s (1989) results suggest that focusing on the strategies used to comprehend the text and monitor their own understanding, instead of focusing on the material covered, proved to be successful for the reading comprehension task.

Kamil (2004) found similar results regarding the power of strategy instruction during an intervention were found when he looked at students' understanding of text structure. In one

study, eye-gaze patterns were altered by an intervention on text comprehension delivered online, resulting in longer durations of eye-fixations on relevant text when compared to pre-test data (Yeh, Schwartz, & Baule, 2011). Further, interventions that provide feedback about processing behaviors (e.g., Nietfeld & Schraw, 2002) have been found to be effective at improving outcome variables such as judgments of performance. In addition to judgments of performance, feedback incorporated into an intervention has also been found to be an effective tool for improving comprehension outcomes (Gutierrez, 2012).

Within this dissertation, the intervention was designed using recommendations from the literature regarding best practices, such as those just described (Gutierrez, 2012; Kamil, 2004; Nietfeld & Schraw, 2002; Yeh et al., 2011). For example, I modeled effective processing behaviors, as suggested by Meyer et al., 1989, by including videos of individuals navigating digital text that highlighted behaviors that are indicative of more deeper-level processing behaviors when compared to the peripheral behaviors. Secondly, the training module included multiple opportunities for feedback, which has been documented in successful interventions (Nietfeld & Schraw, 2002). Finally, the intervention was delivered online. This decision was made in order to rule out any differences in delivery that might occur across participants, while still using a method of delivery that has been documented as effective for interventions and familiar to the proposed population (Yeh et al., 2011).

Conclusion

This literature review examined the relevant theoretical and empirical research related to reading in print and digital mediums and about the processing of multimodal exposition. Moreover, overviews of text comprehension, judgments of learning and calibration, text processing behaviors, and strategy interventions were presented. From this review, lingering

questions and concerns related to reading in print or digitally were specified and considered in the design of this dissertation. Among those lingering questions and concerns that emerged from this literature review are: Will the factors hypothesized to influence comprehension and calibration for competent readers hold when they process multimodal texts? Will an improved measure of comprehension prove informative as to why medium effects in comprehension arise? What profiles will manifest among competent readers after processing multimodal expository text passages in print and digitally? Can processing behaviors, comprehension, and calibration accuracy be improved as a result of an intervention that expressly targets speed and depth of processing when reading digitally?

Within the review of the literature, factors that may influence reading time and processing behaviors were examined. Based on these findings, it is hypothesized that the very wealth of information that resides in the digital universe, combined with the increasing speed and ease of access, may overwhelm intentions to process such digital content critically or analytically (Ophir et al., 2009; Wallis, 2010). Further, this ease of access may be a factor in the findings regarding more shallow-processing behaviors when reading digitally (Ophir et al., 2009; Wallis, 2010).

The literature on comprehension revealed noticeable variability in approaches to its assessment. As it pertains to the processing of multiple-medium, multimodal texts, comprehension measures are needed that probe understanding beyond the gist level and that tap into the different information sources (e.g., photos, diagrams, or text). Given these findings, the comprehension measures included items that require recall of and inferential thinking about specific content within the texts. Moreover, those items were expressly written to require information drawn from the text only or the visuals only, or content found in both modalities.

Next, the literature review overviewed judgments of performance and calibration accuracy. Specifically, the literature stressed the association between these judgments and readers' metacognitive awareness and self-regulatory behaviors. In effect, the judgments that readers form regarding the level of their understanding and the quality of their performance contribute to their decisions to exert further time and strategic effort on learning from text. The procedure for calculating calibration accuracy scores that was implemented in this dissertation was established on recommendations from this review.

With regard to text processing behaviors, this literature review examined how strategic processing has been measured previously (i.e., think-aloud procedures, retrospective interviews, eye-tracking studies, and self-report measures). Additionally, it highlighted the advantages and disadvantages of each method. In this dissertation, I attempted to negate the potential threats to validity encountered with popular methods of strategic intervention measures by collecting real-time processing data as competent readers processed the text.

Next, the literature on deeper and surface-level strategies examined how processing behaviors came to be measured as indicators of strategic processes. Additionally, it discussed occurrences within the literature where researchers tried to create strategic profiles. This dissertation builds on findings from previous studies on processing profiles and strategy research. Finally, the literature review examined reading interventions.

With regard to reading interventions, the literature review highlighted the importance of incorporating strategy instruction into an intervention. Additionally, the literature review highlighted other components of an intervention that have been associated with successful outcomes. The intervention within this dissertation included the following features as per the

recommendations within the literature (a) model effective processing behaviors, (b) create multiple opportunities for feedback, and (c) deliver the material via an online platform.

CHAPTER 3
METHODOLOGY

Purpose

I pursued multiple goals in this two-study dissertation. For one, and as displayed in Figure 1, the purpose was to examine the effects of medium (i.e., print or digital) on the speed at which undergraduate students read expository text and the processing behaviors they display, which were hypothesized to influence their comprehension performance and their ability to judge that performance accurately. For another, there was the purpose of investigating what different groups of readers emerge when their reading behaviors, reading times, and their calibration accuracy were used as the basis for cluster membership. Further, there was the goal of determining if an intervention that focuses on students' regulation of their speed and processing behaviors when reading digitally not only results in improvement in comprehension and calibration accuracy over prior performance, but also instigates changes in cluster membership.

Design

To achieve the aforementioned goals, I designed this investigation to incorporate both variable-centered and person-centered analyses across two experimental studies. In Study I, I conducted variable-centered analyses using Multiple Analysis of Covariance for print and digital reading conditions, while controlling for topic knowledge. For the person-centered analyses, I conducted a Latent Class Analysis for the print medium using processing behaviors, reading time, and calibration accuracy scores to identify clusters and then again for the digital medium using the same variables to identify clusters. This person-centered analysis is useful for identifying variation within a sample where individuals can be grouped with other similar individuals (Laursen & Hoff, 2006).

Study II included as many participants as possible from Study I who were willing to return to the lab about two weeks after completing Study I. Within Study II, both variable and person-centered analyses were again conducted. For the variable-centered analyses, a repeated measures ANOVA compared the students' digital comprehension scores and calibration accuracy scores following an intervention to their digital comprehension scores and calibration accuracy scores in Study I. For the person-centered analyses, a LCA was conducted using the post-intervention processing behaviors, reading time, and calibration accuracy as the manifest variables to identify clusters. In addition, changes in students' cluster membership for the digital condition from Study I to Study II will be tested using a chi-square analysis.

Method: Study I

Participants

Undergraduate students were the target population for this dissertation. For undergraduate students, reading in print and digital mediums is prevalent in academic contexts (Singer & Alexander, 2017b). In spite of this, there is evidence that students at the undergraduate level are often unsuccessful in comprehending as well as when reading in print after reading digitally and are even poorer judges of their performance in that medium then they are in print (Ackerman & Goldsmith, 2011; Singer & Alexander, 2017a; Singer Trakhman et al., 2017).

Participants were 70 undergraduate students enrolled in human development courses at the University of Maryland, College Park. The sample was 69.5% female with a mean age of 19.82 *(SD*= 1.71*)* years. Their average GPA was 3.31 (*SD* = 0. 39) and they represented a variety of majors, primarily in the social (44.8%) and natural (29.7%) sciences. The sample included White (46%), Asian (17%), Hispanic (17%), African American (16%), and Other (4%).

Experimental Texts

Passages from *Geology and the Environment* (Pipkin et al., 2014) were used as the experimental reading materials. The passages covered two topics from this introductory college textbook (i.e., Soil and Weather), and these topics were presented in counterbalanced order. In this way, every participant will have read both passages but half will have read "Soil" in print, while the other half will have read "Weather" in print. The passages are similar in length (about 3 pages long) and readability level (grade 14; Fry Readability; Fry, 1968). Each passage contained four visuals, which are comparable across texts. The first is a diagram that graphically represents information presented in the text (e.g., soil profile shown layer by layer). The next two visuals are photographs, which serve to illustrate the topic explained in the text. Within the weather passage, for example, there is a photograph of a flooded street that accompanies the section on flooding, thus illustrating the effects of severe weather. The final visual is a map, which encapsulates spatial relations briefed in the text and which are difficult to express in words alone (Levie & Lentz, 1982). For example, in the weather passage, a world map displays areas of the world with below and above average rainfall.

Measures

Demographics Questionnaire. Participants completed a demographic questionnaire in which they reported personal and educational background (see Appendix A). They were asked to report their age, sex, race, and whether they are a native English-speaker. Additionally, the students were asked to provide information about their year in school, major, and current grade point average. For this dissertation, demographic variables were collected for reporting purposes only and were not entered into any analyses.

Topic Knowledge Tests. In order to measure students' knowledge of soils and weather, participants completed knowledge measures for each topic. These measures were used in my most recent investigation and performed well (Singer Trakhman et al., 2018a). The measures consist of three key terms from the passages that the students are to define (e.g., define the following term: hydrologists) and three multiple-choice questions that are based on explicit, factual content from the passage (see Appendix B). For example, the Weather pre-test asks the following question:

Why are heavy thunderstorms a summer phenomenon?

a. The land needs water to cool off due to the climate

b. Strong solar heating makes the atmosphere unstable

c. Moisture streams north of the equator due to the Earth's rotation

d. Air masses move differently in high temperatures, causing storms

Comprehension Measures (CM). A comprehension measure was developed for each passage that improves on previous investigations (e.g., Singer et al., 2017a) by removing the main idea question and adding several questions that rely on different information sources in the passage. Students completed the measure immediately after reading the associated passage and took the measure in the same medium in which that passage was read. Each comprehension test consisted of seven short-response questions. The first two questions asked about information found in the text portion of the passages only (e.g., what is a loess? List up to 2 of its physical properties), while the next two questions relied on information in the visuals only (e.g., according to the information visually presented in the passage, where would you find 4 areas of the earth with extensive loess covering?). The final three questions involved content included in both the text and visuals (e.g., name and describe 4 of the 5 layers in an idealized soil profile).

The comprehension questions and scoring guides were reviewed by two science teachers in order to judge the accuracy of the items and the proposed response guides. The questions for both passages, along with the answer keys, can be found in Appendix C.

Prior to scoring, training materials with simulated responses paralleling those for the actual dissertation were developed. Once the second rater reached a scoring proficiency level of 95% or above on the training materials, the actual data were introduced. To establish interrater agreement, I worked with the second rater to score 30% of the data independently and compared scores. Interrater agreement for the data set was 99.6%.

Judgment of Performance (JOP). In order to assess students' ability to judge their comprehension performance accurately, I had them judge how well they did immediately after completing that test in each medium. Specifically, after reading a text and completing the comprehension test in print (or digitally), participants answered the following question presented on a sheet of paper (or online):

Please rate your performance on the reading comprehension test you just completed in print (digitally) on a scale of 0 (comprehended nothing) to 100 (comprehended everything).

Calibration Accuracy. Given that calibration refers to the correspondence between predicted and actual performance (Alexander, 2013; Fischhoff et al., 1977), I descriptively and statistically examined the relation between judgment of performance (JOP) and demonstrated performance on the comprehension measure (CM) for each medium. For this analysis, I first converted JOP ratings and CM scores into percentages so they are on a similar metric. Next, I took the absolute value of the difference in percentages between the JOP rating and CM score for each participant for each medium, with a lower the calibration score representing higher calibration accuracy.

Processing Behavior Codes (PBC). In a previous investigation (Singer Trakhman et al., 2018a), processing behaviors were generated by conducting a pilot study where I and a trained research assistant examined videos of two participants reading in print and two reading digitally. While watching the videos, the research assistant and I independently recorded and observed processing behaviors. We then conferred on the documented behaviors and agreed upon a set of codes that encompassed all observable actions for both print and digital reading. The final processing behavior codes (PBC) used in that study were: *Read* (R); *Re-Read* (RR), *Re-Position* (RP); *Page-Turn* (PT, print only) or *Scroll* (SC, digital only); and *Skim* (S). However, the previous investigation used texts that had no visuals.

To determine the appropriate PBC's for the current texts that include visuals, I again conducted a pilot study in which I worked with a trained research assistant to examine videos of several volunteer undergraduates reading the passages in print and digitally. While watching these videos, we independently recorded any observed processing behaviors. I then met with the research assistant to discuss the documented behaviors and created a set of codes that encompasses all observable actions for reading of a text including visuals. Both myself and the research assistant observed the behaviors that had been document in the previous study, which were *Read* (R); *Re-Read* (RR), *Re-Position* (RP); *Page-Turn* (PT, print only) or *Scroll* (SC, digital only); and *Skim* (S). In addition to these codes, we both noticed four new processing behaviors, which were all related to the presence of visuals in the text. Participants would refer to the visual when cued by the text (e.g., "as seen in Figure 1.2, thunderstorms form through a complex process"). We named this code *Look at Visual When Cued* (CV). Next, we observed students reading captions of the visual when cued by the text. This code was named *Read Caption When Cued* (CC). In contrast, there were instances where a participant would refer to a

visual either when it appeared next on the page, or at random. This action was labeled *Refer to Visual* (RV), and if the caption below the visual was read that processing behavior was labeled as *Refer to Caption* (RC). The full list of real-time processing codes, their definitions, and if they were designed as a peripheral behavior or a behavior showing more depth can be found in Table 1. In coding real-time data, PBC's were scored at a specified time interval, established on the basis of pilot data, previous studies, and the suggestions of Bakeman and Gottman (1997). In the previous study, almost all observed actions occurred in cycles under 4 seconds. Therefore, three-second intervals will be used for data coding. In the previous study, there were very few occurrences (11 times across 22,213 coded intervals) of two behaviors (e.g., *Read* and *Scroll*) occurring within a three-second interval. My co-authors and I decided that behavior observed closest to the three-second mark would be recorded. The same decision rule was applied in this dissertation for any such occurrences. Using the recorded footage, coding of the behaviors will occur after participants finishes their session. Prior to scoring, training materials with simulated videos paralleling those for the actual study were developed. Once the second rater reached a scoring proficiency level of 90% or above on the training materials, the actual data were introduced. To establish interrater agreement, the first author and a second rater coded video data independently and compared scores. Interrater agreement for the full data set was 99.4%.

Reading Duration. To calculate reading time in print, I reviewed GoPro© and Camtasia Screen Record© Software footage after participants finished their session. I watched the video and began timing as soon as students positioned the pen on paper and began underlining the text.

Table 1

Processing Behavior Codes and Descriptions

Processing Behavior	Description	Level of Processing
Read (R)	Underlining text in linear order	Peripheral
Re-Position (RP)	Moving from one section of text to other in non-linear order	More depth than peripheral
Re-Read (RR)	(Only following re-positioning code) Underlining text that has already been read in linear order	More depth than peripheral
Page Turn (PT) /Scroll (SC)	(Print Only) Turning page forward/backwards (Digital Only) Moving forwards/backwards in text	Peripheral
Skim (S)	Progressing through text without underlining	Peripheral
Refer to Visual (RV)	Looking at a visual in the passage	More depth than peripheral
Refer to Caption (RC)	Reading the caption under the visual	More depth than peripheral
Look at Visual When Cued (CV)	When cued by the text (e.g., when the text says “see visual 2.1), looking at the visual in the passage	More depth than peripheral
Read Caption when Cued (CC)	When cued by the text (e.g., when the text says “see visual 2.1), reading the caption under the visual cued	More depth than peripheral

Timing was stopped when students lifted their pen from the page and indicated that they are done. Similarly, I initiated timing for digital reading as soon as students positioned their cursor on the screen and moved the cursor along as directed. Timing was terminated when students moved the cursor to the “next” arrow at the bottom of the screen.

Overview of the Analyses

To achieve the aforementioned goals, this investigation was designed to incorporate both variable-centered and person-centered analyses across two experimental studies. In Study I, preliminary analyses including descriptives were conducted. Next, to answer the first research question. variable-centered analyses were conducted using Multiple Analysis of Covariance for print and for digital reading conditions, while controlling for topic knowledge. For the person-centered analyses in research question two, a Latent Class Analysis was conducted for the print medium using processing behaviors, reading time, and calibration accuracy to identify clusters. This person-centered analysis is useful for identifying variation within a sample where individuals can be grouped with other similar individuals (Laursen & Hoff, 2006). Fit statistics were employed to decide upon the appropriate number of clusters to retain (Schwartz, 1978). The extent to which the groups have meaningful differences in comprehension outcomes were pursued to determine if the LCA is able to accurately create processing groups. Next, a second LCA was conducted for the digital data using the same variables and procedure as the print medium.

Procedure

Students were recruited from human development courses at the University of Maryland and were offered extra credit for participation. For this study, students met individually with the researcher in a designated testing area. The session lasted approximately 50 minutes. Upon entering the lab, the undergraduates received a consent form, which was presented digitally, along with the demographics survey and the topic knowledge measure. Next, they moved to the reading task, which involved the two passage on Soils and Weather presented either in print or digitally, with topic and medium counterbalanced. While reading in print, students followed

along by underlining text as they read and wore a GoPro© camera to record movements. When reading digitally, students used an enlarged cursor to indicate their movements through the text. These data were recorded using Camtasia Screen Record© Software.

Prior to receiving texts in print/digital form, participants were instructed as follows:

You are going to be given a passage to read and asked questions after. You may take as much time as you need with the text, but you will not be able to access the text while you answer the questions. As you read, make sure to "follow along" with your place in the text by using the pen to underline the words as you read them/by using your mouse to run below the words as you read them. If you re-read any text, make sure the pen/cursor follows and underline the text again. If you skip any text, do not underline it and simple move your pen/cursor to the place where you begin reading again and start underlining where you decide to begin reading. If you are looking at a visual, place your pen/cursor somewhere within the visual. If you are reading the caption of a visual, underline the text in the same manner as I explained previously. I am now going to demonstrate this for you using the directions paragraph on this paper/on your screen. Any questions?

When students began reading either in print or digitally, I monitored their actions and intervened if they stopped tracking their reading by prompting them to continue underlining as they read. After reading, they were handed the print comprehension test or clicked the "next" button on their computer screen to begin the digital comprehension test. When the comprehension test was completed, the students were prompted to judge how well they performed. Once they indicated their level of performance on the judgment of performance measure, the task for one medium was finished and they began the same procedure for the other medium. Finally, they were asked if they found their reading skills to be adequate for the

processing of the texts and were prompted to select “yes” or “no”. All of the participants selected “yes” as their response to this question.

Method: Study II

Participants

The primary purpose of Study II, as noted, was to examine the effects of an intervention related to students’ regulation of their speed and processing behaviors when reading digitally. Therefore, the target population for this study were the participants from Study I who were willing to return to the lab about two weeks after completing Study I. The sample was composed of as many participants as possible from Study I. Of the 70 participants in Study I, 65 returned to for the Study II session. Although 5 participants did not return for Study II, there were no discernable changes to the demographics presented in Study I.

