THE JOURNAL OF EXPERIMENTAL EDUCATION
2023, VOL. 91, NO. 4, 599–620
https://doi.org/10.1080/00220973.2022.2092831

Routledge
Taylor & Francis Group

LEARNING, INSTRUCTION, AND COGNITION

The Effects of Processing Multimodal Texts in Print and Digitally on Comprehension and Calibration

Lauren M. Singer Trakhman, Patricia A. Alexander, and Yuting Sun

University of Maryland, College Park, MD, USA

ABSTRACT

The comprehension and calibration of 54 undergraduates were investigated as they read excerpts from an introductory geology textbook on weather and soil in print and digitally. All excerpts were approximately 1600 words in length and contained a graph, a diagram, and three photographs that complemented or extended the written text. Each student read two texts with medium and topic counterbalanced. Prior to reading, the students completed a demographic survey, rated their topic familiarity, and completed two topic knowledge pretests. They next read one chapter on either weather or soil in print or digitally and then answered a series of short-answer questions. The questions drew on content from the written text only, visuals only, or both. The same procedure was then repeated in the other medium. Analyses indicated processing multimodal texts in print was significantly more advantageous than processing those same texts digitally, and this difference was more pronounced for questions focused on visuals only. Students' self-rated topic familiarity was compared to their demonstrated topic knowledge for weather and soil and their predicted comprehension performance was compared to actual comprehension performance. Results showed that undergraduates' calibration was poor overall, but comprehension was overestimated more often when students read multimodal texts digitally.

KEYWORDS

multimodal text; digital reading; reading comprehension; calibration; cognitive processes/development; literacy; comprehension

AS A CONSEQUENCE OF today's rapid-paced society and ever-changing technologies, students at all educational levels are frequently called upon to learn from texts that differ not only in the medium of delivery (e.g., paper or digital) and genre (e.g., expository or narrative), but also in terms of their *modality* or the means by which information is encoded (e.g., word, images, or video; Coiro, 2021; Cromley et al., 2016; McNamara et al., 2012). So much of the text that students are called upon to read and comprehend when learning are multimodal in nature (Firetto & Van Meter, 2018; List et al., 2020; Serafini, 2011). *Multimodal texts* not only rely on written language to convey critical content but also frequently incorporate pictures, tables, graphs, or photographs that either complement or extend the information presented linguistically (Lauer, 2009; Liu, 2013; Mayer, 2002).

Multimodal Texts and Comprehension

Although reading multimodal texts is an everyday experience for today's students, the effects of such materials on comprehension remain an open issue. On the one hand, there are those who

CONTACT Patricia A. Alexander palexand@umd.edu Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD 20742-1131, USA.

see advantages for multimodality in general or for certain learner populations, such as for English language learners or others who struggle with the processing of written language (Rose & Dalton, 2009; Serafini, 2011). For these researchers, multimodal texts can afford learners alternative pathways to acquiring critical content through informative visuals or graphic displays. Yet, there is a substantial literature showing that many students are ill-prepared to make sense of the graphs, maps, diagrams, or other informational visuals they are presented (Canham & Hegarty, 2010; Cromley et al., 2016).

When those visuals are embedded in texts, there are added processing demands. One complicating factor that researchers have chronicled are students' difficulties in constructing an integrated and coherent understanding of content represented in both the text and accompanying visuals (Alvermann & Wilson, 2011; Bartholomé & Bromme, 2009; Kress & Van Leeuwen, 1996; Van Leeuwen, 2015). Even with these added demands, the empirical findings pertaining to comprehension performance for multimodal texts are mixed. For example, Jewitt (2008) reported improved comprehension when visuals were included in texts, whereas Cromley et al. (2016) and Schwamborn et al. (2011) found that students struggled to interpret the meaning of visuals and to build an integrated representation of textual and visual content. The picture becomes even more complicated when the form of comprehension assessment is considered. For instance, in their meta-analysis of multimodal texts, Guo et al. (2020) concluded that visuals such as pictures, diagrams, and flow or procedural charts were generally facilitative of comprehension when assessments consisted of true/false questions. In contrast, when comprehension was assessed by more open-ended items or constructed-response questions, the effects of visuals were again mixed. Guo et al. also reported differential outcomes depending on the types of visuals that had been examined, with pictures being more facilitative to comprehension than diagrams or other forms.

Eye-tracking studies have also provided clues about how readers differ in the attention they afford to visuals in multimodal texts and the way that their eyes move between the visuals and related text (Scheiter et al., 2019). In one such study, Mason et al. (2013b) used eye-tracking to examine fourth graders' processing of multimodal science texts. Employing person-centered analysis, they found that significantly higher comprehension occurred for those students who fixated longer on the prominent illustration on the page when initially encountered and for those whose eyes more frequently moved back and forth between relevant text and the illustration, suggesting greater effort toward integration of pictorial and textual content. Mason et al. (2013a) also used eye-tracking to test whether a science text with a more concrete visual (i.e., skier racing down an incline), a more abstract visual (i.e., generic plane and incline), or no visual at all would affect high-school students' comprehension differentially. They found that having either the more concrete or abstract illustration embedded in the science text was superior to the unillustrated text on the students' comprehension measured immediately after reading. However, at delayed post testing, the researchers found that only students receiving the more concrete picture significantly outperformed students in the no-illustration condition.

Further, what Mayer (2002; Moreno & Mayer, 1999) and others (O'Neil, 2011; Serafini, 2011) investigating multimodal texts have demonstrated is that the placement of visuals in relation to associated textual content can either support or interfere with comprehension. In fact, several of Mayer's cognitive principles of multimedia learning (Mayer & Fiorella, 2014) expressly address this issue of visual-text alignment and integration. For example, the principle of spatial contiguity states that visuals should be presented in close proximity to the associated text to enhance learning. These principles have influenced multimodal investigations, including with the eye-tracking studies by Mason et al. (2013a, 2013b) in which the authors positioned the visual relative to the text to maximize spatial contiguity.

Although the growing literature on multimodal texts has enhanced what is known about learning from text containing visuals, significant limitations of this work remain for understanding how students learn from academic texts under more typical conditions. For one, most of the research examining the effects of multimodal text on comprehension often involve quite short texts with a carefully generated visual that has been intentionally positioned. For example, the inclined plane passage that Mason et al. (2013a) developed for their multimodal study was only 243 words in length, and another text illustrating how air occupies space used in Mason et al. (2013b) was just 219 words long. It should be noted the Mason, Tornatora, et al. study involved fourth graders, which may have been a consideration in the shorter texts the researchers used. That being said, Mason et al. (2013a) used materials of similar length for eleventh graders whom we would describe as more mature readers capable of processing appreciably longer texts. As noted, the placement of the visual in these studies was informed by Mayer's (2002) principles of multimedia design. Even when texts are short and visuals are more intentional positioned, many students do not appear to recognize the informative nature of visuals nor not put much effort into integrating textual and visual content (Van Meter et al., 2017). We acknowledge that some multimedia researchers have used texts of over 1000 words made to resemble textbook readings (Scheiter et al., 2014, 2019). However, studies employing these more extended texts are not the norm within this literature.

Of particular concern to the present investigation, the orchestrated nature of the multimodal materials used in the aforementioned research appears to have little bearing on the materials that students are required to process routinely. Rather, the expository texts that are standard fare in academic domains (e.g., history or biology) from upper elementary through college are not only appreciably longer, but also include multiple visuals of varying types (e.g., diagrams, maps, or pictures). Moreover, the placement of these visuals does not appear to conform to any theoretical or empirically-derived principles (Canham & Hegarty, 2010; Serafini, 2011; Serafini & Ladd, 2008; Unsworth & Cléirigh, 2014).

The previously cited studies by Scheiter et al. (2014, 2019) again varied from this trend in that these researchers attempted to construct materials that resembled the exposition their participants would typically read in their courses. However, we did not encounter studies within the multimedia literature that used longer and unmodified or unmanipulated texts when studying multimedia principles forwarded by Mayer (2002). Consequently, it cannot be presumed that findings from the multimedia literature readily generalize to the conditions that exist in typical academic contexts. If anything, it might be argued that the outcomes reported in much of the multimodal literature are apt to be more optimistic than what would likely be unearthed under more ecologically valid conditions.

There is one more shortcoming of the existing multimodal literature that warrants attention. Specifically, there is a paucity of research that has systematically investigated whether the medium of delivery of multimodal texts has any bearing on what students understand or remember from reading. In effect, the text-processing landscape framed by multiple modalities and multiple mediums remains virtually unexplored.

Multiple Mediums and Text Comprehension

From the research we have been conducting for the past several years, we have garnered a rich and quite consistent picture of the role that medium plays in college students' comprehension of academic texts delivered on paper or on screen (Singer Trakhman et al., 2019; Singer & Alexander, 2017a, 2017b). As part of this research program, we have explored the effects of medium while systematically varying certain text elements such as the academic domain (e.g., psychology and earth science), text length (e.g., 450 to 1800 words), comprehension focus (e.g., main idea or points and supporting details), and question format (e.g., multiple choice and short constructed response items). We have also investigated the medium effects by means of variable-centered and person-centered analyses and have incorporated behavioral data gathered during text processing (Singer Trakhman et al., 2018). Regardless of the variations we have implemented, clear trends have been documented. Those consistent trends include the following:

1. Medium effects are more pronounced with longer texts (500 words or more).
2. College students strongly prefer to read digitally rather than in print.
3. Undergraduates judge their comprehension performance as higher after reading and answering questions digitally than when reading and answering questions on paper.
4. Despite their predictions, comprehension performance for the college students is significantly higher in the print than digital condition for all but gist or main idea questions for which there is no medium effect.
5. Students tend to read faster when the text is on screen rather than on paper.
6. The significant gap between predicted and actual performance when it comes to medium indicates that these college students are not effectively monitoring their performance during reading or when answering questions.

However, while the naturally-occurring texts used in the aforementioned investigations have typically been multimodal in character, with diagrams, maps, and photographs accompanying the written text, we have not expressly factored in those multiple representations in our research designs, comprehension assessments, or analyses. Thus, we had not considered whether the processing of multimodal texts under print or digital conditions would differentially affect students' understanding of the content conveyed in the written text or in accompanying visuals, or their ability to self-monitor their processing or comprehension effectively.

Mediums and Multimodality

In the current study, we sought to extend our prior investigations of the effects of print and digital mediums on students' comprehension and calibration (i.e., accuracy of performance self-judgments) by expressly considering the visual, as well as linguistic, content of the reading materials. We were aware that this consideration of the visual-text interface deviated from much of the modality literature due to: (a) the use of lengthy naturally-occurring texts; (b) the inclusion of multiple and diverse visuals; and (c) the placement of visuals without explicit regard to principles of multimedia learning (Mayer, 2002; Moreno & Mayer, 1999). However, we viewed the opportunity to examine what students comprehend from the integration of textual *and* visual content, as well as the accuracy of their performance judgments, as informative to both the multimodal and multiple mediums literatures.

For one, it would explore visual-text integration under less constrained, albeit less controlled, conditions. We recognized from the outset that there were cost-benefits to the more ecologically valid design of this study. However, understanding better how students process the multimodal materials they are regularly assigned in their coursework was worth the loss of controls that have been frequently imposed in the literature. For another, the current investigation would introduce another element of complexity to prior print-versus-digital investigations that has not been given adequate attention. Specifically, by assessing what students recall from the textual and visual content, we may better understand the differences in comprehension performance that have arisen across print and digital mediums, and we might learn more about how students' self-monitoring (or lack thereof) may be contributing to those differences.

Visuals and Text Comprehension

For the present investigation, we had undergraduates read selections from a college textbook under both print and digital conditions, as we have done in prior studies. The students were told that they would be ask questions about the readings The readings we chose were from chapters on soil and weather in an introductory textbook on *Geology and the Environment* (Pipkin et al., 2014). In selecting these segments, we aimed for comparability on critical characteristics without manipulating the text or visuals in any fashion. These characteristics, detailed in the Method section, included the number and types of visuals (i.e., a graph, a diagram, and three photographs), as well as word length (i.e., 1602 for weather and 1658 for soil).

When it came to assessing students' comprehension of what was read, we focused on different question levels that had proven informative in the past. What we altered for the current study was the source of that information, with one detail question relying on content found only in the written text or only in the visuals, as well as two questions where the targeted content could be found in both the textual or the visual content. By constructing the post-reading comprehension measure in this way, we would have some evidence of whether students were attending carefully to the visuals in the chapter segments.

Calibration

Calibration or individuals' ability to monitor their understanding or judge their performance accurately has a long history within the text processing literature (Delgado & Salmerón, 2021; Glenberg et al., 1985). It has been fairly well documented that those who monitor their performance more effectively during reading or studying do better on measures of comprehension than those who are less effective at monitoring (Dunlosky & Thiede, 2013; Nietfeld & Schraw, 2002). It has been argued that those who are poorly calibrated—leading to an overestimation of what they understand—are less likely to invest the time or strategic energy required to improve their overall performance (Hattie, 2013). Factors such as students' age, text complexity, and prior knowledge have also been shown to effect calibration accuracy and, thus, comprehension performance (Lin et al., 2002).

In our past studies comparing print and digital mediums (e.g., Singer Trakhman et al., 2019; Singer & Alexander, 2017a), calibration accuracy was generally not good for students but significantly poorer when they read on screen. Other researchers have also reported that students are less well calibrated when reading on screen than on paper (Delgado & Salmerón, 2021). Because students in our studies read excerpts from the same source under both conditions, factors like text complexity, prior knowledge, or age could not account for this consistent difference. What we came to regard as potential contributors to lower calibration in the digital condition were speed or ease of processing (Ackerman & Goldsmith, 2011; Koriat et al., 2006), judgments of learning (Nietfeld & Schraw, 2002), and scrolling behaviors (Singer & Alexander, 2017a).

For example, in Koriat et al. (2006) investigation, the less effort that participants exerted in task performance, the higher their judgments of learning. When students move through a digital text more rapidly than a comparable text read in print, they may assume that this differential speed signaled greater ease of processing. In fact, this faster processing rate meant that students had less time to reflect on their level of understanding, which could contribute to miscalibrations. Wickelgren (1977) referred to this as the speed-accuracy tradeoff hypothesis. Delgado and Salmerón (2021) felt that students engage in more shallow information processing when reading on screen than on paper when time is an issue. Of course, it is possible that students' preference for reading online and the time they spend on screens (Singer & Alexander, 2017b) contributes to greater confidence in their performance in the digital versus print condition. However, whether these findings for calibration hold when multimodality interfaces with multiple mediums remain to be seen.