Experimental Text

A passage on Volcanos from the same textbook as used in Study I was employed as the post-intervention experimental text (*Geology and the Environment;* Pipkin et al., 2014). The passage is similar in length (about 3 pages long) and readability level (grade 14; Fry Readability; Fry, 1968) to the texts presented in Study I. Similar to the passages in Study I, the passage on volcanos contains four visuals (1 diagram, 1 map, and 2 photos) which are comparable to the visuals used in Study I.

Digital Reading Intervention

The intervention consisted of five pre-recorded training modules delivered on computer that included narration and video clips. The decision to pre-record the modules was made to rule out any differences in delivery that might occur across participants. The training modules consisted of several sections in which general information on research findings pertinent to

reading in print and digitally were summarized and illustrated (e.g., student's reading preferences, differences between medium preferences and performance, speed of reading, judgment of performance, and different reader profiles). A large portion of the intervention focused on previous research on processing profiles, which emphasized the research on self-regulation, strategic processing, and reading time.

In the section addressing students' reading preferences, participants were asked to predict the medium in which they believe undergraduates prefer to read in. Their response affected the next page they see. If they selected "print" or "no preference" the module moved them to a slide that narrates, "That is incorrect. In a recent investigation of college-aged participants, 89% responded that they preferred to read text digitally." If they selected "digital," the module moved them to a slide that narrated, "That is correct! In a recent investigation of college-aged participants, 89% responded that they preferred to read text digitally."

The module then moved to the next slide, which detailed the research on the differences between preferences for reading digitally and actual performance. After each section, students were asked to respond to a question that dealt with the information covered that was followed by immediate feedback in a similar manner to the example previously described. The entire five modules took about 30 minutes to complete and is detailed in Appendix D.

Measures

Topic Knowledge Test. A knowledge measure covering the topic of volcanoes was created to parallel the topic knowledge measures in Study I was administered in Study II. An example of a question in the Volcano pretest follows:

Which type of rock has the lowest melting point?

a. Granitic and silicic metamorphic rocks

b. Rocks found within the mantle

c. Sandstone and Shale Sedimentary rocks

d. Diorite and Basalt Igneous rocks

Comprehension Measure (CM). Comprehension was assessed in the same format as in Study I. Students answered seven short-response questions digitally immediately after reading the associated passage. The first two questions asked about information found in the text portion of the passage only (e.g., Why does mafic material not melt when it is located in the Earth's mantle?), while the next two questions relied on information in the visuals only (e.g., According to the information visually presented in the passage, name three locations that are not included in the "Ring of Fire" that produce volcano clusters). The final three questions involved content included in both the text and visuals (e.g., What often occurs at the mid-ocean ridge system? What features cause this to happen?). The comprehension measure and answer key for Study II were verified in the same manner by the two science teachers as noted in Study I. The questions for this passage, along with the answer key, can be found in Appendix E.

Judgment of Performance (JOP). Students reported their Judgment of Performance rating in the same procedure as in Study I.

Calibration Accuracy. Calibration accuracy was calculated in the same procedure as in Study I.

Processing Behavior Codes (PBC). Processing behavior codes were measured in the same procedure as in Study I.

Reading Duration. Reading duration was assessed in the same procedure as in Study I.

Overview of the Analyses

Within Study II, both variable and person-centered analyses were again conducted. In Study II, preliminary analyses including descriptives were completed. For the third research question, a variable-centered analyses, a repeated measures ANOVA compared the students' digital comprehension scores and calibration accuracy following an intervention to their digital comprehension scores and calibration accuracy in Study I. For the fourth and final research question, which is a person-centered analysis, a LCA was conducted using the post-intervention processing behaviors, reading time, and calibration accuracy as the manifest variables to identify groups. In addition, changes in students' group membership for the digital condition from Study I to Study II was examined using descriptives and non-parametrics, similar to what was done by Alexander and Murphy in 1998.

Procedure

For Study II, students again met individually with the researcher in a designated testing area about 2 weeks after completing Study I. The session lasted approximately 60 minutes. Upon entering the lab, the undergraduates began the intervention with instructions to navigate through the pre-recorded training modules on the desktop computer. The five training modules consisted of several brief sections in which general information pertinent to reading in print and digitally will be summarize and illustrated (e.g., students' reading preferences, speed of reading, judgment of performance, and different reader profiles). After each section, students were prompted by the module to respond to a question that deals with the information covered that is followed by immediate feedback. The entire task took about 30 minutes to complete. After completing the intervention portion of this study, participants moved to the topic knowledge measure, which was taken digitally. Next, they were read the digital reading task directions and

then read the passage digitally and answered the comprehension and judgment of performance questions.

CHAPTER 4

RESULTS

In answering the four research questions that framed this two-study investigation, I sought to profile undergraduates' real-time reading behaviors and reading duration as they processed multimodal texts (Study 1) and then to intervene to improve on those outcomes (Study 2). Moreover, a variety of statistical methods were used in these two studies, including Multivariate Analysis of Covariance (MANCOVA), Latent Class Analysis (LCA), repeated measures ANOVA, Chi-Square Analysis and non-parametric tests. For Study 1, MANCOVAs were employed to examine the effects of real-time reading behaviors and reading duration on comprehension and calibration accuracy (Research Question 1), while Latent Class Analysis was conducted to determine what distinctive groups among the undergraduate participants emerged from real-time processing behaviors, reading duration, and calibration accuracy for print and digital mediums (Research Question 2). In Study II, a repeated measures ANOVA was run to compare students' digital comprehension scores and calibration accuracy following an intervention on their processing behaviors and reading duration when reading digitally to their digital comprehension scores and calibration accuracy documented in Study I (Research Question 3). Finally, I ran a Latent Class Analysis test to identify clusters based on post-intervention processing behaviors, reading duration, and calibration accuracy and a Chi-Square Analysis to compare students' group memberships to those documented in Study I (Research Question 4).

Research Question 1

How do students' real-time processing behaviors and reading times affect their comprehension performance and calibration accuracy when processing multimodal expository texts in print or digitally, after adjusting for demonstrated topic knowledge?

Descriptive Statistics

Descriptive data for Study 1 measures (e.g., Topic Knowledge tests) and subcomponents (e.g., Soil and Weather) are displayed in Table 2.

Topic Knowledge tests. As shown in Table 2, students scored an average of 1.57 (*SD* = 0.97) on the Soil pretest and 1.62 (*SD* = 1.24) on the Weather pretest, out of a possible six points. This difference between the Soil and Weather pretest was nonsignificant, $t(71) = -0.51$, $p = .61$, indicating that the undergraduates in this study were rather uninformed about the two topics covered in the text passages, as was expected. This nonsignificant difference allowed me to collapse the data for the two Topic Knowledge tests into a single score (0-12) for analysis.

Table 2

Study I Means and Standard Deviations for Measures and Subcomponents by Medium

Measure		Medium			
Subcomponent		Print		Digital	
	max	*M*	*SD*	*M*	*SD*
Topic Knowledge					
Soil	6	1.57	0.97	1.57	0.97
Weather	6	1.62	1.24	1.62	1.24
Processing Time		11m 52s**	2m 58s	10m 31s	2m 43s
Comprehension					
Total	20	13.56*	3.67	12.26	3.61
Text Only	5	4.36**	1.02	3.37	1.17
Visual Only	5	3.24	1.42	2.96	1.46
Text and Visual	10	5.94	2.34	5.60	2.24
Judgment of Performance	100	58.01	19.34	53.97	21.19
Calibration^{+}	0	17.89	12.95	15.83	11.18

Note: *=p<.05, **=p<.01
+Calibration represents the difference between predicted and actual comprehension performance, so the lower the number the better the calibration.

Comprehension tests. First, comprehension scores were examined by comparing scores from the Soil comprehension measure to the Weather comprehension measure. Students scored an average of 12.74 (*SD* = 3.35) on the Soil comprehension measure and an average of 13.08 (*SD* = 3.62) on the Weather comprehension measure. This difference between the Soil and Weather comprehensions tests scores was nonsignificant, $t(71) = -3.73$, $p = .59$, indicating that the comprehension performance was not significantly affected by the subject of the passage. This nonsignificant difference allowed me to collapse the data for the two passage subjects into a single score (0-20) for analysis.

Students' comprehension performance was also examined for print and digital medium and for question type (i.e., text only, visual only, text and visual). These descriptive data for both texts are displayed in Table 2. As these data show, when the students answered questions in the print condition that required information only from the text, their average score was 4.36 points (*SD* = 1.02) out of a possible 5 points. However, when those same questions were answered after reading digitally, the undergraduates scored an average of 3.37 (*SD* = 1.17). Similarly, when the comprehension questions could be answered with information found only in the visuals (5 points), students scored an average of 3.24 (*SD* = 1.42) in the print condition, but an average of 2.96 (*SD* = 1.46) in the digital condition.

A similar performance pattern emerged when the information required to answer the comprehension questions could be found *either* in the text *or* in the visuals. For these dual-source questions, students scored 5.94 (*SD* = 2.34) out of 10 possible points when they read in print, and an average of 5.60 (*SD* = 2.24) when they read digitally. Thus, for their performance across all three question types, the undergraduates in this study scored an average 13.56 (*SD* = 3.36) out of 20 points when they read the printed texts, and an average of 12.26 (*SD* = 3.61)

when they read those same texts digitally. Collectively, this consistent performance pattern suggests that comprehension of the texts was superior when that text was presented in print, no matter the information source (i.e., text only, visual only, or both).

Reading duration, judgment of performance (JOP), and calibration. When reading in print, students in this study spent an average of 11 minutes and 52 seconds (*SD* = 2min 58s) with the texts, whereas when reading digitally the mean duration was 10 minutes 31 seconds (*SD* = 2m 43s). Further, out of a possible 100 points, undergraduates judged their comprehension performance after reading the printed texts to be an average of 58.01 (*SD* = 19.34), as compared to 53.97 (*SD* = 21.19) when they read digitally. The fact that students rated their comprehension after reading printed texts to be higher than after reading digitally deviates from previous investigations of print and digital reading that I have conducted (Singer & Alexander, 2017a; Singer Trakhman et al., 2017; Singer Trakhman et al., 2018a). In those prior studies, students consistently predicted their performance to be better for the digital medium. Possible explanations for this anomalous outcome will be considered in Chapter 5.

Based on the difference between the comprehension scores that students predicted they would receive and the comprehension scores they actually received, I then calculated their calibration accuracy scores. Using this approach, the lower the score students receive, the better calibrated they are. In this study, the undergraduate's calibration score was 17.89 (*SD* = 12.95) for print and 15.83 (*SD* = 11.18) for digital, suggesting that participants were slightly better at judging their performance after reading digitally. In the first three studies that I conducted (Singer & Alexander, 2017a; Singer Trakhman et al., 2017; Singer Trakhman et al., 2018a), calibration accuracy was consistently better for print than for digital reading. The outcome for this study, however, follows the pattern documented in my most recent investigation (Singer

Trakhman et al., 2018b). Although more will be discussed about this finding in Chapter 5, the commonality between this current investigation and the preceding study was the inclusion of visuals.

Comprehension and Calibration Accuracy

Before looking at the effects of real-time processing behaviors and reading time on comprehension and calibration, I first tested whether topic knowledge would serve as an appropriate covariate for the MANCOVAs. To be an appropriate covariate, topic knowledge had to be significantly related to comprehension and calibration accuracy. As hypothesized, topic knowledge had a significant effect on both comprehension scores, [$F(1,41) = .53$, $p < .05$, $\eta^2 = .06$], and calibration accuracy, [$F(1,41) = 2.20$, $p < .05$, $\eta^2 = .04$]. Thus, undergraduates' processing behaviors and reading times were adjusted based on their level of knowledge about soil and weather.

Print condition. To answer Research Question 1, I coded the real-time processing behaviors described in Chapter 3 (e.g., *Read, Refer to Visual*) that were captured on the Go-Pro© footage as participants were reading the printed texts. This coding occurred at three second intervals. The resulting data are displayed in Table 3 as percentages of processing behaviors overall and by medium. Additionally, reading duration, as documented on the Go-Pro© footage, was recorded. The average durations by medium and combined are displayed in Table 2.

Assumptions. Before beginning analysis, the print data were screened and cleaned for univariate and multivariate outliers, and the normality of distributions was accessed. In the data screening process, one univariate outlier with standardized z-score higher than +3.29 was detected and removed from further analysis. No multivariate outliers were detected. In total, 1.43% of the data were removed from the sample prior to analysis. Explorative analysis also

Table 3

Percentages of Processing Behaviors by Medium and Combined for Study I

Behavior	Medium		
	Combined *M (SD)*	Print *M (SD)*	Digital *M (SD)*
Read	81.00 (10.74)	79.77 (3.81)	81.23 (14.44)
Re-Read	5.78 (7.37)	6.41 (7.79)	5.12 (6.91)
Re-Position	1.36 (3.00)	1.61 (3.27)	1.11 (0.27)
Page Turn/ Scroll	2.00 (1.45)	1.20 (0.68)	2.28 (1.56)
Skim	1.34 (3.93)	1.75 (4.15)	0.93 (3.68)
Refer to Visual	3.68 (3.01)	3.73 (3.16)	3.62 (2.89)
Refer to Caption	3.57 (3.23)	4.15 (3.48)	2.93 (0.29)
Look at Visual When Cued	0.72 (0.16)	0.92 (0.17)	0.54 (1.48)
Read Caption When Cued	0.57 (1.22)	0.71 (1.32)	0.42 (1.11)

showed that dependent variables and the covariate were normally distributed. Both skewness and kurtosis on all three variables were in minus one–plus one range. Scatterplot analysis showed that dependent variables were linear related with the covariate. Custom model analysis showed that the topic knowledge covariate did not significantly interact with any independent variable, indicating that the assumption of homogeneity of regression slopes was met. Levene's test showed that the assumption of homogeneity of variance was met for comprehension performance [$F(58,10) = 1.09$, $p < .05$, $\eta^2 = .03$] and calibration accuracy [$F(58,10) = 1.42$, $p < .05$, $\eta^2 = .01$]. Overall, these analyses suggested that all assumptions for conducting a MANCOVA had been met.

Total comprehension (print). None of the processing behaviors included in the MANCOVA significantly affected students' total comprehension performance, which was the sum of the three question types. *Read* was nonsignificant [$F(3,41) = .91$, $p = .41$], as was *Re-Read* [$F(3,41) = .13$, $p = .94$], *Re-Position* [$F(3,41) = .25$, $p = .86$], *Page Turn* [$F(3,41) = .73$, $p = .54$], *Skim* [$F(3,41) = 2.54$, $p = .07$], *Refer to Visual* [$F(3,41) = .65$, $p = .53$], *Refer to Caption* [$F(3,41) = .13$, $p = .88$], *Look at Visual When Cued* [$F(3,41) = 2.14$, $p = .15$], and *Read Caption When Cued* [$F(3,41) = .23$, $p = .64$]. Further, reading duration had no significant effect on total comprehension scores, [$F(3,41) = 2.24$, $p = .08$]. The results of MANCOVA for participants reading in print suggest that none of the independent variables significantly influenced total comprehension performance even after adjusting for topic knowledge.

Text only questions (print). For the comprehension questions requiring information from information found within the text only, two of the processing behaviors significantly affected comprehension of the text only questions. These two processing behaviors were *Refer to Visual*, which is when the reader looked as a visual in the passage, [$F(3,41) = 5.35$, $p < .05$, $\eta^2 = .02$], and *Read Caption When Cued*, which is when the reader followed along with their pen while reading the caption under a visual when the text cued the reader to do so (e.g., when the text says "see visual 2.1"), [$F(3,41) = 4.39$, $p < .05$, $\eta^2 = .01$]. It is interesting that *Refer to Visual* and *Read Caption When Cued* were the two variables that significantly affected comprehension of text only questions involved viewing the visuals in text. The seven remaining processing behaviors were found to have no significant effect on text only comprehension, *Read* [$F(3,41) = 2.22$, $p = .14$], *Re-Read* [$F(3,41) = 3.22$, $p = .07$], *Re-Position* [$F(3,41) = 1.38$, $p = .24$], *Skim* [$F(3,41) = 1.76$, $p = .19$], *Page Turn* [$F(3,41) = .62$, $p = .43$], *Refer to Caption* [$F(3,41) = .45$, $p = .50$], and *Look at Visual When Cued* [$F(3,41) = .62$ $p = .43$]. Reading duration was also found

to have no significant effect on students' responses to the text only questions, [$F(3,41) = 2.61$, $p = .11$].

Visual only questions (print). When the comprehension questions only required information from the visuals in the passage, there was only one text processing behavior that had a significant effect on performance. Considering the nature of the visual only comprehension questions, it is not surprising that this processing behavior was *Refer to Visual* [$F(3,41) = 5.35$ $p < .05$, $\eta^2 = .12$], which is the code for when the reader looked at a visual in the passage. *Read* [$F(3,41) = .34$ $p = .33$], *Re-Read* [$F(3,41) = .01$, $p = .61$], *Re-Position* [$F(3,41) = 1.03$, $p = .38$], *Skim* [$F(3,41) = .48$, $p = .55$], *Page Turn* [$F(3,41) = .13$, $p = .76$], *Refer to Caption* [$F(3,41) = .93$ $p = .41$], *Look at Visual When Cued* [$F(3,41) = .12$, $p = .76$], and *Read Caption When Cued* [$F(3,41) = .003$, $p = .96$] had no significant effect on students' performance on the visual only comprehension question. Similarly, reading duration was found to have no significant effect on performance of the visual only questions, [$F(3,41) = 2.92$, $p = .14$].

Text and visual questions (print). For the comprehension questions requiring knowledge of information found across the text and visuals, none of the processing behaviors had a significant effect on students' performance. Specifically, *Read* [$F(3,41) = 6.11$, $p = .20$], *Re-Read* [$F(3,41) = 2.27$, $p = .44$], *Re-Position* [$F(3,41) = 7.95$, $p = .14$], *Skim* [$F(3,41) = 6.09$, $p = .20$], *Page Turn* [$F(3,41) = .22$, $p = .81$], *Refer to Visual* [$F(3,41) = .01$, $p = .95$], *Refer to Caption* [$F(3,41) = 2.24$ $p = .44$], *Look at Visual When Cued* [$F(3,41) = 5.52$ $p = .23$], and *Read Caption When Cued* [$F(3,41) = 6.50$, $p = .18$] were found to have no significant effect on visual only questions. Reading duration was also found to have no significant effect on comprehension scores for text and visual questions, [$F(3,41) = 14.42$, $p = .05$].

Calibration accuracy scores (print). Even after controlling for topic knowledge, calibration accuracy scores were not significantly affected by any of the real-time processing behaviors, *Read* [$F(3,41) = 1.00$, $p = .40$], *Re-Read* [$F(3,41) = 1.16$, $p = .34$], *Re-Position* [$F(3,41) = .31$, $p = .82$], *Page Turn* [$F(3,41) = .93$, $p = .44$], *Skim* [$F(3,41) = 1.21$, $p = .32$], *Refer to Visual* [$F(3,41) = 2.76$, $p = .08$], *Refer to Caption* [$F(3,41) = .44$, $p = .65$], *Look at Visual When Cued* [$F(3,41) = .21$, $p = .65$], and *Read Caption When Cued* [$F(3,41) = 0$, $p = .95$]. Likewise, reading duration also had no significant effect on calibration accuracy scores, [$F(3,41) = .81$, $p = .53$]. In sum, the results of this MANCOVA analysis showed that after adjusting for demonstrated topic knowledge, students' real-time processing behaviors and reading duration do not significantly affect their calibration accuracy when processing multimodal expository texts in print.