The Present Study

For the current investigation, we continued to target college students' comprehension of naturally-occurring texts read both in print and digitally. The principal difference in this study was the added focus on multimodal elements of the text segments in comprehension assessment. Because there is evidence that multiple representations can increase cognitive demands (Cromley et al., 2016; McNamara et al., 2012), we wanted to test their effects with more competent readers who are likely more equipped to deal with the added complexity than younger or struggling readers. Further, we incorporated an additional calibration measure in this study that centered on topic knowledge. The rationale for that addition was the emphasis placed on students' existing topic knowledge as a significant predictor of their comprehension performance (McCarthy & McNamara, 2021). Thus, calibration in the current study pertained to college students' ability to accurately predict their topic knowledge *prior to* reading, as well as their comprehension performance *after* reading.

Research questions

The specific research questions we sought to address in the present study were:

1. How accurate are students' perceptions of their topic knowledge compared to what they demonstrated on the topic knowledge pretests?

Informed by the literature on calibration (Delgado & Salmerón, 2021; Hattie, 2013; Singer & Alexander, 2017b), we hypothesize that undergraduate students in this investigation would inaccurately predict what they know about weather or soil—the topics about which they would be reading. However, we expected that students would judge their existing knowledge of weather higher than their knowledge of soil, given that weather seems to be a more common topic in and out of school.

2. To what extent are students able to respond to comprehension questions that require information only from the text, only from the visuals, or from both the textual and visual content?

What we expected in terms of the comprehension questions varying by source (text, visual, or both) was that students' performance would be best on those items where the relevant content could be found in the written text and in the visuals. Conversely, we were less confident about students' attention to relevant content found only in the visuals. We grant that findings on the effectiveness of visuals on comprehension have been mixed (Guo et al., 2020). Yet, in multimodal studies employing eye-tracking technology with quite short texts containing carefully positioned visuals, there were those clusters of students who attended only superficially to the visual content (Mason et al., 2013a, 2013b). Therefore, in the present study that includes multiple visuals embedded within an extended text for the reading task, we predict that the visual-only questions will be less correctly answered than the text-only or text-visual questions.

3. What effect does medium have on undergraduates' overall comprehension when processing multimodal texts and on their performance for questions differing in specificity (e.g., main idea or details) and by source (e.g., text-only or visual-only)?

Based on prior medium studies (e.g., Alexander & Singer, 2017; Singer Trakhman et al., 2018), we hypothesize that students' overall comprehension performance for both multimodal texts would be significantly better in the print condition than in the digital condition. This predicted outcome is based on the consistent strength of print over digital processing for longer and complex expository texts. Thus, given the added dimension of multimodality, there is reason to assume that the increased complexity of the visual elements would only strengthen the advantage for the print over digital medium. Further, when it comes to questions differing by specificity, we would expect the advantage for print to hold for the detail but not the main questions, as has been the pattern in prior studies (Singer Trakhman et al., 2019). For questions differing by source, however, there is insufficient history to forward any prediction as to medium effects.

4. To what extent does the medium of delivery influence the accuracy of students' comprehension judgments when processing multimodal texts?

Finally, we would predict that students' calibration ability when it pertains to comprehension performance would be more discrepant for the multimodal texts read digitally than in print. Our rationale for this judgment rests on several factors. For one, the issue of faster reading times recorded for digital versus printed texts reported in prior studies would likely hold in the present study. This more rapid processing would mean less time for students to monitor their understanding. In addition, without intentional placement of visuals in these text, readers' toggling between the visual and textual content in the digital condition, which is important for building a coherent mental model of the information, could well be hampered.

Method

Participants

Participants for this study were 54 undergraduate students enrolled in human development and educational psychology courses at a large mid-Atlantic university. The students were predominantly female (77.8%) with a mean age of 20.78 ($SD = 1.32$) years. The self-reported races for students were White (50.0%), Asian/Pacific Islander (24.1%), Hispanic (7.4%), African American (7.4%), and Other/Multiracial (11.1%). These data on gender and race were comparable to the overall demographics of the university in which the study was conducted. In addition, the undergraduates in this study represented a variety of majors, mainly in the natural (66.7%) and social (31.5%) sciences, and years in school (3.7% freshmen, 9.3% sophomore, 31.5% junior, 53.7% senior, and 1.9% graduate students).

Experimental Texts

For this study, each participant read two multimodal texts: one in print and one digitally. The two texts were chapter excerpts from *National Geographic Learning* (Pipkin et al., 2014), a college textbook for an introductory geology course. The two excerpts covered the topics of weather and soil. In a pilot study, weather was rated as a more familiar topic than soil by undergraduate students. The texts were comparable on a number of dimensions: three pages in length; similar word counts (1602 for weather and 1658 for soil), and readability level (grade 14; Fry Readability: Fry, 1968). In terms of the visuals, both text excerpts contained one graph, one diagram, and two pictures. For the weather text, the graphic showed the global distribution of precipitation and the diagram depicted how the flow of air currents over a mountain range created rainfall on the ascending side and dry conditions on the descending side. The pictures were a montage of extreme weather events and their effects in the northeast US and in Texas. The two pictures included one of a geologist standing by a rock formation, while the second showed a layer of rich volcanic ash. For the soil text, the graphic was a map showing loess-covered regions of the earth, while the diagram showed an idealized soil layer profile and explained how it was formed.

Pre-reading Measures and Variables

Demographic Survey

Prior to beginning the study, students completed a demographic survey. The survey collected information on the participants' age, gender, ethnicity, mother language, grade point average, major, and year in college (e.g., freshman or sophomore).

Self-reported Topic Knowledge

To ascertain the relative novelty of the reading topics, participants were asked to rate their level of knowledge of the weather and soil topics. These ratings were made on a 100 mm scale that ranged from 0 (no knowledge) to 100 (expert). For example, if a student responding to the question "Rate your knowledge on the subject of weather" placed her mark at the 25 mm point, her self-reported topic knowledge score for weather would be recorded as 25. This same process was repeated for the soil topic.

Demonstrated Topic Knowledge Pretests

To see how much knowledge students had about the two geology topics prior to reading, they also completed a knowledge measure for each topic. This demonstrated topic knowledge measures consisted of three terms key to each reading that students were asked to define, along with three multiple-choice questions, one of which had two correct answers. Both the short-answer and the multiple-choice questions were scored as correct (1) or incorrect (0), except for the "select all that apply" multiple-choice question, which was scored as fully correct (2), one correct answer selected (1), or incorrect (0). Thus, each topic knowledge measure had a maximum score of seven. The definitions were scored independently by the first author and a research assistant resulting in a high rate of interrater agreement (.94). The demonstrated topic knowledge pretests are displayed in the Appendix.

Topic Knowledge Calibration

To examine how well students' self-ratings of their topic knowledge aligned with their demonstrated knowledge on the topic pretests, we created a *topic knowledge calibration* variable. We first converted the topic knowledge test scores and the self-reported knowledge ratings into percentages, so the two variables were on the same scale. Next, we took the absolute value of the difference between the percentage scores of the self-reported and demonstrated topic knowledge to represent the accuracy of topic knowledge calibration. For this variable, lower scores indicated higher calibration accuracy for topic knowledge.

Post-Reading Measures and Variables

Reading Comprehension Posttests (RCT)

We were interested in examining the effects of medium (i.e., print and digital) on students' comprehension overall and by level of specificity (main idea or details) and information source (text-only, visual-only, text-visual). Immediately after reading and with no text available, students responded to five short-answer questions for each topic presented in the same medium as the reading. The first question asked them to summarize the main idea of the excerpt. The remaining four questions relied on multiple pieces of information found only in the written text (1), only in the visuals (1), or in both (2), for a maximum score of 17 per topic. The comprehension posttests and the specific points assigned each question appear in the Appendix. To establish interrater agreement, the first author and a research assistant scored all responses independently and their scores were compared. Interrater agreement for the data set was high (.95). Any disagreements were resolved through discussion.

Judgment of Performance (JOP)

In order to assess students' overall calibration for this multimodal, multiple mediums task, we first had them judge their performance on the reading comprehension measure immediately after reading the printed text and again after reading the digital text or vice versa. Specifically, after reading of the weather (or soil) text and completing the comprehension measure in the print (or digital) condition, students answered the following question presented on paper (or online):

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

Comprehension Calibration

We examined the students' *comprehension calibration*—the relation between judgment of performance (JOP) and demonstrated performance on the comprehension measure (RCT)—both descriptively and quantitatively. For these analyses, it was first necessary to convert JOP ratings and RCT scores to a similar metric. Therefore, we transformed the JOP rating and the RCT score into percentages. Descriptively, we wanted to determine how many students accurately judged their performance (i.e., self-ratings within 3 percentage points of comprehension score) and how many underestimated (i.e., self-ratings more than 3 percentage points below comprehension score) or overestimated (i.e., self-ratings more than 3 percentage points above comprehension score) how well they comprehended in the print and digital conditions.

We created a comprehension calibration score for each participant for each medium that corresponded to the definition of calibration as the difference between predicted and actual performance (Alexander, 2013; Fischhoff et al., 1977; Glenberg et al., 1987). To produce that score, we took the absolute value of the difference in percentages between the JOP rating and RCT score for each condition. Because calibration represents the correspondence between predicted and actual performance, a lower calibration score indicated of better calibrated readers.

Procedure

The study was conducted outside of class in a designated area. Students completed the demographics survey, the self-reported topic knowledge measure, and demonstrated topic knowledge pretests digitally. Then, they moved on to the reading portion of the study. Each participant read two texts: one text from each medium (i.e., print and digital) concerning weather and soil. These texts were presented in counterbalanced order. Prior to receiving the first text (i.e., weather or soil) in either print or digital form, students 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 will not be able to access the text while you answer the questions.

As soon as they began reading, a research assistant started a timer to begin logging processing time. When students indicated they had completed the reading, the research assistant stopped the timer and presented them with the comprehension test in the same medium. The students were allowed to record their responses using bullet points or in connected discourse and were given unlimited time to complete the test. Once the comprehension test was completed, students judged their performance in that medium.

Equipment

Participants completed the digital portion of the study using a 15" LCD monitor at a resolution of 1280×1024 pixels. We assumed that these computers would be familiar to participants as they are the computers that students currently and routinely used at the university. This assumption was confirmed by asking students about their familiarity with the equipment prior to initiating the study. The texts presented digitally were presented as PDF-files, read using Adobe Reader for Windows. The printed texts were read from the textbook. The designated passage was bookmarked, and only relevant sections were visible to the reader.

Data Analysis

We used a within-subject design with each student reading one text in print and one digitally on either the topic of weather or soil in each condition (orders counterbalanced). As our research questions mainly involved examining the within-subject differences by topic, by medium, or by information sources, we mainly used paired-sample *t*-tests when the independent variable had two levels (i.e., topic: weather vs. soil; medium: digital vs. print) and applied repeated measures analysis of variance when the independent variable involved more than two levels (i.e., source of information: visual-only, text-only, and text-visual). Before we conducted the analyses for the main research questions, we first did initial analyses of the descriptive statistics and the correlations among variables. We particularly inspected the relations between the prereading topic measures (self-reported and demonstrated topic knowledge), reading time, and outcome variables on comprehension and comprehension calibration to identify potential covariates (self-reported and demonstrated topic knowledge or reading time) that should be adjusted for in analyses addressing the core research questions. In cases where significant correlations were found between a potential covariate and the outcome, in addition to the paired-sample *t*-test, we also conducted analyses controlling for potential influences of the covariates using linear mixed-effects models (on the condition that model assumptions were met). Mixed-effects models allow examination of the fixed effects of the within-subject condition (e.g., medium) and continuous predictors (e.g., topic knowledge) of interest, while taking between-subject variabilities into account (random effects). This approach can account for the dependence in outcomes (e.g., reading comprehension) and covariates (e.g., topic knowledge) across the within-subject conditions (mediums or topics) in our study.

Based on *a priori* power analyses using GPower, a sample size of 54 is adequate for a paired-sample *t*-test (e.g., differences by medium) with a moderate effect size (Cohen's $d = .50$) at a .05 significance level and for repeated measures ANOVA for one group with three measurements (as we would use for analysis by source of information) with a moderate effect size (Cohen's $f = .25$). However, it should be noted that there is a possibility that we may not be able to detect significance if the observed effect size is small.

Results and Discussion

Descriptive Statistics and Preliminary Analyses

Before addressing the research questions pertaining to the interplay of multimodality and medium, we examined the correlations and descriptive statistics, and we explored the results preliminarily by topic to ensure that data met the assumptions for subsequent analyses.

Reading Times by Medium and by Topic

Upon inspection, the one variable that required scrutiny for several reasons was reading time. Specifically, the reading time data for four students could not be retained due to procedural problems that occurred. There were also three students whose processing times in either or both mediums deviated significantly from their peers (2.7 to 3.8 *SDs* above the means of the respective conditions), unduly affecting the average for each medium. Therefore, it was decided that the time data for these seven students should be removed, resulting in a sample of 47 students for the analysis of reading times by medium.

With outliers removed, result of a paired-sample *t*-test indicated a significant difference in reading time as measured in seconds across mediums [$t(46) = 2.84$, $p = .007$], with students spending more time in the digital medium [print, $M = 420.81$, $SD = 141.56$; digital, $M = 471.70$, $SD = 151.14$]. This finding was unexpected, as it is the reverse of what previous medium-only

Table 1. Bivariate Pearson Correlation Matrix.