Digital condition. To answer Research Question 1, I next coded the real-time processing behaviors described in Chapter 3 (e.g., *Read*, *Refer to Visual*) captured on the Camtasia Screen Capture Record© software footage when participants were reading digitally. This coding occurred at three second intervals. The resulting data are displayed in Table 3 as percentages of processing behaviors overall and by medium. Additionally, reading duration as documented on Camtasia Screen Capture Record© Software was recorded. The average reading durations by medium and combined are displayed in Table 2.

Assumptions. Similar to the procedure for conducting the analyses for print scores, the data for digital reading scores first had to be screened before conducting the MANCOVA. In the data screening process one univariate outlier with standardized z-score higher than +3.29 was detected and removed from further analysis. Also, one multivariate outlier was detected. In total, 2.86% of the data were removed from the sample before the main analysis. Explorative analysis

showed that dependent variables and the covariate were normally distributed. Both skewness and kurtosis on all three variables was in minus one–plus one range. Scatterplot analysis showed that dependent variables were in linear relationship with the covariate. Custom model analysis showed that topic knowledge covariate did not significantly interact with any independent variable, indicating that the assumption of homogeneity of regression slopes was met. Levene's test showed that the assumption of homogeneity of variance was met for comprehension performance [$F(60,7) = 1.39$, $p < .05$, $\eta^2 = .05$] and calibration accuracy [$F(60,7) = .83$, $p < .05$, $\eta^2 = .03$]. These analyses confirmed that the data met all assumptions for conducting a MANCOVA.

Total comprehension sores (digital). For total comprehension scores, two of the processing behaviors significantly influenced total comprehension scores, *Refer to Visual* [$F(3,41) = 5.77$, $p < .05$, $\eta^2 = .06$], and *Read Caption When Cued* [$F(3,41) = 5.39$, $p < .05$, $\eta^2 = .05$]. The remaining processing behaviors did not influence total comprehension scores, *Read* [$F(3,41) = 1.70$, $p = .20$], *Re-Read* [$F(3,41) = .34$ $p = .54$], *Re-Position* [$F(3,41) = 1.26$, $p = .30$], *Scroll* [$F(3,41) = .50$, $p = .68$], *Skim* [$F(3,41) = .57$, $p = .64$], *Refer to Caption* [$F(3,41) = .02$, $p = .98$], and *Look at Visual When Cued* [$F(3,41) = 2.14$, $p = .13$]. Additionally, reading duration was also found to have no significant effect on total comprehension scores, [$F(3,41) = .69$, $p = .60$]. Overall, results of the MANCOVA analysis indicated that while two processing behaviors significantly affected total comprehension scores, the majority of variables did not play a significant role. However, it is important to note that the two variables were found to have a significant effect were both related to accessing the visuals in the text passage. This suggests that viewing a passage in its entirety (e.g., reading both the text and visuals) can contribute to successful comprehension.

Text only questions (digital). For the comprehension questions requiring information from information found within the text only in the passage, none of the processing behaviors were found to have a significant effect on text only comprehension scores, *Read* [$F(3,41) = .24$, $p = .65$], *Re-Read* [$F(3,41) = .59$, $p = .48$], *Re-Position* [$F(3,41) = 1.70$, $p = .97$], *Skim* [$F(3,41) = 1.02$, $p = .25$], *Scroll* [$F(3,41) = .41$, $p = .55$], *Refer to Visual* [$F(3,41) = .17$, $p = .71$], *Refer to Caption* [$F(3,41) = .35$, $p = .58$], *Look at Visual When Cued* [$F(3,41) = 5.18$ p=.05], and *Read Caption When Cued* [$F(3,41) = .67$, $p = .45$]. However, reading duration did have a significant effect on comprehension of the text only questions, [$F(3,41) = 4.08$, $p < .05$, $\eta^2 = .07$]. This means that the speed at which participants navigated the digital text had a significant effect on their comprehension of the information found within the text only.

Visual only questions (digital). For the questions requiring information found in the visuals only, three processing behaviors had significant effects on the corresponding comprehension scores, *Read* [$F(3,41) = 8.84$, $p < .05$, $\eta^2 = .02$], *Skim* [$F(3,41) = 8.77$, $p < .05$, $\eta^2 = .03$], and *Look at Visual When Cued*, [$F(3,41) = 5.17$, $p < .05$, $\eta^2 = .08$]. These significant effect of these three processing behaviors suggests that the way in which participants navigated the digital passage did in fact have an impact on comprehension of the visuals within the text passage. The remaining processing behaviors, *Re-Read* [$F(3,41) = 3.72$, $p = .12$], *Re-Position* [$F(3,41) = 2.78$, $p = .18$], *Scroll* [$F(3,41) = .27$, $p = .68$], *Refer to Visual [*$F(3,41) = .04$, $p = .88$], *Refer to Caption* [$F(3,41) = 1.93$ $p = .27$], and *Read Caption When Cued* [$F(3,41) = .82$, $p = .47$] had no significant effect on visual only comprehension scores. Similarly, reading duration had no significant effect on comprehension scores for the visual only questions, [$F(3,41) = 1.58$ $p = .31$]

Text and visual questions (digital). For the comprehension questions requiring knowledge of information found across the text and visuals, six of the processing behaviors had a significant effect on the comprehension scores for this question type. These processing behaviors were *Read* [$F(3,41) = 73.25$, $p < .05$, $\eta^2 = .01$], *Re-Read* [$F(3,41) = 39.90$, $p < .05$, $\eta^2 = .07$], *Scroll* [$F(3,41) = 60.87$, $p < .01$, $\eta^2 = .02$], *Refer to Caption* [$F(3,41) = 13.85$, $p < .05$, $\eta^2 = .11$], *Look at Visual When Cued* [$F(3,41) = .66$, $p < .05$, $\eta^2 = .13$], and *Read Caption When Cued* [$F(3,41) = 1.59$, $p < .05$, $\eta^2 = .13$]. This suggests that a majority of the real-time processing behaviors played a significant part in text and visual comprehension scores. Despite this, the following processing behaviors did not have a significant effect on this outcome, *Re-Position* [$F(3,41) = 5.56$, $p = .21$], *Scroll* [$F(3,41) = 6.87$, $p = .17$], and *Refer to Visual* [$F(3,41) = 4.79$, $p = .24$]. Reading duration was also found to have no significant effect on the comprehension questions requiring knowledge of information found across the text and visuals, [$F(3,41) = 1.10$, $p = .57$].

Calibration accuracy scores (digital). For calibration accuracy, the results of the MANCOVA showed that only one processing behavior, *Read Caption when Cued* [$F(3,41) = 4.97$, $p < .05$, $\eta^2 = .05$] had a significant effect on calibration accuracy scores. The remaining processing behaviors did not significantly influence calibration accuracy, *Read* [$F(3,41) = .34$, $p = .72$], *Re-Read* [$F(3,41) = .85$, $p = .43$], *Re-Position* [$F(3,41) = 1.07$, $p = .37$], *Scroll* [$F(3,41) = 1.34$, $p = .28$], *Skim* [$F(3,41) = 2.08$, $p = .12$], *Refer to Visual* [$F(3,41) = .42$, $p = .52$], *Refer to Caption* [$F(3,41) = .27$, $p = .76$], and *Look at Visual When Cued* [$F(3,41) = 2.75$, $p = .08$]. Reading duration also had no significant effect on calibration accuracy scores, [$F(3,41) = .30$, $p = .88$]. Although only one variable had a significant effect on calibration accuracy scores for the

digital medium, it is interesting to note that the act of referring to a caption when cued by the text appears to make students better attuned to their own performance.

Summary

MANCOVAs were employed to examine the effects of real-time reading behaviors and reading duration on comprehension and calibration accuracy. For the print data, the findings from the MANCOVAs suggest that, *Refer to Visual and Read Caption When Cued* significantly affected comprehension of text only questions. When the comprehension questions only required information from the visuals in the passage, there was only one text processing behavior that had a significant effect on performance, *Refer to Visual*. For the digital data, reading duration did have a significant effect on comprehension of the text only questions. For the questions requiring information found in the visuals only, three processing behaviors had significant effects on the corresponding comprehension scores, *Read*, *Skim*, and *Look at Visual When Cued*. For the comprehension questions requiring knowledge of information found across the text and visuals, six of the processing behaviors had a significant effect on the comprehension scores for this question type (*Read, Re-Read, Scroll*, *Refer to Caption*, *Look at Visual When Cued*, and *Read Caption When Cued*). Only one processing behavior, *Read Caption when Cued,* had a significant effect on calibration accuracy scores.

Research Question 2

What verifiable clusters emerge from real-time processing behaviors, reading time, and calibration accuracy for undergraduates' reading of print and digital texts? Are there significant differences in comprehension performance for the resulting clusters for print and digital text processing?

In order to address Research Question 2, I first set out to examine what clusters emerged from real-time processing behaviors, reading time, and calibration accuracy after reading in print and digitally by conducting a Latent Class Analysis (LCA). Subsequently, I assessed whether these resulting clusters differed in terms of their comprehension performance by conducting Multiple Analyses of Variance (MANOVA) and post-hoc analyses using Tukey HSD. The findings for those analyses are first detailed for the print reading condition and then for the digital reading condition. First, let me briefly describe the LCA procedure.

Latent Class Analysis

LCA is a multivariate method designed to identify unobserved (or latent) subpopulations of individuals based on multiple observed measures (Lubke & Muthén, 2005). LCA uses maximum likelihood estimation to fit a hypothesized model in which membership in a specified number of latent classes is related to performance on the included measures. A latent class can be interpreted as a subpopulation of students sharing important similarities on the multiple observed indicators included in the analysis, such as processing behaviors, reading times, and calibration accuracy in this dissertation. Moreover, each latent class can be meaningfully distinguished from other classes that emerged from the analysis as a result of different performance patterns for those observed measures. It is also expected that the latent classes identified through LCA will be theoretically justifiable.

For this investigation, LCA was conducted in R using the poLCA function, which was selected for its ability to estimate latent class and latent class regression models for polytomous outcome variables like those in this study. The poLCA function with R evaluates the LCA model 10 times with changing start parameters. The model-fitting process begins with a one-class model (i.e., a model in which all readers are hypothesized to cluster into one group), with

additional classes added sequentially. Finally, the LCA reports statistical tests at each step in order to determine if the additional classes added significantly improve the goodness of fit of the model.

In LCA, the final judgment as to the model and associated classes to be retained is based on fit indices: log-likelihood (LL), the Bayesian information content (BIC), and the Akaike's information criterion (AIC). Each fit statistic is an indication of the degree to which a particular model fits the data well, with a lower estimate equating to a better model fit (Lazarsfeld & Henry, 1968). Of the three indices, it is recommended that log-likelihood should not be the primary fit statistic because it is unduly affected by smaller sample sizes like the one in this investigation. Similarly, when there are relatively few parameters used in construction the resulting groups, as in the present study, the AIC has the tendency to overidentify viable clusters (Dayton, 1998). Thus, given the number of parameters and sample size for this investigation, the BIC is the fit statistic that should be most informative (Lazarsfeld & Henry, 1968), although the theoretical interpretability and viability of the resultant classes must also be carefully considered (Lubke & Muthén, 2005).

Print Cluster Models

When the data for the print reading condition were submitted to LCA, five potential models emerged (see Table 4). The fit statistics for those five models were then inspected, with greater weight given to the BIC statistic due to the relatively small sample size. As shown in Table 4, the two-cluster model for print reading had the best fit based on the BIC. However, the difference in the BIC between the two-cluster and three-cluster models was only .01.

Table 4

Fit Statistics for Print and Digital Exploratory Latent-Class Models

Number of Clusters	Fit Statistics		
	LL	AIC	BIC
	Print		
1	-452.64	927.29	952.02
2	-427.85	901.70	953.42
3	-407.36	884.73	953.43
4	-391.82	877.65	983.32
5	-381.25	880.49	1013.57
	Digital		
1	-387.87	797.74	822.47
2	-359.70	765.40	817.12
3	-344.06	758.12	836.82
4	-334.85	763.70	869.38
5	-327.36	772.71	905.38

Note. LL, log-likelihood; AIC, Akaike's information criterion; BIC, Bayesian information criterion

Further, upon examining the memberships for the two-cluster model, their composition did not seem theoretically viable. Specifically, one of the two groups that emerged in the two-cluster model contained 61 of the 70 participants. In addition, the performance patterns for the manifest variables for this group were not particularly discernible. For example, in terms of duration, some of the members of this cluster had the longest reading durations of the entire population (24m 45s), while others had the shortest (7m 28s). Similarly, calibration accuracy scores in this cluster ranged from the most accurate (0) to the least accurate (54). Also, some members of this 61-person cluster had real-time processing behaviors that suggested only surface-level reading (e.g., *Skim* and *Page Turn*). In contrast, other members of this large cluster displayed relatively frequent use of behaviors showing more depth when compared to the more peripheral behaviors (e.g., *Refer to Caption When Cued* and *Re-Read*).

By comparison, the distribution of participants in groups for the three-cluster model was more balanced, and patterns of performance were more theoretically justifiable. For example, the membership numbers for the three clusters were 12, 18, and 40 individuals, respectively. Further, the real-time processing behaviors, reading time, and calibration accuracy scores for cluster members were more closely aligned. Thus, considering all of the fit statistics and theoretical viability, the three-cluster model was adopted for the print data.

With the number of clusters decided, I examined whether these resulting clusters differed from each other both for overall comprehension performance and for the specific question types (i.e., text only, visual only, text and visual). The results of the MANOVA with total score as the dependent measure confirmed a significant difference in overall comprehension for the three clusters [F (67, 2) = 4.77, $p < .01$, $\eta^2 = .13$]. Post-hoc analyses using Tukey HSD tests revealed that the source of that difference was between the first and third clusters and the second and third clusters (Cluster 1: $M = 14.33$, $SD = 2.96$; Cluster 2: $M = 15.00$, $SD = 3.05$; Cluster 3: $M = 12.90$, $SD = 2.10$). Also, when comprehension questions required information found only in the passage visuals, the performance of the three clusters again differed significantly [F (67,2) = 3.93, $p < .05$, $\eta^2 = .12$]. Post hoc analyses determined that the source of that difference was again between the second and third clusters (Cluster 2: $M = 3.94$, $SD = 1.11$; Cluster 3: $M = 2.90$, $SD = 1.58$). In contrast, for the questions based solely on information in the text, and for questions where relevant information was located in both the text and visuals, there were no significant differences for the clusters, [text only: $F(67,2) = 0.86$, $p = 0.92$; and text and visuals: $F(67,2) = 1.67$, $p = .19$]. More detailed descriptions of the three clusters for the print condition are provided in the subsequent sections.

Cluster 1. The first cluster identified for the print data consisted of 12 participants. A breakdown of their means and standard deviations for reading time and calibration accuracy by cluster are presented in Table 5, and mean and standard deviations for real-time processing behaviors can be found in Table 6. In terms of reading duration, participants grouped into Cluster 1 spent more time on the printed text than any other cluster (Cluster 1: M = 13m 21s, SD = 3m 46s; Cluster 2: M = 12m 48s, SD = 3m 41s; Cluster 3: M = 10m 50s, SD = 1m 55s). The difference in duration between Cluster 1 and Cluster 3 was statistically significant, [F (67, 2) = 5.10, $p < .05$, $\eta^2 = .04$]. Individuals in Cluster 1 were also more highly calibrated than those in Clusters 2 or 3 (Cluster 1: M = 13.25, SD = 14.10; Cluster 2: M =18.28, SD = 9.02; Cluster 3: M = 19.10, SD = 18.95). Likewise, when compared to Clusters 2 and 3, individuals belonging to Cluster 1 spent more of their time engaging in behaviors that are indicative of relative deeper processing when it came to the visuals in the passage, such as looking at visual when cued by the text. Based on these data, Cluster 1 could be best characterized as highly capable and well calibrated readers.

Table 5

Study I Means for Measures by Clusters and Medium

Measure Subcomponent	Clusters			
	Combined *M* (*SD*)	1 *M* (*SD*)	2 *M* (*SD*)	3 *M* (*SD*)
		Print		
n	70	12	18	40
Comprehension				
Total	13.56 (3.67)	14.33[a] (2.96)	15.00[a] (3.05)	12.90[b] (2.10)
Text Only	4.36 (1.02)	4.33 (0.98)	4.44 (0.92)	4.32 (1.10)
Visual Only	3.24 (1.42)	3.58 (0.79)	3.94[a] (1.11)	2.90[b] (1.58)
Text and Visual	5.94 (2.34)	6.50 (2.15)	6.61 (1.94)	5.58 (2.38)
Duration	11m 52s (2m 58s)	13m 21s[a] (3m 46s)	12m 38s[a] (3m 41s)	10m 50s[b] (1m 55s)
Judgment of Performance	58.01 (19.34)	67.08 (17.41)	62.83 (18.18)	53.13 (19.27)
Calibration	17.89 (12.95)	13.25 (14.10)	18.28 (9.02)	19.10 (18.95)
		Digital		
n	70	12	8	50
Comprehension				
Total	12.26 (3.61)	14.42[a] (2.75)	13.13 (2.75)	11.60[b] (3.73)
Text Only	3.37 (1.17)	4.00 (1.35)	3.75 (1.16)	3.36 (1.14)
Visual Only	2.96 (1.46)	3.75[a] (1.14)	3.50 (0.76)	2.68[b] (1.53)
Text and Visual	5.60 (2.24)	6.67 (1.44)	5.66 (1.36)	5.30 (2.44)
Duration	10m 31s (2m 43s)	11m 31s (3m 50s)	11m 54s (1m 49s)	10m 15s (2m 12s)
Judgment of Performance	53.97 (21.19)	59.75 (17.82)	58.63 (13.09)	51.84 (22.81)
Calibration	15.83 (11.18)	14.83 (10.90)	16.00 (11.22)	16.09 (11.37)

Note. [a] and [b] represent significant differences between clusters, $p < .05$.