Measure	2	3	4	5	6	7	8	9	10	11	12
1. Self-reported TK[a]	.36**	−.33**	.15	.14	.14	.21*	−.05	.20*	.29**	−.15	.05
2. Demonstrated TK	–	.74**	.19	.20*	.17	.24*	.02	.13	.16	−.10	−.06
3. TK calibration		–	.07	.08	.07	.09	.04	−.03	−.08	.03	−.03
4. Total comprehension			–	.44**	.99**	.70**	.11	.88**	.22*	.42**	.04
5. Main idea item				–	.31**	.27**	−.03	.36**	.24*	.10	.06
6. Total detail items					–	.70**	.13	.88**	.19*	.42**	.04
7. Visual-only items						–	.08	.49**	.17	.30**	.14
8. Text-only items							–	.14	−.07	.13	.06
9. Text-visual items								–	.19*	.42**	−.02
10. Judgment of performance									–	−.72**	.22*
11. Comp. calibration										–	−.15
12. Reading time											–

Note:
[a]Topic knowledge.
$^{*}p < .05$, $^{**}p < .01$.

studies have found about processing time, where students spent more time reading in print than digitally (Singer & Alexander, 2017a, 2017b). As reading time was not correlated with any of the comprehension measures or calibration for overall comprehension (rs $< .15$, ps $> .05$), we determined that reading time would not be included in further analyses when addressing the core research questions. Additionally, we did not find any significant difference in reading time between text topics, $t(46) = 1.12$, $p = .268$. As there were no missing data for the other variables and reading time was not entered as a covariate, the full data set ($n = 54$) was used for subsequent analyses.

Topic Differences for Comprehension Outcomes

We also conducted an initial analysis of topic differences in students' comprehension performance by total score, by specificity level (main idea or details), and by information source (text, visual, or both) without regard to medium. We found no significant differences between the text topics for total comprehension, or for the main idea or detail questions. Nor did we find any significant topic effect for questions that tested comprehension of information referenced in text only, visuals only, or text-visual. Further, there was no significant topic difference for comprehension calibration. Therefore, we collapsed the two topics when addressing the core research questions regarding comprehension and comprehension calibration in ensuring analyses.

Correlations between Self-Reported and Demonstrated Topic Knowledge and Comprehension

Finally, we explored the interrelations among the pre-reading variables of self-reported topic knowledge, demonstrated topic knowledge, and post-reading comprehension, and calibration scores. As shown in Table 1, self-reported topic knowledge and demonstrated topic knowledge had a moderate positive correlation, $r = .36$, $p < .001$. However, neither of these topic knowledge measures were significantly correlated with total comprehension score or calibration accuracy. Students who self-reported higher topic knowledge tended to have higher ratings of their comprehension performance ($r = .29$, $p = .002$), but topic knowledge calibration accuracy was not correlated with the comprehension scores or comprehension calibration accuracy. As a result of these findings, the overall comprehension and comprehension calibration scores required no adjustments for either self-reported or demonstrated topic knowledge for subsequent analyses.

For comprehension at different levels of specificity, neither self-reported nor demonstrated topic knowledge was correlated with comprehension of text details. However, main idea comprehension was significantly correlated with demonstrated topic knowledge ($r = .20$, $p = .035$), but

Table 2. Descriptive Statistics for Self-Reported and Demonstrated Topic Knowledge.

Measure	Topic						
	Weather				Soil		
	Max.	*M*	*SD*	Range	*M*	*SD*	Range
Self-reported topic knowledge	100	20.15	18.14	0–80	9.39	9.58	0–35
Demonstrated topic knowledge	7	3.06	1.28	1–6	3.39	0.98	2–6
Topic knowledge calibration	100	32.01	17.98	2.67–80	47.09	15.69	20–80
Raw discrepancy	100	−30.78	20.06	−80–20	−47.09	15.69	−80–(−20)

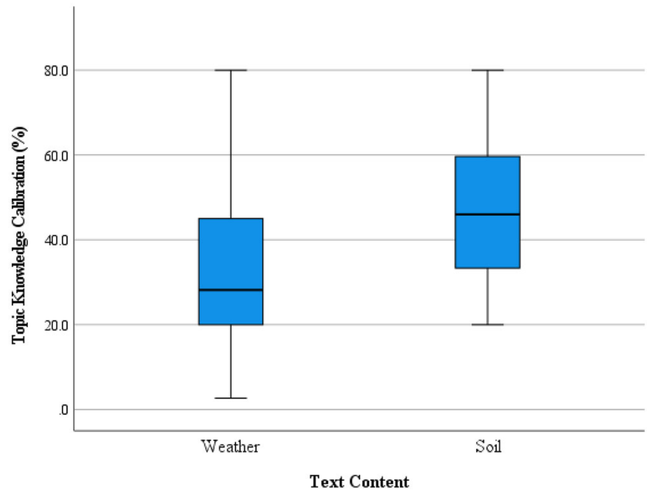

Figure 1. Topic Knowledge Calibration Differences Between Weather and Soil Topics.

not with self-reported topic knowledge ($r = .14$, $p = .139$). To rule out the potential influence of topic knowledge on students' ability to grasp main ideas when reading in different mediums, we controlled for demonstrated topic knowledge in the subsequent analysis of main idea questions by medium.

Correlations also differed by source of information for comprehension questions. Comprehension performance on visual-only questions was significantly positively correlated with self-reported topic knowledge ($r = .21$, $p = .026$) and demonstrated topic knowledge ($r = .24$, $p = .013$). Text-visual question performance was also positively correlated with self-reported topic knowledge ($r = .20$, $p = .035$). But no significant correlations were found between text-only items and self-reported or demonstrated topic knowledge. Topic knowledge calibration was not correlated with comprehension by any type of information source. As a result of these findings, we adjusted for self-reported or demonstrated topic knowledge for visual-only questions and self-reported topic knowledge for text-visual items in subsequent analysis.

Research Question 1: Differences between Self-Reported and Demonstrated Topic Knowledge

We first analyzed the accuracy of the undergraduates' self-reported topic knowledge compared to their demonstrated topic knowledge (see Table 2). As displayed in Figure 1, the scores for students' self-reported knowledge with the two geology topics, although low, were significantly higher for *weather* than for *soil*, as predicted, $t(53) = 5.75$, $p < .001$, Cohen's $d = .78$. This was consistent with the findings of the pilot study and confirmed that students perceived one of the two topics as significantly better known than the other. However, the results for the topic knowledge pretest showed no statistically significant differences in students' demonstrated knowledge for these soil topics, $t(53) = 1.77$, $p = .083$. We then examined the accuracy of students' topic knowledge calibration. We found that while students tended to underestimate their knowledge for both topics, they were more likely to do so when they judged the topic as less well-known. The topic knowledge calibration score, which was the absolute value difference between students' self-reported and demonstrated topic knowledge, was higher for the soil topic than for the weather topic [$t(53) = 5.02$, $p < .001$, Cohen's $d = .68$], indicating lower calibration accuracy for the topic of soil.

Research Question 2: Comprehension Performance by Source of Information

Next, we examined students' comprehension of information conveyed in the text only, the visuals only, or in both. As the maximum scores for text-only, visual-only, and text-visual questions varied, we converted these scores into percentages for the purpose of analysis. In that way, we could compare comprehension performance for information from different sources on the same metric. A repeated measures ANOVA was conducted to compare the performance on the three information sources. The assumption of sphericity was violated, as shown by Mauchly's test, $\chi^2(2) = 19.56$, $p < .001$, so a Huynh-Feldt adjustment was used.

The result showed significantly different comprehension outcomes for the three sources, $F(1.74, 185.83) = 6.31$, $p = .004$, partial $\eta^2 = .06$. Post hoc pair-wise comparisons with Bonferroni correction of level of significance for multiple comparisons (corrected $\alpha = .017$) showed that performance on text-only questions ($M = .82$, $SD = .24$) was significantly higher than performance on text-visual questions ($M = .71$, $SD = .20$), $t(107) = 3.82$, $p < .001$, Cohen's $d = .37$. No statistically significant differences were found between performance on visual-only ($M = .77$, $SD = .30$) and text-only questions [$t(107) = -1.36$, $p = .177$] or between visual-only and text-visual questions [$t(107) = 2.26$, $p = .026$]. It should be noted that our analysis could be underpowered for detecting a significant difference with a small effect size, as we observed in the case of visual-only and text-visual questions (Cohen's $d = .22$). Nonetheless, these results deviated from our expectations that higher comprehension scores would result for the text-visual questions and lower scores would likely occur for the visual-only questions.

Research Question 3: Comprehension Performance by Medium

The third research question guiding this investigation focused on the effects of medium on the undergraduate students' comprehension performance for multimodal texts on the topics of weather and soil. We first analyzed students' overall comprehension performance by medium (see Table 3). When total comprehension score was the outcome measure, students performed significantly better when they read in print ($M = 13.93$, $SD = 2.54$) than digitally ($M = 12.07$, $SD = 3.03$), $t(53) = 3.49$, $p < .001$, Cohen's $d = .48$. This result parallels the outcomes from prior investigations (Delgado & Salmerón, 2021; Mangen et al., 2013; Singer Trakhman et al., 2019). Next, we examined medium differences for performance on comprehending the main ideas as opposed to more specific content in the multimodal texts. As in our prior studies (Singer & Alexander, 2017a, 2017b), medium had no effect on students' ability to identify the main ideas of each text, $t(53) = .26$, $p = .799$. This result held consistent even when we controlled for demonstrated topic knowledge, which was significantly correlated with main idea comprehension, by using a mixed-effects model with medium as the fixed-effect factor, demonstrated topic knowledge as a covariate, and participant ID as a random effect, $F(1, 52) = .04$, $p = .843$. In contrast, medium proved influential when students answered the four short-answer questions addressing more detailed content. For those detail questions their performance was significantly better in the print ($M = 12.15$, $SD = 2.33$) than in the digital condition ($M = 10.31$, $SD = 2.90$), $t(53) = 3.63$, $p < .001$, Cohen's $d = .49$. This superiority of comprehension when reading in the print compared

Table 3. Descriptive Statistics for Comprehension and Calibration by Medium.

Measure	Max.	Medium					
		Print			Digital		
		M	*SD*	Range	*M*	*SD*	Range
Total comprehension	17	13.93	2.54	8–17	12.07	3.03	5–17
Main idea	2	1.78	0.42	1–2	1.76	0.43	1–2
Details	15	12.15	2.33	7–15	10.31	2.90	4–15
Text-only	4	3.44	0.88	1–4	3.07	1.03	0–4
Visual-only	4	3.43	1.06	0–4	2.70	1.25	0–4
Text + Visual	7	5.37	1.14	3–7	4.54	1.51	1–7
Comprehension calibration	100	31.11	20.99	1.18–80.00	21.57	18.96	0.71–74.12
Judgment of performance	100	52.56	21.96	6–100	55.96	22.42	2–100

to the digital medium for understanding of particulars in the reading again echoes findings in previous studies (Singer Trakhman et al., 2018).

Finally, we investigated whether students' performance for items based on information in the text only, the visuals only, or referenced in both would differ by medium. We ran separate paired-sample t-tests for the text-only, visual-only, and text-visual questions. The results showed that students performed better in the print than the digital condition when responding to questions that relied on information conveyed in visuals only [$t(53) = 3.70$, $p < .001$, Cohen's $d = .50$] or in both the text and the visuals [$t(53) = 3.13$, $p = .003$, Cohen's $d = .43$]. However, when information was found only in the text, the comprehension performance between mediums was not significantly different, $t(53) = 1.87$, $p = .067$. It should be noted that the mean score for text-only questions was higher in the print ($M = 3.44$, $SD = .89$) than in the digital condition ($M = 3.07$, $SD = .1.03$; see Table 3), but our relatively small sample size may have limited the statistical power to detect a significant difference with a smaller effect size (Cohen's $d = .25$). Nevertheless, the results of these analyses showed that the advantage in comprehension for reading in print (versus digitally) was more pronounced when questions involved content conveyed in the visuals.

As noted, we found significant correlations between topic measures (self-reported and demonstrated topic knowledge) and visual-only items. To rule out the potential influence of prior knowledge in this instance, we created two separate mixed-effects models to control for demonstrated and self-reported topic knowledge respectively. In both models, medium was entered as the fixed-effect factor, visual-only question response as the dependent variable, and participant ID entered as a random-effect factor. Demonstrated topic knowledge was entered as a covariate in the first model and self-reported topic knowledge in the second. The results showed that when demonstrated topic knowledge was controlled, the medium difference remained significant, $F(1, 52) = 13.37$, $p = .001$, partial $\eta^2 = .21$, with better performance in the print condition on visual-only items. This advantage of reading in print remained significant when self-reported topic knowledge was controlled, $F(1, 52) = 9.43$, $p = .003$, partial $\eta^2 = .15$. Finally, we examined medium effect on text-visual questions with adjustment of self-reported topic knowledge. We similarly tested a mixed-effects model, with medium as the fixed-effect variable, text-visual question response as the dependent variable, participant ID as a random-effect factor, and self-reported topic knowledge as a covariate. The superiority of print over digital condition in comprehending information delivered through both text and visual modalities remained significant, $F(1, 52) = 6.21$, $p = .016$, partial $\eta^2 = .11$. All model assumptions were satisfied.

Research Question 4: Medium Differences for Calibration for Overall Comprehension

The final research question addressed the effect of medium on students' calibration of their reading performance. To investigate this question, we ran a paired-sample t-test to compare comprehension calibration accuracy between mediums. For this analysis, lower scores were indicative of less discrepancy between predicted and actual performance and thus better calibration. Thus, when it came to overall comprehension, results revealed that students were significantly better calibrated when they read digitally ($M = 21.57$, $SD = 18.96$) than when they read in print ($M = 31.11$, $SD = 20.99$), $t(53) = 2.97$, $p = .004$, Cohen's $d = .41$. This outcome did not conform to our hypothesis that students would be better calibrated when reading multimodal texts in print than digitally.

In interpreting this result, we turned to students' judgment of performance (JOP) and their actual comprehension performance (RCT) scores, both denoted in percentages (see Table 2). By comparing their JOP and RCT scores, we found that while students generally tended to underestimate their performance across mediums, their prediction of their comprehension performance was slightly lower when they read the texts in print ($M_{JOP} = 52.56$, $SD = 21.96$) than digitally ($M_{JOP} = 55.96$, $SD = 22.42$), although this difference was not statistically significant [$t(53) = 1.32$, $p = .194$]. In contrast, their actual comprehension performance for the print condition ($M_{\mathrm{RCT}} = 81.92$, $SD = 14.94$) was significantly higher than for the digital condition ($M_{\mathrm{RCT}} = 71.02$, $SD = 17.84$), $t(53) = 3.49$, $p < .001$.