Table 6

Study I Processing Behavior Percentage Means by Clusters and Medium

Processing Behaviors	Clusters			
	Combined *M* (*SD*)	1 *M* (*SD*)	2 *M* (*SD*)	3 *M* (*SD*)
		Print		
Read	79.77 (3.81)	78.00 (7.27)	66.65 (8.63)	86.20 (5.88)
Re-Read	6.41 (7.79)	6.61 (5.33)	11.16 (5.73)	4.22 (4.77)
Re-Position	1.61 (3.27)	1.25 (0.89)	3.96 (5.73)	0.67 (0.19)
Page Turn	1.20 (0.68)	1.28 (0.76)	1.41 (0.83)	1.07 (0.56)
Skim	1.75 (4.15)	0.89 (2.45)	4.51 (6.82)	3.72 (3.00)
Refer to Visual	3.73 (3.16)	2.89 (4.86)	5.83 (3.06)	3.04 (2.06)
Refer to Caption	4.15 (3.48)	1.32 (1.52)	7.01 (3.51)	3.72 (0.30)
Look at Visual When Cued	0.92 (0.17)	4.24 (1.41)	0.12 (0.30)	0.28 (0.71)
Read Caption When Cued	0.71 (1.32)	3.14 (1.26)	0.18 (0.53)	0.23 (0.57)
		Digital		
Read	81.23 (14.44)	70.48 (5.27)	78.82 (10.21)	84.19 (15.31)
Re-Read	5.12 (6.91)	15.04 (7.04)	7.50 (9.43)	2.36 (3.15)
Re-Position	1.11 (0.27)	0.84 (1.21)	0.62 (0.32)	1.30 (4.13)
Scroll	2.28 (1.56)	2.78 (1.22)	3.23 (1.47)	3.54 (2.45)
Skim	0.93 (3.68)	0.00 (0.00)	0.00 (0.00)	2.77 (1.66)
Refer to Visual	3.62 (2.89)	5.59 (3.77)	1.16 (2.05)	2.54 (2.45)
Refer to Caption	2.93 (0.29)	4.03 (2.65)	0.93 (1.40)	2.98 (2.97)
Look at Visual When Cued	0.54 (1.48)	4.25 (1.89)	0.19 (0.35)	0.03 (0.13)
Read Caption When Cued	0.42 (1.11)	2.19 (1.25)	0.20 (0.50)	0.04 (0.19)

Cluster 2. The second cluster for the print data was made up of 18 participants. While participants in Cluster 2 spent less time reading than their peers in Cluster 1, they did read for a statistically significant longer duration than members of Cluster 3, F (67, 2) = 5.10, $p < .05$, $\eta^2 = .03$. What is interesting about this cluster is that they engaged in some processing behaviors indicative of relative deeper processing at percentages higher than the other clusters (e.g., *Re-Read* and *Re-Position*). However, this picture contrasts with the pattern that emerged when this cluster dealt with the visuals in text. Specifically, although Cluster 2 members evidenced some attention to the visuals in the reading, they rarely did so when directed to by the text (e.g., "As shown in…."). Cluster 2 would be labeled as capable but less multimedia savvy print readers.

Cluster 3. The third cluster for the print data consisted of the majority of participants (n = 40). On average, this cluster spent less time reading the printed text than any other cluster (Cluster 3: M = 10m 50s, SD = 1m 55s; Cluster 2: M = 12m 48s, SD = 3m 41s; Cluster 1: M = 13m 21s, SD = 3m 46s). Members of Cluster 3 had calibration accuracy scores which were the poorest among clusters, (Cluster 3: M = 19.10, SD = 19.85; Cluster 2: M = 18.28, SD = 9.02; Cluster 1: M = 13.25, SD = 14.10). As noted, this finding is not surprising given that members of Cluster 3 spent the least amount of time with the text, F (67, 2) = 5.10, $p < .05$, $\eta^2 = .03$. Based on these data, Cluster 3 is considered the low performing and poorly calibrated print readers.

Digital Cluster Models

When the data for the digital reading condition were submitted to LCA, five potential clusters emerged (see Table 4). Similar to the process for the print data, the fit statistics (i.e., LL, AIC, and BIC) were examined with the most attention given to the BIC statistic due to the relatively small sample size in this study. Once again similarly to the fit statistics for the print data, the two-cluster model had the best fit in terms of BIC (Table 4). However, the two-cluster model sorted the participants into groups that did not make theoretical sense. Specifically, one of the two groups that emerged in the two-cluster model contained 64 of the 70 participants. In addition, similar to the two-cluster model for the print data, the performance patterns for the manifest variables for this cluster were not particularly discernible. For example, in terms of duration, some of the members of this cluster had the longest reading durations of the entire population (22m 13s), while others had the shortest (5m 11s).

Similarly, calibration accuracy scores in this cluster ranged from the most accurate (0) to the least accurate (40). Further, some members of this 64-person cluster had real-time

processing behaviors suggestive of surface-level reading (e.g., *Skim,* and *Scroll*). In contrast, other members of this large cluster displayed relatively frequent use of relative deeper processing behaviors (e.g., *Refer to Caption When Cued* and *Re-Read*). With this in mind, I re-examined the model options. The three-cluster model had a higher BIC than the two cluster-model, but it was still smaller than the other available model options. Also, the distribution of participants in groups for the three-cluster model, and patterns of performance were more theoretically justifiable. For example, the membership numbers for the three clusters were more evenly distributed at 12, 8, and 50 individuals, respectively. Moreover, the real-time processing behaviors, reading time, and calibration accuracy scores for cluster members were more closely aligned. Thus, considering all of the fit statistics and theoretical viability, the three-cluster model solution was adopted for the digital data.

With the number of clusters decided for the digital reading data, I examined whether these resulting clusters differed from each other both for overall comprehension performance and for the specific question types (i.e., text only, visual only, text and visual). The results of the MANOVA with total score as the dependent measure confirmed a significant difference in overall comprehension for the three clusters [F (67, 2) = 3.43, $p < .05$, $\eta^2 = .14$]. Post-hoc analyses using Tukey HSD tests revealed that the source of that difference was between the first and third clusters (Cluster 1: M = 14.42, SD = 2.75; Cluster 3: M = 11.60, SD = 3.73). Also, when comprehension questions required information found only in the passage visuals, the performance of the three clusters again differed significantly [F (67,2) = 3.46, $p < .05$, $\eta^2 = .11$]. Post hoc analyses determined that the source of that difference was again between the first and third clusters (Cluster 1: M = 3.75, SD = 1.14; Cluster 3: M = 2.68, SD =1.53). In contrast, for the questions based solely on information in the text, and for questions where relevant

information was located in both the text and visuals, there were no significant effects for clusters, [text only: $F(67,2) = 1.61$, $p = 0.21$; text and visuals: $F(67,2) = 1.87$, $p = .16$]. More detailed descriptions of the three clusters for the digital condition are provided in the subsequent sections.

Cluster 1. The first cluster identified for the digital data consisted of 12 participants. A complete breakdown of their mean and standard deviations for reading time and calibration accuracy by cluster are presented in Table 5, and mean and standard deviations for real-time processing behaviors can be found in Table 6. Individuals in Cluster 1 were more highly calibrated than those in Clusters 2 or 3 ($M = 14.83$, $SD = 10.90$; $M = 16.00$, $SD = 11.22$; and $M = 16.09$, $SD = 11.37$, respectively). Also, individuals belonging to Cluster 1 spent more of their time engaging in behaviors that are indicative of relative deeper processing when it came to the visuals in the passage, such as *Look at Visual When Cued* when compared to the entire digital population. With this in mind, Cluster 1 could be best characterized as highly capable and calibrated digital readers.

Cluster 2. The second cluster for the digital data consisted of eight participants. In terms of reading duration, participants grouped into Cluster 2 spent more time with the digital text than any other cluster (Cluster 1: $M = 11$m 31s, $SD = 3$m 50s; Cluster 2: $M = 11$m 54s, $SD = 1$m 49s; Cluster 3: $M = 10$m 15s, $SD = 2$m 43s). Overall, individuals in Cluster 2 engaged in real-time processing behaviors that are suggestive of relative deeper-processing at a percentage higher than the mean of the entire digital population. But, the mean percentages for two processing behaviors that are indicative of relative deeper processing behaviors, *Look at Visual When Cued* and *Refer to Caption When Cued,* were much lower than the Clusters 1 and 3 (see Table 6). With these findings in mind, Cluster 2 is considered the capable but less multimedia savvy digital readers.

Cluster 3. The third cluster for the digital condition consisted of the majority of participants (n = 50). Individuals in Cluster 3 spent less time reading the digital text than the other two clusters (Cluster 3: M = 10m 15s, SD = 2m 43s; Cluster 2: M = 11m 54s, SD= 1m 49s; Cluster 1: M = 11m 31s, SD = 3m 50s). Further, calibration accuracy scores for members of Cluster 3 were the poorest of any cluster, (Cluster 3: M = 16.09, SD = 11.37; Cluster 2: M = 16.00, SD = 11.22; Cluster 1: M = 14.83, SD = 10.90). This finding is not surprising given that members of Cluster 3 spent the least amount of time with the text. Based on these data, Cluster 3 is classified as the low performing and poorly calibrated digital readers.

Within-Subject Cluster Mobility

Finally, I examined the consistency of individuals' cluster memberships while reading in print and digital mediums. While the data suggest that some clusters functioned well regardless of whether the reading occurred in print or digitally, I had yet to ascertain whether there was consistency in cluster membership. In effect, I had yet to determine whether students who populated the highly capable and well calibrated reader, capable but less multimedia reader, or low performing and poorly calibrated reader clusters for the print condition were likely to be in the comparable clusters when they read digitally. To make this determination, I tracked the movement of each member of a cluster across mediums, which is graphically depicted in Figure 2. Also, a chi-square analysis was performed, [$\chi^2(2) = 22.75$, $p < .05$], suggesting that within-subject movement across clusters was significantly different than what could be attributed to chance.

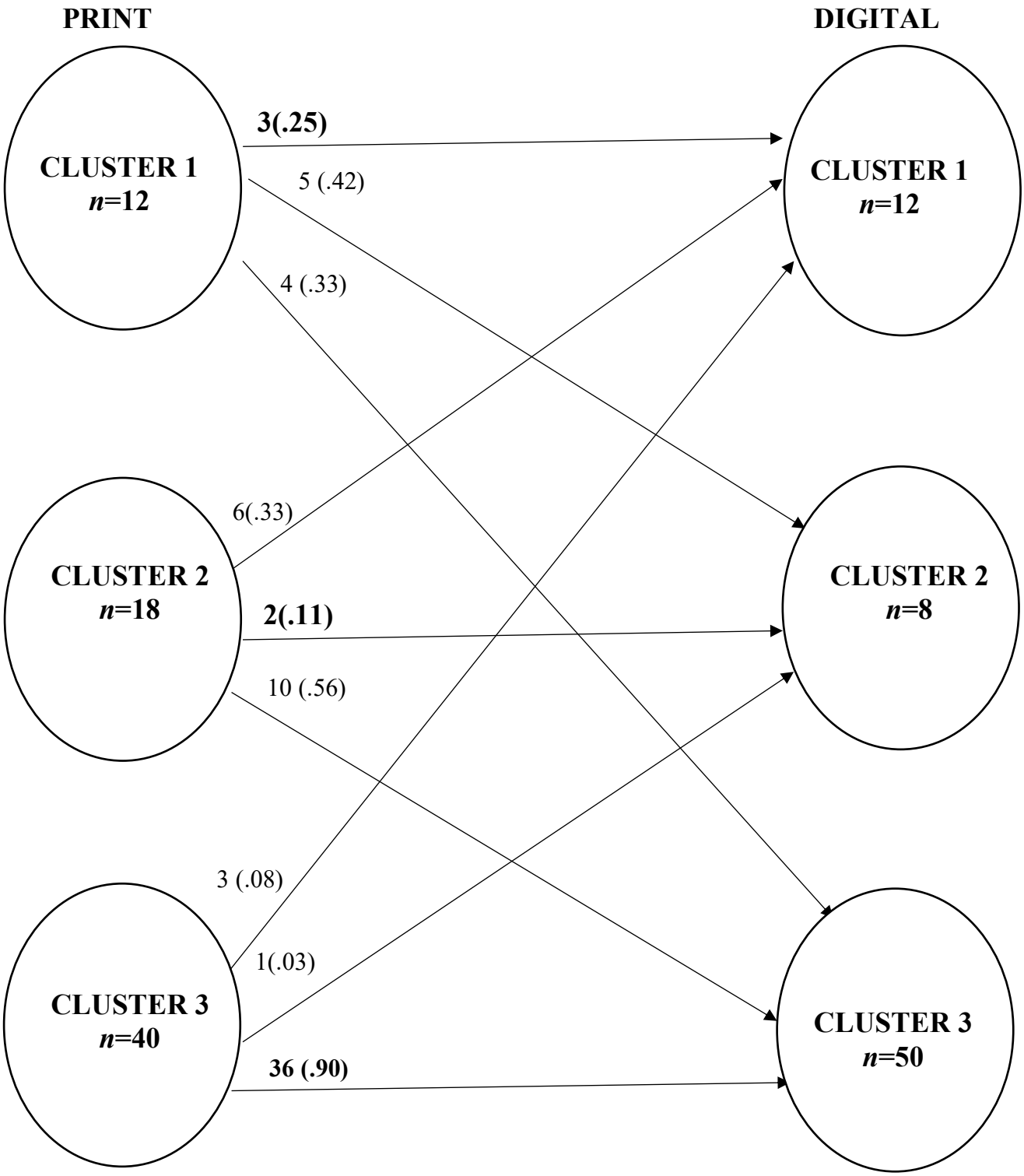

Figure 2. Number and proportion of participants comprising the three clusters when reading in print, and the repositioning of participants within comparable profiles when reading occurred digitally.

As shown in Figure 2, students populating the low performing and poorly calibrated reader clusters (i.e., Cluster 3) tended to display similar processing behaviors across mediums. In contrast, there was more mobility exhibited by undergraduates in Cluster 1 (highly capable and well calibrated readers) and Cluster 2 (capable but less multimedia readers) across mediums. Specifically, 42% of students who made up Cluster 1 for print reading moved to Cluster 2 when reading digitally, which means that those highly capable and well calibrated print readers were among the capable but less multimedia savvy digital readers. Another 33% of the highly capable and well calibrated print readers in Cluster 1 became low performing and poorly calibrated readers in the digital condition (i.e., Cluster 3).

When examining within-subject cluster movement for individuals in Cluster 2 (i.e., capable but less multimedia savvy readers), 33% of those individuals moved to the highly capable and well calibrated reader cluster (i.e., Cluster 1) when reading digitally. The final 56% of Cluster 2 became members of Cluster 3, the low performing and poorly calibrated reader cluster for the digital medium.

Summary

Latent Class Analysis was used to examine what clusters emerged from real-time processing behaviors, reading time, and calibration accuracy after reading in print and digitally. For the print data, a three-cluster model was adopted. The first cluster (n = 12) could be best characterized as highly capable and well calibrated readers. Cluster 2 (n = 18) was labeled as capable but less multimedia savvy print readers. The third and final print cluster (n = 40) is considered the low performing and poorly calibrated print readers. For the digital data, again a three-cluster model was the best fit. Similar to the print data, the first cluster (n = 12) could be best characterized as highly capable and well calibrated digital readers. Cluster 2 (n = 8) is

considered the capable but less multimedia savvy digital readers. Cluster 3 for the digital data (*n* = 50) is classified as the low performing and poorly calibrated digital readers. Finally, I examined the consistency of individuals' cluster memberships while reading in print and digital mediums. Students populating the low performing and poorly calibrated reader clusters (i.e., Cluster 3) tended to display similar processing behaviors across mediums. In contrast, there was more mobility exhibited by undergraduates in Cluster 1 (highly capable and well calibrated readers) and Cluster 2 (capable but less multimedia readers) across mediums.

Research Question 3

What are the effects of an intervention targeting self-regulation of digital reading time and depth of processing on undergraduates' comprehension performance and calibration accuracy when reading digitally?

To answer Research Question 3, the digital reading data from Study I for 65 undergraduates who agreed to participate in Study II served as the baseline for the intervention phase of this investigation. In what follows, I first review the digital reading preliminary intervention. Then, I provide the pre-intervention descriptives for those 65 students, along with their post-intervention data in Table 7. Next, I used a repeated measures ANOVA to investigate if the differences in comprehension and calibration accuracy pre-intervention to post-intervention were statistically significant.

Table 7

Pre-Intervention and Post-Intervention Means and Standard Deviations for Measures and Subcomponents

Measure		Pre-Intervention		Post-Intervention	
Subcomponent	max	*M*	*SD*	*M*	*SD*
Topic Knowledge	6	1.59	1.28	1.63	1.02
Processing Time		10m 46s	2m 33s	12m 6s	3m 37s
Comprehension					
Total	20	12.38	3.61	13.95*	3.81
Text Only	5	3.72	1.67	3.46	1.37
Visual Only	5	2.98	1.45	2.96	1.82
Text and Visual	10	5.71	2.19	7.57*	2.16
Judgment of Performance	100	54.68	20.32	41.89*	16.52
Calibration	0	15.98	11.06	28.54*	20.15

Note: * = $p < .01$

Comprehension and Calibration Accuracy Pre-Intervention to Post-Intervention

As shown in Table 7, the average overall comprehension score for the 65 participants at pre-intervention was 12.38 (*SD* = 3.61), as compared to 13.95 (*SD* = 3.81) at post-intervention. A repeated measures ANOVA with a Greenhouse-Geisser correction determined that the difference pre to post intervention was significant, [$F(1.19, 21.26) = 21.01$, $p < .01$], which indicates that the participants comprehension performance improved significantly as a result of the intervention. Similarly, comprehension scores for the text and visuals questions also increased significantly from pre-intervention (*M* = 5.71, *SD* = 2.19) to post-intervention (*M* = 7.57, *SD* = 2.16), $t(64) = 5.93$, $p < .01$, $d = 0.61$. Interestingly, after intervention, there was a slight decrement in performance for the text only questions (pre-intervention: *M* = 3.71, *SD* = 1.67; post-intervention: *M* = 3.46, *SD* = 1.37) and visual only questions (pre-intervention: *M* =

2.98, *SD* = 1.45; post-intervention: *M* = 2.96, *SD* = 1.82) slightly decreased. However, this difference in text only comprehension scores was not statistically significant, $F(3.01, 11.2) = 9.10$, $p = .12$.

Prior to the intervention, the judgment of performance ratings averaged 54.68 (*SD* = 20.32), but dropped significantly following the intervention to 41.89 (*SD* = 16.52), $F(11.09, 16.87) = 3.09$, $p < .05$. Thus, the undergraduates in this study had a more negative view of their comprehension performance after they had taken part in the intervention. This less favorable impression of their performance carried over into students' calibration accuracy scores. These scores, which indicate the degree to which students accurately predict their performance vis-à-vis the actual performance data, were found to be significantly less accurate after intervention than before (pre-intervention: *M* = 15.98, *SD* = 11.06; post-intervention: *M* = 28.54, *SD* = 20.15, $F(7.19, 6.06) = 3.87$, $p < .01$. Plausible explanations for these unexpected outcomes for both judgment of performance and calibration accuracy will be discussed in greater detail in Chapter 5.

Summary

A repeated measures ANOVA was ran to investigate if the differences in comprehension and calibration accuracy pre-intervention to post-intervention were statistically significant. The differences were significant, and although comprehension improved after the intervention, calibration accuracy scores worsened.