As a result, the discrepancy between predicted (JOP) and observed (RCT) values of comprehension performance was higher and thus the calibration accuracy was poorer in the print condition, as compared to the digital condition, which was not expected. Based on ratings, it appears that students were generally more conservative judges of their performance when reading multimodal texts in print than digitally than has been the case. This pattern relates to yet another unexpected outcome; the better comprehenders in this study were less well calibrated. Comprehension calibration accuracy was significantly positively correlated with overall comprehension ($r = .42$, $p < .001$) and comprehension of text details ($r = .42$, $p < .001$). When examined by source of information, comprehension calibration was associated with responses to visual-only items ($r = .30$, $p = .002$) and text-visual items ($r = .42$, $p < .001$), but not with text-only items ($r = .13$, $p = .188$). [It should be remembered that the higher the calibration score the less accurate the accuracy.]

To further analyze the role of medium in students' calibration, we categorized their self-ratings of comprehension performance into three categories: underestimation (i.e., self-judgment more than 3 percentage points below actual performance), accurate (i.e., self-judgment within 3 percentage points of actual performance), and overestimation (i.e., self-judgment more than 3 percentage points above actual performance). This categorization revealed that overall students were poorly calibrated, regardless of medium. Specifically, only 7.4% ($n = 4$) of the students were classified as accurately calibrating their performance in the print condition and 9.3% ($n = 5$) in the digital condition. There were appreciably more students who underestimated their performance when reading in print (83.3%, $n = 45$) than when reading digitally (64.8%, $n = 35$). Conversely, more students overestimated their performance in the digital (25.9%, $n = 14$) than in the print (9.3%, $n = 5$) condition.

Conclusions and Implications

For this investigation, our goals were to broaden our program of research into the effects of medium on comprehension to explore whether processing multimodal texts changed outcomes that have consistently shown print to be preferable to digital texts. We also wanted to expand our exploration of students' calibration accuracy, which has been tied to self-regulatory behaviors and performance outcomes (Bol et al., 2005). Specifically, we gauged the accuracy of undergraduates' judgments of topic knowledge about weather and soil against the level of topic knowledge they demonstrated on the pre-reading assessment. In addition, we assessed the accuracy of students' judgments about how well they performed on the post-reading comprehension measures for the multimodal weather and soil texts read in print and digitally against their actual performance on those measures. Before we consider the implications that arose from this investigation, we want to acknowledge certain delimitations and limitations.

Delimitations and Limitations

The most evident delimitation of the current study rests on our decision to use naturally-occurring excerpts from an introductory geology textbook without any manipulations. There are evident costs that derive from this decision, most notably the lack of strict controls with regard to the precise characteristics of the visuals and their positioning relative to any associated textual content that typifies the multimodal literature. However, it was our contention that the benefits of exploring college students' processing of texts that are more realistic than the very short passages with carefully drawn and positioned visuals common in the multimodal literature was worth the costs. What we also acknowledge as a cost of this more naturalistic approach is our inability to monitor students' movements within the excerpts within the text and visuals and between these multimodal components. Finally, we recognize that our decision to continue to investigate comprehension of multimodal and multiple medium texts among college students as delimitation. Yet, we wanted to explore the nature of comprehension and calibration with these demanding multimodal texts read on paper and on screen for mature, capable readers before turning to younger, less experienced, and potentially less skilled text processers.

One limitation of this investigation that we had not anticipated was the small sample size that restricted our ability to appropriately judge comprehension of information delivered through text only across the two mediums and difference between performance on visual-only and text-visual questions. Although the number of participants was not large, we had determined via power analysis that these numbers would be sufficient for the within-subjects analyses we had planned. However, our power analysis was based on expected effects of moderate sizes (Cohen's $d = .50$). When the observed effect size was small, as in the cases of text-only items across mediums (Cohen's $d = .25$) and between visual-only and text-visual items (Cohen's $d = .22$), we may not have been able to detect statistical significance due to low power. Thus, the findings for these analyses on information sources must be interpreted with caution.

Key Outcomes

What emerged as perhaps the most salient outcome of this multimodal, multiple mediums investigation had to do with the continued primacy of reading in print versus reading digitally with regard to what students understood and recalled from naturally-occurring expository texts. It was not just the fact that processing multimodal texts in print was statistically more advantageous than processing those same texts digitally. Rather, what we determined was that the contributors of the print over digital advantage in this study were not entirely as expected either in terms of students' comprehension performance or their calibration accuracy. For the comprehension questions that differed by information source, for example, it was not the text-only items that produced the effect for medium as we anticipated. The advantage for reading in print over digitally emerged with content found only in the visuals or in the visuals and the text. Whether a medium effect for text-only questions would have emerged with a larger sample size remains unresolved.

There were also unexpected results when it came to the two measures of calibration in this study—topic knowledge and comprehension calibration. First, in both instances, students were far more likely to underestimate either what they knew about the topics or how well they answered the comprehension questions than we had observed in prior investigations. Second, although we continue to find that students are not very effective judges of their existing knowledge nor their comprehension performance, they turned out to be more accurate in their judgments about comprehension performance when they read digitally than in print. This finding is antithetical to what we have reported in prior investigations (Singer & Alexander, 2017a; Singer Trakhman et al., 2019). Third, while calibration accuracy for topic knowledge was unrelated to comprehension performance, the same was not true for calibration accuracy for comprehension performance. Yet, what was counterintuitive about the significant correlation for comprehension calibration and comprehension performance was that the better comprehenders were more poorly calibrated—not what would be expected from the literature on calibration (Dunlosky & Thiede, 2013; Gutierrez & Schraw, 2015; Nietfeld & Schraw, 2002).

In reflecting on these unexpected outcomes, we can forward several plausible explanations for their occurrence pertaining to underestimations of prior knowledge, awareness of text-processing demands, text-visual contiguity issues, increased digital reading times, scrolling requirements, and conservative self-judgments. From the outset of this study, when these undergraduates were asked to rate their knowledge of the study topics, it was evident that they felt they knew very little about weather and even less about soil. In actuality, they performed moderately well on the topic knowledge pretests. However, it is likely that their perceptions of low knowledge may have lowered these undergraduates' efficacy judgments about this reading task, which manifested in their judgments of performance underestimations. These initial concerns about the task were likely exacerbated when students began to read the excerpts, which were moderately long, somewhat technical, punctuated with terminology, and consisting of informative visuals.

What also helps to understand the current pattern of results that deviated from prior investigation was the way in which the text and related visuals were displayed particularly in the digital condition. When students read the printed document, they were able to see the entire page of these large size textbooks. Even when visuals were not placed in close proximity to the related text (Mayer, 2002; Mayer & Fiorella, 2014), they typically appeared somewhere on the page or the facing page. The students in the digital condition, by contrast, could view only a portion of any page at a time. Therefore, this meant students often could not see a visual and related text on the same screen and were required to scroll back and forth between the two. Researchers have reported the detrimental effects of scrolling on comprehension in the digital medium in general (Mangen et al., 2013; Støle et al., 2020). Yet, in this instance, with the importance of text-visual integration heightened, the need for scrolling would likely be of more concern to students, leading to increased processing demands when reading digitally as well as increased interference in comprehension when visual content was the focus.