Research Question 4

What verifiable clusters emerge from real-time processing data, reading time, and calibration accuracy for undergraduates' reading digital texts after an intervention? Are there significant

differences for comprehension performance for the resulting clusters from pre-intervention to post-intervention?

In order to address Research Question 4, I set out to investigate what clusters emerged from real-time processing behaviors, reading time, and calibration accuracy for students completing the intervention. Latent Class Analysis (LCA) was again used to determine the number of statistically and theoretically viable clusters emerging from the data. Next, I examined whether these resulting clusters differed in terms of their comprehension performance by conducting Multiple Analyses of Variance (MANOVA) and post-hoc analyses using Tukey HSD. I also detail the findings for the post-intervention digital reading groups, and then compared those groups to the pre-intervention digital reading clusters described in Study I. To better understand the various clusters that emerged after the intervention, I will first revisit the nature of the intervention that was key to this portion of study.

Digital Reading Intervention

The intervention for this study consisted of five pre-recorded training modules delivered on computer that included narration and video clips. As detailed in Chapter 3, each training module was designed to address one of the key outcomes or challenges that emerged from the series of studies on print and digital reading my colleagues and I conducted over the past six years (e.g., student's reading preferences, differences between medium preferences and performance, speed of reading, judgment of performance, and different reader profiles). In effect, a large portion of the intervention focused on previous research on processing profiles, which emphasized the research on self-regulation, strategic processing, and reading time.

Although the text upon which topic knowledge, processing time, comprehension, and real-time processing behaviors were calculated in Study II (Volcanos) differed from the texts

used in Study 1 (Soil and Weather), direct comparison was deemed appropriate for several reasons. First, the passage was from the same textbook, *Geology and the Environment* (Pipken et al., 2014). Also, the text was similar in length (3 pages) and readability (grade 14; Fry Readibility; Fry, 1968) to the texts presented in Study I. Further, the passage used in Study II contained the same number and types of visuals used for the Soil and Weather passages in Study I (i.e., 1 diagram, 1 map, and 2 photos).

Post-Intervention Cluster Models

After the 65 undergraduates completed the intervention, they read the Volcano (Pipken et al., 2014) text digitally. As in Study 1, I next coded the real-time processing behaviors described in Chapter 3 (e.g., *Read*, *Refer to Visual*) captured on the Camtasia Screen Capture© Software footage when participants were reading digitally. This coding occurred at three second intervals. Additionally, reading duration as documented on Camtasia Screen Capture© software was recorded. Further, data was collected which was used to determine participants overall comprehension, comprehension by question type, judgment of performance, and calibration accuracy scores. To determine the number of theoretical and statistically viable clusters of readers, the processing behavior, reading time, and calibration accuracy were submitted to LCA, with comprehension performance as an external indicator that the groups were truly distinct. Five potential models emerged from the analysis (see Table 8). Similar to the process for Study I, the fit statistics (i.e., LL, AIC, and BIC) were examined with the most attention given to the BIC statistic due to the relatively small sample size. As shown in Table 8, the two-cluster model had the best fit based on the BIC. However, the difference in the BIC between the two-cluster and three-cluster models was only .22.

Further, upon examining the memberships for the two-cluster model, their composition did not seem theoretically viable. Specifically, one of the two groups that emerged in the two-cluster model contained 60 of the 65 participants. In addition, similar to the two- cluster models for the print and digital data from Study I, the performance patterns for the manifest variables for this group were not particularly discernible. For example, in terms of duration, some of the members of this cluster had the longest reading durations of the entire population (25m 40s), while others had the shortest (6m 11s). Similarly, calibration accuracy scores in this cluster ranged from the most accurate (0) to the least accurate (83). Further, while only a relatively small portion of this 60-person cluster relied predominately on relative deeper processing (approximately 20%) or peripheral processing behaviors during reading (approximately 28%), others used some combination of both (approximately 52%).

Table 8

Fit Statistics for Study II Exploratory Latent-Class Models

	Fit Statistics		
Number of Clusters	LL	AIC	BIC
1	-446.84	915.67	939.59
2	-408.25	862.49	912.51
3	-384.81	839.62	912.73
4	-371.90	837.80	940.00
5	-359.02	836.05	964.33

Note. LL, log-likelihood; AIC, Akaike's information criterion; BIC, Bayesian information criterion

In light of these outcomes for the two-cluster model, I re-examined model options. The three-cluster model had a marginally higher BIC than the two cluster-model, but that fit statistic was still smaller than the four- and five-cluster models. Also, the distribution of participants into

groups for the three-cluster model, and those groups' performance patterns were more theoretically justifiable than seen for the two-cluster model. For example, the membership numbers for the three clusters were more evenly distributed at 16, 18, and 31 individuals, respectively. Moreover, the real-time processing behaviors, reading time, and calibration accuracy scores for cluster members were more closely aligned. Thus, considering all of the fit statistics and theoretical viability, the three-cluster solution was adopted for the digital data. It is important to note that although I decided on three-cluster models based on the LCA fit statistics for both the digital condition pre-intervention and for the post-intervention data, the defining attributes for the three are not the same pre-intervention to post-intervention, as they were pre-intervention print to pre-intervention digital.

With the number of clusters decided for the post-intervention digital reading data, I examined whether these resulting clusters differed from each other both for overall comprehension performance and for the specific question types (i.e., text only, visual only, text and visual). The results of the MANOVA with total score as the dependent measure confirmed a significant difference in overall comprehension for the three clusters [F (62, 2) = 13.23, $p < .01$, $\eta^2 = .25$]. Post-hoc analyses using Tukey HSD tests revealed that the source of that difference was between the first and second clusters and the first and third clusters (Cluster 1: $M = 10.51$, $SD = 3.14$; Cluster 2: $M = 16.00$, $SD = 3.07$; Cluster 3: $M = 14.55$, $SD = 3.38$). Clusters 2 and 3 did not differ significantly from each other.

Also, when comprehension questions required information found only in the text, the performance of the three clusters again differed significantly [F (62,2) = 4.16, $p < .05$, $\eta^2 = .19$]. Post hoc analyses determined that the source of that difference was between the first and second clusters only (Cluster 1: $M = 2.69$, $SD = 1.40$; Cluster 3: $M = 3.58$, $SD = 1.36$). For the visual

only comprehension questions, the performance of the three clusters again differed significantly [F (62,2) = 7.67, $p < .01$, $\eta^2 = .25$]. Post hoc analyses determined that the source of that difference was between the first and second clusters and the first and third clusters (Cluster 1: M = 1.63, SD = 1.71; Cluster 2: M = 3.78, SD = 1.26; Cluster 3: M = 3.16, SD =1.81). There was no significant difference between Clusters 2 and 3.

Finally, when the comprehension questions required knowledge from the text and visuals, significant differences across the three clusters were again identified, [F (62,2) = 5.43, $p < .01$, η^2 = .18]. Post hoc analyses determined that the source of that difference was between the first and second clusters, and the first and third clusters (Cluster 1: M = 6.19, SD = 2.26; Cluster 2: M = 8.39, SD = 2.12; Cluster 3: M = 7.81, SD =1.83). As with the other question types, Clusters 2 and 3 did not differ from each other in any significant way.

Generally speaking, Cluster 1 can be best characterized as low comprehenders who were seemingly aware of their poor performance, whereas Cluster 2 included students who were high performing, but poorly calibrated digital readers. Finally, Cluster 3 was comprised of rather competent readers who seemingly lacked multimedia savvy and were poorly calibrated. Before detailing the character of these three clusters, I considered the mobility exhibited by the participating students from pre-intervention to post-intervention.

Pre-Intervention to Post-Intervention Mobility

Using the same procedure employed in Study I to examine the cluster membership of students when reading in print versus digitally, I explored the consistency of cluster membership of the 65 undergraduates before and then after intervention. The primary purpose in this examination was to identify any possible effects from students' exposure to the 5 training modules that pertained to certain characteristics of these clusters both pre-intervention and post-

intervention. The movement of each member of a cluster pre- to post-intervention is graphically depicted in Figure 3. A chi-square analysis was performed to determine if the movement in clusters pre-intervention to post-intervention was beyond what could be attributed to chance. The result of this analysis indicated that within-subject movement across clusters pre- to post-intervention was significantly greater than chance [$\chi^2(2) = 28.98, p < .01$]. This outcome suggests that the observed changes in cluster membership were attributable to students' participating in the intervention.

As shown in Figure 3, students populating Cluster 1 for the pre-intervention digital condition ($n = 12$) were characterized as highly capable and well calibrated digital readers. After the intervention, the 42% of the participants belonged to Cluster 3, which is the cluster composed of strong comprehenders who, nonetheless, lacked multimedia savvy as indicated by their treatment of the visuals in texts, and who were poorly calibrated. Four individuals (33%) from Cluster 1 pre-intervention moved to Cluster 2, which is the cluster consisting of students who had the highest comprehension scores but who were poorly calibrated.

There were eight individuals populating the cluster described as capable but less multimedia savvy digital readers before the intervention (Cluster 2). Of these eight individuals, the majority (63%) were members of Cluster 2 after the intervention. That means these students became high performing, but poorly calibrated readers. Prior to the intervention, Cluster 3 contained 45 participants. Members in Cluster 3 were classified as low performing and poorly calibrated digital readers. After the intervention, 24 (53%) students in this pre-intervention Cluster 3 became members of the post-intervention Cluster 3, which included students who were competent readers who seemingly lacked multimedia savvy and were poorly calibrated,

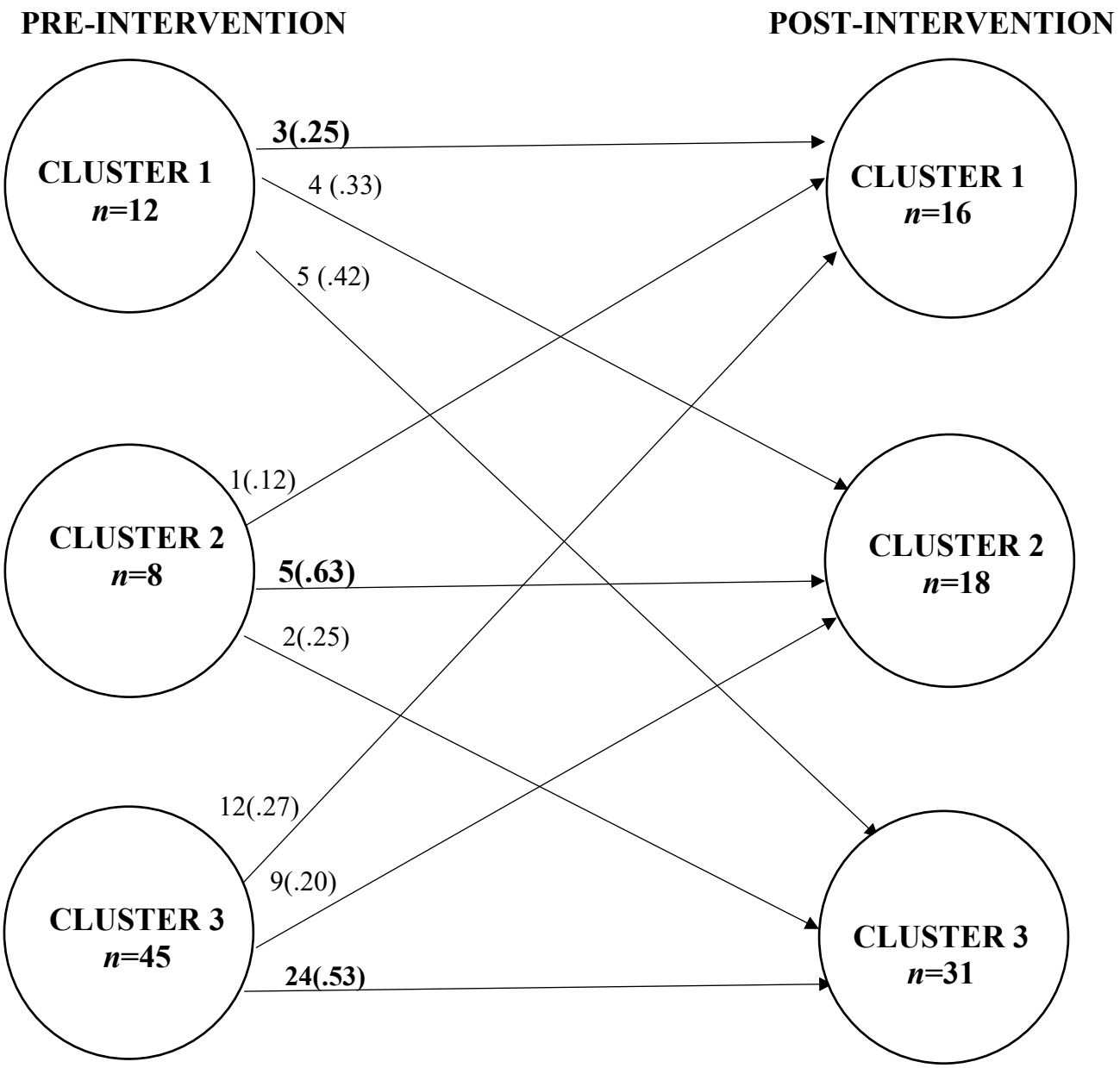

Figure 3. Number and proportion of participants comprising the three clusters when reading digitally pre-intervention, and the repositioning of participants when reading occurred digitally post-intervention.

Post-Intervention Cluster Characterizations

Cluster 1. The first cluster of digital readers identified for the post-intervention data consisted of 16 students. A breakdown of their means and standard deviations for reading duration and calibration accuracy by cluster are presented in Table 9, and mean and standard deviations for real-time processing behaviors can be found in Table 10. The undergraduates in Cluster 1 spent the least amount of time with the digital text than any other cluster (Cluster 1: *M* = 10m 29s, *SD* = 3m 27s; Cluster 2: *M* = 11m 56s, *SD* = 3m 18s; Cluster 3: *M* = 13m 1s, *SD* = 3m 42s). Further, as shown in Table 10, individuals comprising Cluster 1 had the highest percentages for two behaviors that suggest more surface-level processing, *Scroll* and *Skim*.

Members of Cluster 1 were also more highly calibrated than those in Clusters 2 or 3 (Cluster 1: *M* = 6.56, *SD* = 5.39; Cluster 2: *M* = 33.50, *SD* = 20.19; Cluster 3: *M* = 37.00, *SD* = 16.58). The difference in calibration accuracy between Cluster 1 and Clusters 2 and 3 were statistically significant, [F (62, 2) = 20.64 $p < .05$, $\eta^2 = .23$]. When compared with members of Clusters 2 and 3, members of Cluster 1 had the lowest total comprehension score, [F (62, 2) = 13.23, $p < .05$, $\eta^2 = .18$]. Based on these data, Cluster 1 could be best characterized as low comprehenders who were seemingly aware of their poor performance.

In understanding the performance of Cluster 1 in relation to the focus of the intervention, I can make several observations. Despite their poor comprehension abilities, it would appear that individuals in Cluster 1 learned from the intervention to be more aware of their limitations when it comes to comprehending digital text. However, it seems that they did not apply the lessons regarding speed of processing, considering members of this cluster had the lowest reading durations.

Table 9

Study II Means and Standard Deviations for Measures by Clusters

Measure Subcomponent	Clusters			
	Combined	1	2	3
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
N	65	16	18	31
Comprehension				
Total	13.95 (3.81)	10.51[b] (3.14)	16.00[a] (3.07)	14.55[a] (3.38)
Text Only	3.46 (1.37)	2.69[b] (1.40)	3.94[a] (1.11)	3.58 (1.36)
Visual Only	2.95 (1.82)	1.63[b] (1.71)	3.78[a] (1.26)	3.16[a] (1.81)
Text and Visual	7.57(2.16)	6.19[b] (2.26)	8.39[a] (2.12)	7.81[a] (1.83)
Duration	12m 6s	10m 29s	11m 56s	13m 1s
	(3m 37s)	(3m 27s)	(3m 18s)	(3m 42s)
Judgment of Performance	41.89	45.31	46.50	37.50
	(16.52)	(15.43)	(18.22)	(15.35)
Calibration	28.54	6.56[a]	33.50[b]	37.00[b]
	(20.15)	(5.39)	(20.19)	(16.58)

Note. [a] and [b] represent significant differences between clusters, $p < .05$.

Table 10

Study II Processing Behavior Percentage Means and Standard Deviations by Clusters

Processing Behaviors	Clusters			
	Combined	1	2	3
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Read	72.44 (16.74)	76.31 (14.73)	66.79 (16.34)	73.72 (17.58)
Re-Read	8.38 (8.62)	5.80 (8.63)	9.58 (8.19)	9.02 (8.98)
Re-Position	1.39 (2.11)	1.37 (2.42)	1.78 (1.75)	1.18 (2.17)
Scroll	3.42 (1.79)	4.11 (1.79)	3.51 (2.06)	3.02 (1.56)
Skim	0.81 (3.42)	2.58 (6.61)	0.13 (0.54)	0.28 (0.81)
Refer to Visual	4.55 (3.53)	3.98 (3.40)	3.74 (4.69)	5.31 (2.67)
Refer to Caption	4.56 (4.78)	4.09 (4.69)	3.86 (3.69)	5.20 (5.40)
Look at Visual When Cued	1.53 (3.65)	0.25 (1.00)	5.30 (5.34)	0.00 (0.00)
Read Caption When Cued	1.26 (2.57)	0.42 (1.68)	4.16 (3.14)	0.00 (0.00)

Cluster 2. The second cluster identified for the post-intervention data was comprised of 18 members. Students in Cluster 2 had the highest total comprehension score, which was significantly higher than that for Cluster 3, [F (62, 2) = 13.23, $p < .05$, η^2=.18]. Further, the students in Cluster 2 had calibration accuracy scores that were higher than Cluster 3 (although only descriptively), but statistically significantly worse than individuals in Cluster 1 [F(62, 2) = 20.64, $p < .05$, $\eta^2 = .23$]. As seen in Table 10, members of Cluster 2 had the highest instances of two processing behaviors that are indicative of relative deeper-processing behaviors, *Look at*

Visual When Cued and *Read Caption When Cued*. In effect, these data suggest that Cluster 2 students could best be characterized as high performing, but poorly calibrated digital readers.

It would appear that members of Cluster 2 were able to gain a better understanding from the intervention of how to navigate digital text in order to improve comprehension. While the individuals populating Cluster 2 were able to outperform their peers on the comprehension measure, they still struggled to accurately judge their own performance. It seems that after completing the intervention, these participants were not attuned to their improvements in navigating and comprehending the text.

Cluster 3. The third and final cluster for the post-intervention digital data consisted of 31 students—the largest of the three groups. As noted, comprehension scores for Cluster 3 fall lower than individuals in Cluster 2 but not statistically significantly so. However, comprehension scores for Cluster 3 were statistically significantly higher than members in Cluster 1, [F (62, 2) = 13.23, $p < .05$, $\eta^2 = .18$]. Further, their calibration abilities were the poorest of all three clusters, and significantly below that of Cluster 1, [F (62, 2) = 20.64, $p < .05$, η^2=.23]. Moreover, the majority of students in this clusters (n = 30, 96.77%) judged their performance to be much worse than it actually was.

What is interesting about this cluster is that this group of students engaged in some processing behaviors indicative of relative deeper processing at percentages higher than the other clusters, such as *Refer to Visual* and *Refer to Caption* (see Table 10). However, this picture contrasts with the pattern that emerged when this cluster dealt with the visuals in text. In effect, although Cluster 3 members evidenced some attention to the visuals in the reading, they rarely did so when directed to by the text (e.g., "As shown in…."). Based on these data, Cluster 3

members can best be described as rather competent readers who nonetheless lack multimedia savvy and who are poorly calibrated digital readers.

For the individuals populating Cluster 3, their performance after the intervention can be characterized as readers who are able to understand the digital text, but who still struggle to judge their own performance. Members of Cluster 3 saw increases in comprehension scores as a consequence of the intervention. However, it is noteworthy that these students' calibration accuracy *decreased* as a consequence of the intervention. This decrease in calibration ability may reflect participants overgeneralization of the content from the intervention, which stressed how competent readers from previous studies often overjudged their comprehension after reading digitally.

Summary

Latent Class Analysis was used to examine what clusters emerged from real-time processing behaviors, reading time, and calibration accuracy after completing the intervention. With the data from the 65 participants who returned for the intervention, a three-cluster solution was selected. Cluster 1 (n = 16) could be best characterized as low comprehenders who were seemingly aware of their poor performance. Cluster 2 (n = 18) students could best be characterized as high performing, but poorly calibrated digital readers. Finally, Cluster 3 (n = 31) members can best be described as rather competent readers who nonetheless lack multimedia savvy and who are poorly calibrated digital readers. Next, using the same procedure employed in Study I to examine the cluster membership of students when reading in print versus digitally, I explored the consistency of cluster membership of the 65 undergraduates before and then after intervention.

CHAPTER 5
CONCLUSION AND DISCUSSION

Investigation Overview

For the past six years, I have been focused on understanding the role that medium, be it print or digital, plays in student learning and development. My particular interests have concerned learning from text and I have explored how medium shapes how students process and understand what they read, and how they judge their own performance. The studies that I undertook to explore these broad questions have afforded critical insights into the effects of medium on even competent readers' comprehension and calibration accuracy that I would not have expected from the outset. Despite what has been learned about medium effects on comprehension, there is still much to be learned regarding the differences in comprehension and calibration when students read in print and digitally. There are also puzzling questions as to *why* some of the rather consistent findings my colleagues and I have reported exist (e.g., Singer & Alexander, 2017a; Singer & Alexander, 2017b; Singer Trakhman et al., 2018b).

These unanswered questions and puzzlements were the catalyst for the two-study investigation I have reported here. As was hypothesized, the outcomes of this dissertation reinforced key findings that have carried forward across the prior studies I have conducted (Singer & Alexander, 2017a; Singer Trakhman et al., 2017; Singer Trakhman et al., 2018a; Singer Trakhman et al., 2018b). Moreover, the unique design and features of this dissertation have also led to important new insights about medium effects and why they may occur. However, before sharing those insights and offering recommendations for future research and practice, let me address limitations of this work that must acknowledged.

Key Findings: Continuing Trends and New Insights

I have organized the key findings of this two-study dissertation into two categories. The first pertains to the findings of the various analyses undertaken in Study I and Study II that replicate outcomes that have been very consistent across the four studies that have been conducted and reported in the literature. The second identifies what was novel in the design of the current dissertation and what new insights that novelty produced.

Continuing Trends

Over the past decade, the empirical literature has found differences in reading comprehension attributable to the medium of text delivery (e.g., Kerr & Symons, 2006; Mangen et al., 2013; Singer & Alexander, 2017a; Singer Trakhman et al., 2018b). Also, findings of a systematic literature review on this topic (Singer & Alexander, 2017b) suggested that medium plays a more or less influential role based on certain text or task conditions or as a result of certain reader characteristics. One of the key findings in my previous works (e.g., Singer & Alexander, 2017a; Singer Trakhman et al., 2018b) is that students comprehend more when reading in print for any questions that probe beyond the main idea, where comprehension was equivalent across mediums. In this dissertation, comprehension scores for the print medium was superior to digital, no matter the question type (i.e., text only, visual only, text and visual). This suggests that the medium in which the text is presented serves as the determiner of performance over the mode in which the information is conveyed (i.e., text or visual).

Further, participants read the printed texts at a much slower pace, spending on average 90 seconds longer with the printed text. This finding is in line with some of my co-authors and I previous investigations (e.g., Singer et al., 2017a; Singer Trakhman et al., 2018a) and suggests that there is something about reading digitally that seemingly increases the speed at which

students move through the text. Further, the shorter reading times for the digital medium support Wickelgren's (1997) speed-accuracy tradeoff hypothesis; that is, when the speed at which a task is performed increases, there may well be a decrement in performance to consider. That tradeoff was certainly evident amongst the participants in this dissertation. As these competent readers move from print to the digital medium, the time they spent with the text decreased significantly, and so did their subsequent comprehension scores.

When reading in print, participants engaged in higher percentages of relative deeper-processing behaviors than when reading digitally. Specifically, participants engaged more frequently in the following processing codes: *Re-Read*, *Refer to Visual, Refer to Caption, Look at Visual When Cued,* and *Read Caption When Cued*. Despite the difference across mediums in the frequency of peripheral or relative deeper-level processing behaviors, it would appear that reading digitally consists of the same behaviors as reading in print, and that there is little need to articulate a new theory of digital reading from the standpoint of processing behaviors as texts migrate from page to screen. In effect, the peripheral or relative deeper-level processes and resulting profiles captured reading in both mediums. What seemingly changed when the texts were transferred from the page to the screen was the distribution of these processing behaviors. Thus, those seeking to articulate models or theories of reading digitally cannot continue to contend that reading digitally is a different act than reading from a printed page.

In this dissertation, I endeavored to analyze competent readers' interactions with the text by documenting the various behaviors they manifested while engaged with print and digital texts. Over the course of prior studies investigating medium effects on undergraduates' comprehension and calibration abilities, my co-authors and I found that these processing behaviors are key to understanding how and how well students learn from the assigned readings (e.g., Singer

Trakhman et al., 2018a). Specifically, we have been interested in understanding peripheral or relative deeper-level processing behaviors, which serve as an important line of research aimed at understanding how students navigate texts and what effect these behaviors have on comprehension. Although research and popular culture would consider the participants of this dissertation to be "digital natives" (Prensky 2001, 2011), these participants employed lower amounts of relatively deeper-level processing behaviors when navigating the text digitally. As shown in Table 3, processing behaviors manifested differently dependent on which medium the text was presented in. Now of course, some may argue that this is because they are so facile at navigating digital texts that they don't need to employ as many relatively deeper-level processing behaviors when reading digitally. However, when looking at comprehension outcomes by medium, that cannot be the case.

New Insights

Prior to this dissertation, my co-authors and my most recent investigation contained an anomalous finding regarding calibration accuracy across mediums (Singer Trakhman et al., 2018b). That investigation was the first time that those reading digitally were significantly better calibrated than those reading in print. In the present dissertation, participants were again more highly calibrated when reading digitally. While not hypothesized, this new trend in regard to calibration after multiple studies (Singer et al., 2017a; Singer Trakhman et al., 2018a) finding participants to be more highly calibrated after reading in print is encouraging as more and more tasks move online.

There are a few features of this dissertation that differentiate it from previous works while adding more sophistication to the design and subsequent findings. For one, this is my first investigation in which prior topic knowledge was measured using a six-question measure that did

not rely on self-report data. Further, these topic knowledge scores for text subject (i.e., Soil, Weather, and Volcanoes) were able to be used a covariate in the MANCOVAs used in Research Questions 1 and 3. Moreover, the decision to incorporate prior topic knowledge into the conceptual model (see Figure 1) is consistent with Kintsch's (1988) model in which prior knowledge of the text topic is projected to influence comprehension for even competent readers. To be an appropriate covariate, topic knowledge had to be significantly related to comprehension and calibration accuracy. As hypothesized, topic knowledge did have a significant effect on both comprehension and calibration outcomes across the two studies. Thus, the resulting outcomes are able to be interpreted with more clarity in regard to the mediums effect on comprehension and calibration accuracy scores without worrying about the effect of topic knowledge.

The next advancement in this dissertations design relates to the comprehension measures used. For one, the comprehension questions and scoring guides were reviewed by two science teachers in order to judge the accuracy of the items and the proposed response guides. This extra step in validating the measures helps to ensure that this dissertation was measuring appropriate science content that could be found within the text passages, a coup many similar studies have yet to achieve according to a review on the relevant literature (Singer & Alexander, 2017b). Further, the comprehension measures in this dissertation improved on my own previous investigations (e.g., Singer et al., 2017a, Singer Trakhman et al., 2017; Singer Trakhman, 2018a) by removing the main idea question, which was a question that students consistently were able to answer correctly no matter the medium. Further, I was able to add several questions that rely on different information sources in the passage (i.e., text only, visual only, text and visual), which allowed me to see where the differences in comprehension may lie.

Finally, to my knowledge, this dissertation features the first real-time examination of processing behaviors as students read multimodal texts across multiple mediums. I collected real-time processing behaviors using a GoPro© Camera and Camtasia Screen Record© Software to capture undergraduates' movements through the texts. As in the prior study, the students signaled their position in the text by using either a pen or mouse curser to indicate the specific words they were reading (Singer Trakhman et al., 2018b). These processing behaviors, along with reading time and calibration accuracy, served as the database for Latent Class Analysis (LCA). This is a more sophisticated method of person-centered analyses than previously used.

With regard to the clusters identified, the outcomes indicate that there is no uniformity even among competent readers' engagement with and learning from domain-specific exposition texts. As shown in Figure 3, movement in cluster membership when the medium changed was ubiquitous for even members of the competent reader and calibrators cluster. This raises the question, what is it about the medium in which the text is presented that alters the way in which students navigate and comprehend the material?

Despite the puzzling findings regarding movement in individual cluster membership across mediums, there are still discernible patterns in how the readers engage with the multimodal texts. More often than not, readers choose to look at the visuals in the text or read the captions under these visuals when they appear in the text, instead of when the text cues them to do so. As described in Chapter 2, there is abundant evidence that adding a visual element to a text can support and enhance learning for certain populations, such as the population for this dissertation (Ainsworth, 2008; Goldman, 2009). When done correctly, learning from two representations (i.e., text and visuals) can be better than text alone (Goldman, 2003). It would appear that this tendency to examine the visuals and corresponding captions in the text when

they appear, instead of when the relevant text calls their attention to them, is evidence that the readers in this study did not apply the key self-regulatory processes necessary to enhance learning from the multimodal texts (Azevedo & Cromley, 2004). This finding suggests that students in this investigation did not take advantage of the multimodal representations offered within the text passages.

Even a brief intervention can effect changes to the reading duration, processing behaviors, comprehension scores, and calibration accuracy abilities for competent readers. The 30-minute pre-recorded training modules were able to raise reading duration, increase comprehension, increase the use of relatively deeper-level processing behaviors, and for some, increase calibration accuracy. When looking at the three clusters that manifested for the post-intervention data, I noticed that each cluster highlights a part, or parts, of the intervention that was successful for that group. For the 16 individuals populating Cluster 1, their comprehension abilities were below average (M = 10.51, SD = 3.14), but their calibration accuracy was stellar (M = 6.56, SD = 5.39). Although it appears that the intervention did not successfully improve their comprehension abilities, their ability to reflect on their own poor performance is surely an improvement. For the members of Cluster 2, there was a marked improvement in total comprehension scores after completing the intervention. In fact, their comprehension score (M = 16.00, SD = 3.07) was the highest of any cluster pre or post-intervention, no matter the medium. This radical increase in comprehension ability is a testament to the effect even a 30-minute intervention can have for certain students. Although their total comprehension scores were not as high as members of Cluster 2, the individuals populating Cluster 3 still earned total comprehension scores that were higher than any of the digital clusters pre-intervention (M = 14.55, SD = 3.38). Although Cluster 2 is known as the stellar comprehenders, it should be noted

that members of Cluster 3 greatly improved their comprehension scores as a result of the intervention, too.

What is particularly noteworthy about the increase in total comprehension scores for participants after the intervention is the source of the differences pre-intervention to post-intervention. Specifically, as shown in Table 9, the only statistically significant increase in comprehension scores by question type was for questions requiring information found in both the text and visuals (pre-intervention: $M = 5.71$, $SD = 2.19$; post-intervention: $M = 7.57$, $SD = 2.16$). As discussed in Chapter 2, researchers have frequently suggested that students often do not apply the key self-regulatory processes and mechanisms such as effective cognitive strategies and metacognitive monitoring necessary when learning from multimodal texts (e.g., Azevedo & Cromley, 2004; Azevedo et al., 2004). So, it is encouraging that an intervention on digital reading was able to cause such a dramatic improvement in questions requiring comprehension of multimodal content, especially considering these texts and questions are the kind of reading materials that undergraduates encounter in their academic studies most frequently (Alvermann & Wilson, 2011).

Limitations

The first limitation in this dissertation concerns the novel way in which processing behaviors were captured for both print and digital reading. Using the Go-Pro© and Camtasia Screen Record© Software while participants followed along as they read seems to be a strength and weakness of this dissertation. On the one hand, it captured the data in a way that allowed me to see how students navigate the texts without interfering with their process. However, it is possible that the laboratory setting and the awareness that their movements were being recorded, in addition to having to "follow along" with the pen or cursor may have altered their behaviors in

unintended ways, This data-capturing system might also have encouraged participants to be more or less diligent than they typically would have been when reading.

The second limitation regarding this dissertation pertains to perhaps its most important design component: the explicit training that sought to intervene with regard to students' digital processing behaviors and reading speed. To my knowledge, this intervention was the first of its kind to attempt to change the way in which students navigate digital text. However, the intervention was very brief, lasting about 30 minutes. Further, although the intervention was designed with the findings from the literature on interventions in mind (e.g., provide opportunities for feedback, online delivery, and focus on strategies; Meyer et al., 1989; Nietfeld & Schraw, 2002), it was never pilot tested. Moreover, while there were elements of the intervention that proved successful at getting the participants to spend more time with the digital text and engage in behaviors that extended beyond surface level movement through the text, there were less positive outcomes that were identified and that could be traced to components of the intervention. Specifically, the decline in calibration accuracy scores observed after participants completed the intervention appears to be an unplanned effect of the intervention. It was hypothesized that after completing the intervention module on judgments of performance and calibration accuracy, judgments of performance would be more closely aligned with actual performance, thus improving the calibration accuracy scores. In this intervention, participants learned about the speed-accuracy tradeoff and its implications for judgments of performance and calibration accuracy (Wickelgren, 1977). However, after the intervention average judgment of performance ratings dropped over ten points. This suggests that instead of being more reflective in gauging their own performance, the participants merely just lowered their judgments of performance after reading the post-intervention digital text.

Finally, this dissertation only captured post-intervention data immediately after students completed the training. There was no extended follow up in order to see if the improvements in the speed, reading behaviors, and comprehension of digital texts would still exist weeks or months after the intervention had been completed. As such, I have no sense of whether the participants would revert back to their initial processing behaviors the next time they encounter digital texts.

Implications for Research

Even considering the aforementioned limitations, findings from this dissertation provide valuable information in regard to the effect of medium on multimodal texts. Specifically, this investigation largely replicated the outcomes from the previous studies my colleagues and I have undertaken—outcomes that point to the negative consequences of reading digitally on comprehension. Participants in Study 1 spent considerably less time with the texts and engaged in lower percentages of relatively deeper-processing behaviors when reading the passages digitally, and their comprehension suffered accordingly.

Moreover, it would appear that the multimodal character of the expository texts in this study exacerbated some of the patterns documented in the prior investigations. For example, this dissertation highlighted the influence that the location of visuals within a passage has on when they are attended to. Within both Study I and Study II, participants overwhelmingly accessed visuals and their corresponding captions when they appeared in the passage, as opposed to when the text directed them to do so (e.g., “see Figure 1.3 for a diagram depicting the soil profile”). Although the participants in this study are enrolled in a University and considered to be competent readers, most of them were unable to coordinate the visuals in the passage to the text

in the passage in a meaningful way. This suggest that the majority of participants were not strategic about the role that visuals serve in a passage.

An additional implication for research pertains to the intervention presented in Study II. As mentioned, although the intervention was largely successful, there are many implications to help improve future attempts at intervening on digital reading. For one, the role of visuals in text passages merits attention in future interventions. The findings of this dissertation show that students are generally not accessing the visuals at the time when the subject of the visuals is being read, suggesting that students still treat texts and visuals in one passage as separate entities (Mayer, 2002).

Further, as discussed in Chapter 2, researchers have frequently suggested that students often do not apply key self-regulatory processes and mechanisms such as effective cognitive strategies and metacognitive monitoring necessary when learning from multimodal texts (e.g., Azevedo & Cromley, 2004; Azevedo et al., 2004), so it is imperative that future interventions on digital reading of multimodal texts ensure to explicitly demonstrate how to apply cognitive strategies and metacognitive monitoring when reading multimodal texts.

In addition to including material on the role of visuals in text within future interventions, the findings from this dissertation suggest that there needs to be refinement to the judgment of performance and calibration accuracy modules. Within this dissertation, participants caught onto the idea that people generally overestimate their performance when reading digitally. Therefore, their rather non-strategic response appears to have been to simply lower their self-reported judgments of performance following the training.

Further, as mentioned in the limitation section, it is possible that the laboratory setting and the awareness that their movements were being recorded, in addition to having to “follow

along" with the pen or cursor may have altered their behaviors in unintended ways. This data-capturing design might also have encouraged participants to be more or less diligent than they typically would have been when reading similar texts. With this in mind, researchers in the future should consider procedures for more effective collecting real-time data during print and digital reading in as non-intrusive and naturalistic a way as possible.

Implications for Practice

The findings of this dissertation have multiple implications for practitioners. For example, as the findings from Study I suggest, educators should not assume that preference for and familiarity with digital texts equates to proficient reading with digital texts. Although the sample in this dissertation was undergraduates who all responded after completing Study I that they believed they had the reading skills necessary for the processing of the texts, it is clear that they do not bring the necessary strategic and metacognitive behaviors required to be effective comprehenders to digital reading tasks. This means that even though a student is competent enough to be in college-level courses, they still may be lacking the skills necessary to understand the readings and other materials presented digitally within the course. Within Study I, 50 of the 70 participants were in Cluster 3 for the digital medium, which is the cluster for those who performed poorly on the comprehension measure and had poor calibration accuracy skills. As an educator, one must keep in mind that the majority of their students may not have the skills necessary to comprehend and to be metacognitive about tasks when they are presented digitally.

Another implication for practitioners can be seen when examining the findings of Study II, which showed that comprehension of digital texts can be improved through an effective intervention. Although the effects of the intervention varied, there were gains in comprehension for participants in this dissertation. Further, even this rather brief intervention was successful in

getting students to slow down when reading digitally, which is very encouraging considering the difference in reading speeds across mediums and subsequent consequences for performance. As mentioned, educators cannot assume that the skills and processing behaviors students carry when reading in print are automatically transferred to the same task when presented digitally. With this in mind, it is encouraging to know that the skills and processing behaviors needed to successfully comprehend and calibrate are in fact teachable. Or, educators should provide opportunities for their students to read for their class in print.

APPENDICES

APPENDIX A

Demographic Questionnaire

Answer the following questions.

1. What is your class standing?
a. Freshman
b. Sophomore
c. Junior
d. Senior

2. What is your major?

3. What is your approximate GPA?

4. What is your age?

5. What is your ethnicity (select all that apply)?
a. African American/Black
b. American Indian
c. Asian/Pacific Islander
d. Caucasian/Non-Hispanic White
e. Hispanic
f. Other

6. What is your Gender?
a. Female
b. Male
c. Non-binary
d. Prefer not to answer

7. Are you a native English speaker?
a. Yes
b. No

APPENDIX B

Study I Pretests with *Answers*

SOILS PRETEST

1. Define the following terms[1]:
   a. Soil Horizons

   A layer of soil parallel to the soil surface, whose physical characteristics differ from the layers above and beneath.

   b. Loess

   A loosely compacted deposit of windblown sediment of which extensive deposits occur

   c. Residual Soils

   Soil formed in situ by rock decay and left as a residue after the leaching out of more soluble products.
2. What is the solid rock underlying loose deposits such as soil known as?
   a. Top Layer
   b. Regolith
   c. Soil horizons
   d. Bedrock*
3. What influences the thickness and development of soil profiles? Select all that apply.
   a. Climate*
   b. Time*
   c. Temperature
   d. Weather
4. What happens to residual soils over time?
   a. Once formed, they maintain their size and shape
   b. They are always subject to erosion and removal by geological processes*
   c. They only sometimes can be removed by geological processes
   d. They grow in shape and size

[1]Each correct question is worth 1 point.
*Indicates the correct answer.

WEATHER PRETEST

2. Define the following terms[1]:

   a. Inundation

   A flood of water, the most common impact of floods causing large movements of sediment.

   b. Storm Surges

   When hurricanes or nor'easters make landfall, they push before them a wall of water. These surges of water that can rapidly raise sea level by meters.

   c. Hydrologists

   Scientists who deal with water: its properties, circulation, and distribution on Earth and in the atmosphere

3. Why are heavy thunderstorms a summer phenomenon?

   a. The land needs water to cool off due to the climate
   b. Strong solar heating makes the atmosphere unstable*
   c. Moisture streams north of the equator due to the Earth's rotation
   d. Air masses move differently in high temperatures, causing storms

4. Which of the following is known to cause floods? Select all that apply.

   a. Precipitation*
   b. Dams
   c. Earthquakes*
   d. Climate warming*

5. Why are mountains wetter than adjacent valleys?

   a. Cool air ascends and condenses*
   b. Warm air descends causing cloud dissipation
   c. Clouds form over the adjacent valleys
   d. Solar heating moves more quickly over mountains

[1]Each correct question is worth 1 point.

*Indicates the correct answer.

APPENDIX C

Study I Comprehension Measures and *Answers*

SOIL COMPREHENSION TEST

20 points possible

1. What is a loess? List up to 2 of its physical properties.

3 points possible – (text only)

Loess is a windblown silt composed of very small grains of feldspar, quart, calcite, and mica.

Loess is very fertile.

Loess has a loamy text, allows roots to penetrate

Loess retains water well.

Minerals in loess are very small, which helps to release nutrients to plants that grow there.

2. How does soil form? (text only)

2 points possible – (text only)

Soil forms by the weathering of regolith, the fragmental rock material at and just below Earth's surface

3. According to the information visually presented in the passage, where would you find 4 areas of the earth with extensive loess covering?

3 points possible - (visual only)

Pampas of Argentina

Central Plains

East of Mississippi Valley

North Africa

Central Asia

North European Plain

4. What is an arroyo?

2 *points possible- (visual only)*

An incised streambed in which water flows only rarely.

5. Name and describe 4 of the 5 layers in an idealized soil profile.

4 points possible - (text+visual)

O – Humus

Provides CO_2 and organic compounds to soil below

A- Topsoil

Uppermost soil – mineral matter dissolved by water

B – Subsoil

Where dissolved elements accumulate

C- Bedrock

The weathered transition zone that grades downward into the lowest layer

Fresh parent material

The bottom of the soil profile, - the thickest layer

6. Name a type of transported soil and where you are likely to find it. (Text+ visual)

2 points possible - (text+visual)

Alluvia soil – rivers

Eolian soils – mountains

Glacier soils – glaciers

Volcanic soils – near where a volcano erupted

7. List up to 4 additional facts or ideas that you remember from the passage.

4 points possible

WEATHER COMPREHENSION TEST

20 points possible

1. Describe 2 ways in which factors beyond rainfall can contribute to flooding.

4 points possible - (text only)

Earthquakes – tsunami waves can wash ashore

Earthquakes – can change land levels

Climate change – sea level in North America has moved 100 meters

Climate Change – rapidly melting ice in oceans can cause rapid flooding

2. Why are heavy thunderstorms a summer phenomenon?

2 points possible - (text only)

In the summer, the atmosphere is unstable as a result of strong solar heating.

3. As displayed in the map from the passage, identify 2 areas of the world with above average rainfall **and** 2 areas of the world with below average rainfall.

4 points possible – (visual only)

Above average:

Middle of Africa

East Australia

Bottom of South America

East South America

Below Average:

West South America

North Africa

West Australia

North Canada

4. Name one of the two states features in the photographs related to Nor'easters.

1 point possible - (visual only)

Vermont or Virginia

5. Why are mountains always wetter than valleys?

3 points possible - (text+ visual)

Mountains are always wetter than valleys because of the air movement around them. Specifically, air masses are forced to move up and over mountain ranges and cool, which allows the moisture them to condense and rain out. Once the air mass moves over the mountain peak into a valley, the air becomes warm, drying the landscape below (a valley).

6. Why is Southern Texas considered a flooding hot spot?

2 points possible- (text+ visual)

It is where dramatic topography and storms meet – the Gulf of Mexico contributes warm winds – which turn dry riverbends into flood zones very quickly. Also known for highest rainfall in shortest time.

7. List up to 4 additional facts or ideas that you remember from the passage.

4 points possible

APPENDIX D

Intervention Training Outline

The outline below lists the topics represented in the modules, the *approximate durations*, the material that was overviewed and the feedback procedure.

MODULE I: Reading Preferences *(4 minutes)*

Part 1: Do You Read Like Your Peers?

The participant was first asked which medium (i.e., print or digital) they prefer to read in. They will then be asked to predict the medium in which they believe undergraduates prefer to read in. The module will prompt them to click on the answer they believe to be correct (i.e., print, digital, or no preference). If they selected "print" or "no preference" the module moved them to a slide that narrates, "That is incorrect. In a recent investigation of college-aged participants, 89% responded that they preferred to read text digitally." If they selected "digital," the module moved them to a slide that narrated, "That is correct! In a recent investigation of college-aged participants, 89% responded that they preferred to read text digitally."

Part II: Learning about Reading Preferences

Part II illustrated undergraduates' preference for digital reading found in previous studies (Singer Trakhman et al., 2017). A list of the reasons for their preference for digital text was shown (i.e., more interesting, saves paper, price, convenience, easier to search for specific text, can multitask; Millar & Schrier, 2015). This section also highlighted quotes from a qualitative study on preferences for reading digitally (e.g., "I love reading on my Kindle–even my textbooks are on there!" Millar & Schrier, 2015).

Part III: Performance Feedback

After learning about preferences for reading digitally, the participant was asked the following question:

Which of the following reasons were NOT given by participants in the Millar and Schrier (2015) study for why they prefer to read digitally?

a. *Easier to search for specific text*
b. *Don't have to go to library or bookstore*
c. *Better for the environment*

d. More affordable than print

If the participant selected the correct answer (don't have to go to the library or bookstore), they heard the following: "That is correct! Participants discussed preferring to read digitally for a plethora of reasons, but the burden of going to the library or bookstore was not one of them."

If the participant selected an incorrect answer, they heard the following: "That is incorrect. Participants discussed preferring to read digitally for a plethora of reasons, but the burden of going to the library or bookstore was not one of them." The training module showed the list of reasons for preferring digital as this narration occurs.

MODULE II: Print and Digital Comprehension Performance *(4 minutes)*

Part I: Differences between medium preferences and performance

This section contrasted preferences for digital reading with findings showing superior performance in print across age groups (Mangen, 2010; Sellen & Harper, 2002). Next, the training module highlighted findings from studies on undergraduate populations that find superior performance in print despite preferences for digital (Rideout et al., 2010; Singer Trakhman 2018a). Finally, possible explanations for superior performance in print were explained (e.g., more superficial processing, ergonomic differences when reading digitally, sequential versus continuous reading due to scrolling demands in digital).

Part II: Performance Feedback

After learning about the discrepancy between preferences for digital reading and higher comprehension scores for print reading, the participant was asked the following question:

Which of the following possible explanations for superior performance in print was NOT discussed?

a. Scrolling demands when reading digitally

b. Ergonomic differences when reading digitally

c. Navigating digital text more rapidly

d. More superficial processing of digital text

If the participant selected the correct answer (Navigating digital text more rapidly), they heard the following: "That is correct! Multiple reasons were given to explain the

possible difference in preference and performance, but the speed at which the text was read was not discussed."

If the participant selected an incorrect answer, they heard the following: "That is incorrect. Multiple reasons were given to explain the possible difference in preference and performance, but the speed at which the text was read was not discussed." The training module showed the list of possible explanations for this discrepancy in preference and performance.

MODULE III: Reading Speed *(4 minutes)*

Part I: Research on Reading Speed

This section began with an explanation as to why researchers examined speed of reading for print and digital texts (i.e., to look for a possible explanation for the differences in comprehension across mediums). Findings from past studies showing that people read more quickly in digital were discussed (Dyson & Haselgrove, 2000; Singer Trakhman et al., 2018b). The training module used a table to illustrate the differences in average reading times by medium in these studies.

Part II: Performance Feedback

After learning about differences in reading speed as a consequence of medium, the participant was asked the following question:

What reason was given as to researchers decide to measure reading speeds across mediums?

a. *to try and see which medium (i.e., print or digital) was better for speed reading*
b. *to look for a possible explanation for the differences in comprehension across mediums*
c. *to prove a point that reading digitally is more efficient*
d. *to help convince school districts to switch to e-textbooks*

If the participant selected the correct answer (to look for a possible explanation for the differences in comprehension across mediums), they heard the following: "That is correct! Researchers looked at the speed of processing for print and digital texts in order to try and unearth why comprehension was often superior in print."

If the participant selected an incorrect answer, they heard the following: "That is incorrect. Researchers looked at the speed of processing for print and digital texts in order to try and unearth why comprehension was often superior in print."

MODULE IV: Judgment of Performance and Calibration *(6 minutes)*

Part I: Overview of JOPs and Calibration

Part I explained the speed-accuracy tradeoff and its implications for JOPs and calibration accuracy (Wickelgren, 1977). The importance of JOPs and calibration accuracy for task performance was overviewed. Next, the training module overviewed the ways in which JOPs and calibration have been measured and calculated. It highlighted how previous studies in the lab (e.g., Singer et al., 2017a, Singer Trakhman et al., 2018a) calculate calibration accuracy by taking the absolute value of the difference between JOP and comprehension score. This was illustrated with examples that show the steps taken to calculate a calibration accuracy score, highlighting that the lower the calibration score, the more calibrated the learner.

Part II: Performance Feedback

After learning about JOPs and calibration, the participant was asked the following question:

How are calibration accuracy scores calculated?

a. *JOP divided by comprehension score = calibration*
b. *JOP multiplied by comprehension score = calibration*
c. *Absolute value of the difference between JOP and comprehension score = calibration*
d. *Absolute value of JOP divided by comprehension score= calibration*

If the participant selected the correct answer (absolute value of the difference between JOP and comprehension scores = calibration) they heard the following: "That is correct! Calibration accuracy scores are formed by taking the absolute value of the difference between JOP scores and comprehension scores."

If the participant selected an incorrect answer, they heard the following: "That is incorrect. Calibration accuracy scores are formed by taking the absolute value of the difference between JOP scores and comprehension scores." The

training module will then show another example of how a calibration accuracy score is calculated.

MODULE V: Profiling Readers *(12 minutes)*

Part I: Reader Profiles

Sample videos were embedded in the training module that showed the four profiles found in Singer Trakhman et al., 2018a navigating digital text.

While the videos from the Go-Pro© are playing, the narrator described the profile being shown (i.e., Regulator, Plodder, Glider, or Sampler).

Part II: Performance Feedback

After learning about the four profiles found in Singer Trakhman et al., 2018a, participants were shown a video of a participant navigating text in a manner that is consistent with the Plodder profile and were asked the following question:

Which profile was represented in the video?

a. Regulator

b. Plodder

c. Glider

d. Sampler

If the participant selected the correct answer, they heard the following: "That is correct! The slow, but linear movements shown in the video are indicators that this is a Plodder".

If the participant selected an incorrect answer, they heard the following: "That is incorrect. The slow, but linear movements shown in the video are indicators that this is a Plodder". The slide also showed the video of the Plodder reading again so they could see the slow, linear movements again.

APPENDIX E

Study II Materials and *Answers*

VOLCANOES PRETEST

1. Define the following terms[1]:

   a. Volatiles

   Mobiles substances under very high pressure that form gases and liquids at or near Earth's surface.

   b. Thermal Advection

   The process of heat transfer from one part of the Earth's interior to another by moving magma.

   c. Solfataras

   An opening of a volcano that produces sulfur.

2. What percent of the world's potentially active volcanoes reside in the "Ring of Fire"?

   a. 44%

   b. 55%

   c. 66%*

   d. 77%

3. Which of the following substances can form gas and liquids at or near Earth's surface, creating magma? Select all that apply.

   a. H_2O*

   b. CO_2*

   c. SO_2*

   d. Sea Water*

4. Which type of rock has the lowest melting point?

   a. Granitic and silicic metamorphic rocks*

   b. Rocks found within the mantle

   c. Sandstone and Shale Sedimentary rocks

   d. Diorite and Basalt Igneous rocks

[1]Each correct question is worth 1 point.

*Indicates the correct answer.

VOLCANOES COMPREHENSION TEST

20 points possible

Using the space provided, answer the following questions about the volcanoes passage.

1. List and describe one factor that contributes to the production of magma.

two points possible - (text only)

Decompression – natural reduction in pressure causes plates to move and magma to melt.

Subduction – movement in ocean causes the diving plate to be hot enough to release large amounts of seawater, causing volatiles.

2. Why does mafic material **NOT** melt when it is located in the Earth's mantle?

two points possible - (text only)

It is under too much pressure from overlying rocks to melt.

3. According to the information visually presented in the passage, name three locations that are **NOT** included in the "Ring of Fire" that produce volcano clusters. (visual only)

three points possible - (visual only)

East Africa

West Africa

Mid-Atlantic Ridge

North of Australia

Bottom of North America

West South American coast

4. Where is the Cascadia subduction zone located?

two points possible – (visual only)

Western coast of United States

5. How do volatiles lead to volcanic activity? Explain the process by which this occurs.

four points possible – (text+visual)

Volatiles are mobiles substances under very high pressure that form gas and liquids at or near the Earth's surface. These volatiles form "fluids" that exist under the Earth in between a gas and a liquid. As the volatile escapes the mantle wedge, rock can melt more easily, causing magma to escape the volcanoes solfatara.

6. What often occurs at the mid-ocean ridge system? What features cause this to happen? (text+ visual)

   three points possible – (text+visual)

   Earth quakes – this is where the ring of fire is

   It is the perfect storm of continental rifts meeting plate boundaries that separate

7. List up to 4 additional facts or ideas that you remember from the passage.

   four points possible

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus paper. *Journal of Experimental Psychology: Applied, 17*(1), 18-32.

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*(5), 364-373.

Ainsworth, S. (2008). How do animations influence learning? In Robinson, D., & Schraw, G. (Eds.), *Current Perspectives on Cognition, Learning, And Instruction: Recent Innovations in Educational Technology That Facilitate Student Learning* (pp.37-67). Charlotte, NC: Information Age Publishing.

Ainsworth, S., & Loizou, A. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science*, *27*(4), 669-681.

Akbar, R., Al-Hashemi, A., Taqi, H., & Sadeq, T. (2013). Efficacy of learning: Digital sources versus print. *Journal of Education and Practice, 4*(8), 98-114.

Alexander, P. A., (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M. L. Maehr & P.R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 213-250). Greenwich, CT: JAI Press.

Alexander, P. A. (2003). Profiling the developing reader: The interplay of knowledge, interest, and strategic processing. In C. M. Fairbanks, J. Worthy, B. Maloch, J. V. Hoffman, & D. L. Schallert (Eds.), *The fifty-first yearbook of the National Reading Conference*. Oak Creek, WI: National Reading Conference.

Alexander, P. A. (2006). *Psychology in learning and instruction*. Prentice Hall.

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction, 24*, 1-3.

Alexander, P. A., & Fox, E. (2011). Adolescents as readers. In M.L. Kamil, P.D. Pearson, E.B. Moke, & P.P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pp. 157-176). New York, NY: Routledge/Taylor & Francis.

Alexander, P. A., Grossnickle Peterson, E. M., Dumas, D., & Hattan, C. (2018). A retrospective and prospective examination of cognitive strategies and academic development: Where have we come in twenty-five years? In A. O'Donnell (Ed.), *Oxford handbook of educational psychology*. New York: Oxford University Press. Online Publication Date: May 2018

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III* (pp. 285-310). Mahwah, NJ: Lawrence Erlbaum Associates.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, *58*(4), 375-404.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research, 61*(3), 315-343.

Alexander, P. A., & Singer Trakhman, L. M. (2017, October). The enduring power of print for learning in a digital world. *The Conversation.* https://theconversation.com/the-enduring-power-of-print-for-learning-in-a-digital-world-84352

Alvermann, D. E., & Wilson, A. A. (2011). Comprehension strategy instruction for multimodal texts in science. *Theory into Practice*, *50*(2), 116-124.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. *Handbook of reading research*, *1*, 255-291.

Ariasi, N., Hyönä, J., Kaakinen, J. K., & Mason, L. (2017). An eye-movement analysis of the refutation effect in reading science text. *Journal of Computer Assisted Learning, 33*(3), 202-221.

Ariasi, N., & Mason, L. (2011). Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional Science, 39*(5), 581-601.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia?. *Journal of Educational Psychology*, *96*(3), 523-535.

Azevedo, R., Guthrie, J. T., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research*, *30*(1-2), 87-111.

Baker, L., Dreher, M. J., & Guthrie, J. T. (Eds.). (2000). *Engaging young readers: Promoting achievement and motivation*. Guilford Press.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis.* New York: NY: Cambridge University Press.

Barnett, J. E., & Hixon, J. E. (1997). Effects of grade level and subject on student test score predictions. *The Journal of Educational Research*, *90*(3), 170-174.

Bartholomé, T., & Bromme, R. (2009). Coherence formation when learning from text and pictures: What kind of support for whom?. *Journal of Educational Psychology*, *101*(2), 282-293.

Benedetto, S., Drai-Zerbib, V., Pedrotti, M., Tissier, G., & Baccino, T. (2013). E-readers and visual fatigue. *PloS one, 8*(12), 1-6.

Biancarosa, G., & Snow, C. E. (2004). *Reading next: A vision for action and research in middle and high school literacy: A report from Carnegie Corporation of New York*. Alliance for Excellent Education.

Biggs, J. B. (1978). Individual and group differences in study processes. *British Journal of Educational Psychology*, *48*(3), 266-279.

Block, D. (2009). *Second language identities*. Bloomsbury Publishing.

Bodmann, S. M., & Robinson, D. H. (2004). Speed and performance differences among computer-based and paper-pencil tests. *Journal of Educational Computing Research*, *31*(1), 51-60.

Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, *20*(5), 372-382.

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education*, *69*(2), 133-151.

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice. acheivement level. and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73*(4), 269-290.

Bryan, T., & Bryan, J. (1991). Positive mood and math performance. *Journal of Learning Disabilities, 24*(8), 490-494.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*(1), 60-77.

Burton, R. F. (2001). Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers. *Assessment & Evaluation in Higher Education, 26*(1), 41-50.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*(1), 182-197.

Butler, D. K. (1998). Metacognition and learning disabilities. In B. Y. L. W. (Ed.) (Ed.), *Learning about learning disabilities* (2nd ed., pp. 227-307). New York: Academic Press.

Carrell, P. L. (1985). Can reading strategies be successfully taught?. *Australian Review of Applied Linguistics, 21*(1), 1-20.

Cassidy, J., & Cassidy, D. (2007). *What's hot, what's not for 2007*. Reading Today, 24(4), 1-12.

Catrysse, L., Gijbels, D., Donche, V., De Maeyer, S., Lesterhuis, M., & Van den Bossche, P. (2018). How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *British Journal of Educational Psychology, 88*(1), 118-137.

Chen, P., & Zimmerman, B. (2007). A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *The Journal of Experimental Education*, *75*(3), 221-244.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (Eds.). (2008). *Handbook of research in new literacies*. Mahwah, NJ: Lawrence Erlbaum Associates.

College Board. (2009). *The Scholastic Aptitude Test.* New York, NY: Author.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *102*(3), 687-700.

DeStefano, D., & LeFevre, J. A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*(3), 1616-1641.

Dillon, T. W., & Emurian, H. H. (1995). Reports of visual fatigue resulting from use of a video display unit. *Computers in Human Behavior, 11*(1), 77-84.

Dinsmore D. L., & Alexander., P. A. (2016). A multidimensional investigation of deep-level and surface-level processing. *Journal of Experimental Education: Learning, Instruction, and Cognition, 84*(2), 213-244.

Dinsmore, D. L., & Alexander, P. A. (2012). A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. *Educational Psychology Review*, *24*(4), 499-567.

Dinsmore, D., Alexander, P. A., Loughlin, S. (2008), Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Researcher, 20*(4)*,* 391-409.

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, *24*, 4-14.

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction*, *22*(3), 293-331.

Dougherty, M. R., & Franco-Watkins, A. M. (2003). Reducing bias in frequency judgment by improving source monitoring. *Acta Psychologica, 113*(1), 23-44.

Dougherty, M. R., & Sprenger, A. (2006). The influence of improper sets of information on judgment: How irrelevant information can bias judged probability. *Journal of Experimental Psychology: General, 135*(2), 262-281.

Dundar, H., & Akcayir, M. (2012). Tablet vs. paper: The effect on learners' reading performance. *International Electronic Journal of Elementary Education, 4,* 441-450.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228-232.

Dyson, M., & Haselgrove, M. (2000). The effects of reading speed and reading patterns on the understanding of text read from screen. *Journal of Research on Reading, 23*(2), 210-223.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, *52*(4), 551-565.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2009). G* Power 3: A flecible statistical power analysis program for the social, behavior, and biomedical sciences. *Behavior Research Methods, 41*(4), 1149-1160.

Fischhoff, B., Slovic, P. & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552-564.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*, 906-911.

Franklin, M., & Zdonik, S. (1998). "Data in your face": Push technology in perspective. *ACM SIGMOD Record, 27*(2), 516-519.

Fried, C. B. (2008). In-class laptop use and its effects on student learning. *Computers & Education, 50*(3), 906-914.

Fry, E. (1968). A readability formula that saves time. *Journal of Reading, 11*(7), 513-578.

Forzani, E., & Maykel, C. (2013). Evaluating Connecticut students' ability to critically evaluate online information. *CARReader, 10,* 23-37.

Garland, K. J., & Noyes, J. M. (2004). Computer experience: a poor predictor of computer attitudes. *Computers in Human Behavior, 20*(6), 823-840.

Garner, R. (1987). *Metacognition and reading comprehension*. Ablex Publishing.

Gill, K., Mao, A., Powell, A. M., & Sheidow, T. (2013). Digital reader vs. print media: The role of digital technology in reading accuracy in age-related macular degeneration. *Eye, 27*(5), 639-643.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116*(2), 119-136.

Goldman, S. R. (2003). Learning in complex domains: When and why do multiple representations help?. *Learning and Instruction*, *13*(2), 239-244.

Goldman, S. R. (2009). Explorations of relationships among learners, tasks, and learning. *Learning and Instruction*, *19*(5), 451-454.

Goldstone, B. P. (2004). The postmodern picture book: A new subgenre. *Language Arts*, *81*(3), 196–204.

Graham, S., & Harris, K. R. (1989). Improving learning disabled students' skills at composing eassys: Self-instructional strategy training. *Exceptional Children, 56*(3), 201-214.

Gutierrez, A. P. (2012). Enhancing the calibration accuracy of adult learners: A multifaceted intervention. *UNLV Theses, Dissertations, Professional Papers, and Capstones.*

Hacker, D. J. (1998) Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacogition in educational theory and practice* (pp. 1-23). Mahwah, NJ: Erlbaum.

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, *3*(2), 101-121.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction*, *21*(4), 209-249.

Holsanova, J. (2010). Reception of multimodality: Applying eye tracking methodology in multimodal research. *Routledge handbook of multimodal analysis*, 285-296.

Hyönä, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. *Learning and Instruction, 14*(2), 131-152.

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: towards a model of real-world reading. *Journal of Computer Assisted Learning*, *33*(3), 193-201.

Kamil, M. L. (2004). Vocabulary and comprehension instruction: Summary and implications of the National Reading Panel findings. *The voice of evidence in reading research*, 213-234.

Kendeou, P., Bohn-Gettler, C., White, M. J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading, 31*(3), 259-272.

Kendeou, P., & Van Den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, *97*(2), 235-245.

Kendeou, P., Van Den Broek, P., White, M. J., & Lynch, J. (2007). Comprehension in preschool and early elementary children: Skill development and strategy interventions. *Reading comprehension strategies: Theories, interventions, and technologies,* 27-45.

Kerr, M. A., & Symons, S. E. (2006). Computerized presentation of text: Effects on children's reading of informational material. *Reading and Writing, 19*(1), 1-19.

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

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

Klassen, R. (2002). Writing in early adolescence: A review of the role of self-efficacy beliefs. *Educational Psychology Review, 14*(2), 173-203.

Kleeck, A. V. (2008). Providing preschool foundations for later reading comprehension: The importance of and ideas for targeting inferencing in storybook-sharing interventions. *Psychology in the Schools*, *45*(7), 627-643.

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 595-608.

Kress, G. (2003). *Literacy in the new media age*. London: Routledge.

Kurniawan, S.H., Zaphiris, P. (2001). Reading Online or on Paper: Which is Faster? Abridged Proceedings of the 9th International Conference on Human Computer Interaction, 220-222.

Larking, J. H., & Simon, H. A. (1987). Why a diagram is sometimes worth ten thousand words. *Cognitive Science, 11*(1), 65-99.

Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitundinal data. *Merrill-Palmer Quarterly*, *52*(3), 377-389.

Lee, D.-S., Ko, Y.-H., Shen, I.-H., & Chao, C.-Y. (2011). Effect of light source, ambient illumination, character size and interline spacing on visual performance and visual fatigue with electronic paper displays. *Displays, 32*(1), 1-7.

Lenhard, W., Schroeders, U., & Lenhard, A. (2017). Equivalence of screen versus print reading comprehension depends on task complexity and proficiency. *Discourses Processes, 54*(5-6), 427-445.

Leslie, L., & Caldwell, J. (1995). *Qualitative Reading Inventory-Ii (Qri-Ii)*. Harper Collins College Publishers.

Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The new literacies of online research and comprehension: Rethinking the reading acheivement gap. *Reading Research Quarterly, 50*(1), 37-59.

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2017). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. *Journal of Education*, *197*(2), 1-18.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *ECTJ, 30*(4), 195-232.

Levine, L. E., Waite, B. M. & Bowman, L. L. (2007). Electronic media use, reading, and academic distractibility in college youth. *CyberPsychology & Behavior, 10*(4), 560-566.

Lin, L. M., Moore, D., & Zabrucky, K. M. (2001). An assessment of students' calibration of comprehension and calibration performance using multiple measures. *Reading Psychology, 22*(2), 111-128.

Lin, L. M., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, *23*(4), 345-391.

Lindblom-Ylänne, S., & Lonka, K. (1998). Individual ways of interacting with the learning environment—are they related to study success?. *Learning and Instruction*, *9*(1), 1-18.

Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, K., & Cook, D. (2009). Using standards to improve middle school students' accuracy at evaluating the quality of their recall. *Journal of Experimental Psychology: Applied*, *15*(4), 307-318.

Liu, A., Aamodt, S., Wolf, M., Gelernter, D., & Mark, G. (2009, October 14). Does the brain like e-books? *The New York Times*, 17.

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, *82*(6), 1778-1787.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differnces in absolute and relative metacomprehension activity. *Journal of Educational Psychology, 97*(4), 723-731.

Mangen, A. (2010). Point and click: Theoretical and phenomenological reflections on the digitization of early childhood education. *Contemporary Issues in Early Childhood, 11*(4), 415-431.

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

Margolin, S. J., Driscoll, C., Toland, M. J., & Kegler, J. L. (2013). E-readers, computer screens, or paper: Does reading comprehension change across media platforms? *Applied Cognitive Psychology, 27*(4), 512-519.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology*, *46*(1), 4-11.

Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, *41*(1), 172-187.

Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85-139). Academic Press.

Mayer, R. E. (2009). Based Solutions to Three Problems in Web-Based Training. In *Learning, training, and development in organizations* (pp. 225-249). Routledge.

McNamara, D. S. (Ed.). (2012). *Reading comprehension strategies: Theories, interventions, and technologies*. Psychology Press.

McNamara, D. S., de Vega, M., & O'Reilly, T. (2007). Comprehension skill, inference making, and the role of knowledge. *Higher Level Language Processes in the Brain: Inference and Comprehension Processes,* 233-251.

Meyer, B. J. F., & Ray, M. N. (2011) Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education, 4*(1), 127-152.

Meyer, B. J. F., Young, C. J., & Bartlett, B. J. (1989). *Memory improved: Enhanced reading comprehension and memory across the life span through strategic text structure.* Hillsdale, NJ: Lawrence Erlbaum.

Millar, M. & Schrier, T. (2015) Digital or printed textbooks: Which do students prefer and why?, *Journal of Teaching in Travel & Tourism*, *15*(2), 166-185.

Moore, P. J., & Scevak, J. J. (1997). Learning from texts and visual aids: A developmental perspective. *Journal of Research in Reading*, *20*(3), 205-223.

Murphy, L., & Tenenberg, J. (2005). Do computer science students know what they know?: A calibration study of data structure knowledge. *ACM SIGCSE Bulletin*, *37*(3), 148-152.

National Assessment Governing Board (2009). *Reading Framework for the 2009 National Assessment of Educational Progress.* Washington, DC: Author.

National Center for Education Statistics (2013).*The Nation's Report Card: A First Look: 2013 Mathematics and Reading, (NCES 2014-451).* Retrieved from Washington, D.C.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109-133.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science*, *2*(4), 267-271.

Nietfeld, J. L., & Schraw. G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research, 95*(3), 131-142.

Norman, E., & Furnes, B. (2016). The relationship between metacognitive experiences and learning: Is there a difference between digital and non-digital study media? *Computers in Human Behavior, 54*, 301-309.

Noyes, J., Garland, K., & Robbins, L. (2004). Ppaer-based versus computer-based assessment: Is workload another test mode effect? *British Journal of Educational Technology, 35,* 111-113.

Olshavsky, R. W. (1979). Task complexity and contingent processing in decision making: A replication and extenstion. *Organization Behavior and Human Performance, 24*(3), 300-316.

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences, 106*(37), 15583-15587.

Pajares, F., & Kranzel, J. (1995). *Competence and confidence in mathematics: The role of self-efficacy, self-concept, anxiety, and ability.* Paper presented at the meeting of the American Educational Research Association, San Francisco.

Pajares, F. & Miller, M. D. (1994). The role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193-203.

Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices - Past, present and future. *Children's reading comprehension and assessment*, 13-69.

Pipkin, B., Trent, D. D., Hazlett, R., & Bierman, P. (2014). *Geology and the Environment* (7th ed.). Belmont, CA: Brooks/Cole.

Pöchhacker, F. (2010). The role of research in interpreter education. *Translation & Interpreting*, *2*(1), 1-10.

Pressley, M. (1988). Comprehension strategies instruction. In J. Osborn & F. Lehr (Eds.), *Literacy for all: Issues in teaching and learning* (pp. 113–133). New York: Guilford.

Pressley, M. (1990). *Cognitive strategy instruction that really improves children's academic performance.* Cambridge, MA: Brookline Books.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 545–561). Mahwah, NJ: Erlbaum.

Proaps, A. B., & Bliss, J. P. (2014). The effects of text presentation format on reading comprehension and video game performance. *Computers in Human Behavior, 36*, 41-47.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Washington, DC: RAND Education.

Rapp, D. N., Broek, P. V. D., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, *11*(4), 289-312.

Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*(4-5), 559-579.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457-1506.

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin*, *137*(1), 131-148.

Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M2: Media in the lives of 8-to 18-year-olds*. Oakland, CA: Henry J. Kaiser Family Foundation.

Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology, 88*(1), 174-188.

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., . . . … & Tenopir, C. (2008). The Google generation: The information behavior of the researcher of the future. *Aslib Proceedings, 60*(4), 290-310.

Rutherford, T. (2017). The measurement of calibration in real contexts. *Learning and Instruction*, *47*, 33-42.

Sarroub, L., & Pearson, P. D. (1998). Two steps forward, three steps back: The storm history of reading comprehension assessment. *The Clearing House, 72*(2), 97-105.

Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review, 14*(1), 101-120.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitie monitoring. *Metacognition and Learning, 4*(1), 33-45.

Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction*, *24*, 48-57.

Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition, 22*(1), 63-69.

Schunk, D. H. (2000). Coming to terms with motivation constructs. *Contemporary Educational Psychology, 25*(1), 116-119.

Sellen, A. J., & Harper, R. H. R. (2002). *Myth of the Paperless Office.* MIT Press.

Singer L. M., & Alexander, P. A. (2017a). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *The Journal of Experimental Education, 85*(1), 155-172.

Singer, L. M., & Alexander, P.A. (2017b). Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research, 87*(6), 1007-1041.

Singer Trakhman, L. M., Alexander., P. A., & Berkowitz, L. E. (2017). Effects of processing time on comprehension and calibration in print and digital mediums. *Journal of Experimental Education.* Advance online publication.

Singer Trakhman, L. M., Alexander, P. A., & Silverman, A. B. (2018a). Profiling reading in print and digital mediums. *Learning and Instruction, 57,* 5-17.

Singer Trakhman, L. M., Alexander, P. A., Sun, Y., & Silverman, A. B. (2018b). The effects of processing multimodal texts in print and digitally on students' comprehension and calibration. *Manuscript in preparation.*

Sprenger, A., & Dougherty, M. R. (2006). Differences between probability and frequence judgments: The role of individual differences in working memory capacity. *Organizational Behavior and Human Decision Processes, 99*(2), 202-211.

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, *49*(1), 77-96.

Van de Vijver, F. J., & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *Journal of Applied Psychology, 79*(6), 852-859.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategis of discourse comprehension.* Academic.

Van Meter, P. N., Cameron, C., & Waters, J. R. (2017). Effects of response prompts and diagram comprehension ability on text and diagram learning in a college biology course. *Learning and Instruction*, *49*, 188-198.

Verdi, M. P., Crooks, S. M., & White, D. R. (2014). Learning effects of print and digital geographic maps. *Journal of Research on Computing in Education, 35*(2)*,* 290-302.

Vermunt, J. D., & Donche, V. (2017). A learning patterns perspective on student learning in higher education: State of the art and moving forward. *Educational Psychology Review, 29*(2), 269-299.

Wallis, C. (2010). *The impacts of media multitasking on children's learning and development: Report from a research seminar*. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Wallsten, T. S., & Gu, H. (2003). Distinguishing choice and subjective probability estimate processes: Implications for theories of judgment and for cross-cultural comparisons. *Organizational Behavior and Human Decision Processes, 90*(1), 111-123.

Wästlund, E. (2007). *Experimental studies of human-computer interaction: Working memory and mental workload in complex cognition*. Goteborg: Department of Psychology, Gotenburg University, Sweden.

Wästlund, E., Reinikka, H., Norlander, T., & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. *Computers in Human Behavior, 21*(2), 377-394.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dyamics. *Acta Psychological, 41*(1), 67-85.

Wiederholt, J. L., & Bryant, B. R. (1992). *Gray Oral Reading Tests: GORT-3*. Austin, TX: Pro-ed.

Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Mahwah, NJ: Erlbaum.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551-572.

Winne, P. H., Jamieson-Noel, D., & Muis, K. (2002). Methodolical issues and advances in researching tactics, strategies, and self-regulated learning. *Advances in motivation and achievement: New directions in measures and methods, 12,* 121-155.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. (2001). *Woodcock-Johnson R III NU Tests of Achievement*. Itasca, IL: Riverside.

Yeh, L., Schwartz, A. I., & Baule, A. L. (2011). The impact of text-structure strategy instruction on the text recall and eye-movement patterns of second language English readers. *Reading Psychology, 32*(6), 495-519.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13-39).