Support for the contentions of greater processing demands in the digital condition and concomitant decrement in comprehension due to the lack of text-visual contiguity can be found in task performance data and students' observed behaviors. For instance, the surprising fact that reading times were longer in the digital condition than in the print condition in this investigation—which has never occurred in our prior studies—seems evidence of increased task demands. There were also student behaviors that we observed during task performance that speak to this outcome. For example, we witnessed students' apparent frustration when required to scroll back and forth in the digital condition between text and visuals. Not only did some students comment on this situation, but a few also requested permission to zoom out on the digital display so that they could view much more of each page. Despite the longer time students took when processing digital texts in this study, their more conservative judgments of performance, and their apparent efforts to integrate textual and visual content, these mature readers still performed better in the print condition overall and for questions that referred to visual content. We consider the implications of this key finding for future research and for instructional practices.

The overall text processing demands in this investigation may also have played a role in what we referred to as the counterintuitive finding for calibration accuracy. The fact that the better comprehenders exhibited worse calibration accuracy was initially quite perplexing. Nonetheless, we came to reconcile this outcome when we delved into the literature on variability in self-judgment and self-assessment accuracy (León et al., 2021; Tirso et al., 2019; Urban & Urban, 2021).

What this literature rather consistently demonstrates is that higher ability students exhibit the tendency to underestimate their capabilities or performance relative to given task, whereas lower ability students are more prone to overestimation when asked to self-judge or self-assess capabilities or performance. Whether in this study the miscalibration arose from the students' heightened awareness of task complexity (Tirso et al., 2019), concern over their existing topic knowledge (McCarthy & McNamara, 2021), a diminished self-confidence given task unfamiliarity or lack of clear performance criteria (Hattie, 2013), or even the specific measurement scale we used (Hartwig & Dunlosky, 2014) remains an open question.

Implications for Future Research and Instructional Practice

There are several implications that have emerged from this investigation that pertain to subsequent empirical studies needed to address unresolved or newly exposed issues, as well as to situations where multimodal texts are part of instruction in tertiary or secondary education classes. For one, in the current study we examined the interplay of multimodal texts with medium under what we regard as more naturalistic conditions: rather long excerpts from a college textbook containing a variety of visuals. It was also the case that the alignment of visuals with relevant written content in these naturally-occurring texts did not seem to follow any of the multimedia principles that Mayer (2002) contends improves comprehension and diminishes cognitive load for readers. We certainly witness the effects of non-contiguity between text and visual firsthand in the digital condition where increased scrolling and frustration could be observed.

Given our experiences in this study, we would suggest that in the future, studies on print and digital mediums consider testing the effects of text-visual placement but within otherwise unmodified, longer text excerpts similar to what high-school or college students use in their coursework. We would also call on those engaged in multimodal research to expressly consider the medium of text delivery on student comprehension and learning outcomes. Those who have looked intensely at medium have rather consistently demonstrated the added challenges that come with reading on screen rather than on paper. What this means for the generalizability of multimedia principles remains to be seen. Beyond what we observed when visual and related text were not both visible on the screen, what are the effects on comprehension and learning when multimedia principles are instantiated on paper versus on screen? Does the primacy of reading on paper still hold?

In addition to future investigations in which we experimentally test the placement of visuals in otherwise naturally occurring texts, valuable data could be acquired through the use of eye-tracking procedures such as those employed by Mason et al. (2013a) to corroborate the integration of visual and written content in more ecologically valid expository texts. Mixed-method studies that combine think-aloud or interview data with comprehension, time processing, and calibration information would afford an even broader and deeper portrait of students' navigation of multimodal texts in print and digital mediums. What might account for the underestimation of background knowledge by college students for topics like weather and soil? How do better comprehenders explain their serious underestimation of the comprehension performance?

Finally, what stands out to us from this investigation was the absence of any clear rationale for the number, form, and placement of visuals in the commercially-produced materials that students are required to read in college courses. This raises the question as to how aware authors and publishers are of guidelines or principles pertaining to optimizing student learning. Moreover, at many institutions, including our own, course instructors are urged to adopt textbooks that are produced in digital forms. However, the added challenges of processing these course materials on screen seem to garner little if any attention. Thus, we urge the development of guidelines for the effective use of multimodal texts read digitally for students and instructors alike. While some of that critical work is currently underway, much more such inquiry is needed (Firetto & Van Meter, 2018; Van Meter et al., 2017). There is unquestionably much more to be learned about the interplay of multimodality and medium in students' learning from expository texts from different academic domains. Nonetheless, it is essential that this work moves forward, if the ultimate aim is to improve students' learning from text and their ensuing academic development.

ORCID

Lauren M. Singer Trakhman http://orcid.org/0000-0003-4763-3270
Patricia A. Alexander http://orcid.org/0000-0001-7060-2582
Yuting Sun http://orcid.org/0000-0001-8926-347X

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

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. https://doi.org/10.1016/j.learninstruc.2012.10.003

Alexander, P. A., & Singer, 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. https://doi.org/10.1080/00405841.2011.558436

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. https://doi.org/10.1037/a0014312

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education*, *73*(4), 269–290. https://doi.org/10.3200/JEXE.73.4.269-290

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2), 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Coiro, J. (2021). Toward a multifaceted heuristic of digital reading to inform assessment, research, practice, and policy. *Reading Research Quarterly*, *56*(1), 9–31. https://doi.org/10.1002/rrq.302

Cromley, J. G., Weisberg, S. M., Dai, T., Newcombe, N. S., Schunn, C. D., Massey, C., & Merlino, F. J. (2016). Improving middle school science learning using diagrammatic reasoning. *Science Education*, *100*(6), 1184–1213. https://doi.org/10.1002/sce.21241

Delgado, P., & Salmerón, L. (2021). The inattentive on-screen reading: Reading medium affects attention and reading comprehension under time pressure. *Learning and Instruction*, *71*, 101396. https://doi.org/10.1016/j.learninstruc.20

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, *24*, 58–61. https://doi.org/10.1016/j.learninstruc.2012.05.002

Firetto, C. M., & Van Meter, P. N. (2018). Inspiring integration in college students reading multiple biology texts. *Learning and Individual Differences*, *65*, 123–134. https://doi.org/10.1016/j.lindif.2018.05.011

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. https://doi.org/10.1037/0096-1523.3.4.552

Fry, E. (1968). A readability formula that saves time. *Journal of Reading*, *11*(7), 513–578. https://www.jstor.org/stable/40013635

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 702–718. https://doi.org/10.1037/0278-7393.11.1-4.702

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory & Language*, *26*(1), 69–83. https://doi.org/10.1016/0749-596X(87)90063-5

Guo, D., Zhang, S., Wright, K. L., & McTigue, E. M. (2020). Do you get the picture? A meta-analysis of the effect of graphics on reading comprehension. *AERA Open*, *6*(1), 233285842090169–233285842090120. https://doi.org/10.1177/2332858420901696

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *The Journal of Experimental Education*, *83*(3), 386–404. https://doi.org/10.1080/00220973.2014.907230

Hartwig, M. K., & Dunlosky, J. (2014). The contribution of judgment scale to the unskilled-and-unaware phenomenon: How evaluating others can exaggerate over-(and under-) confidence. *Memory & Cognition*, *42*(1), 164–173. https://doi.org/10.3758/s13421-013-0351-4

Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction*, *24*, 62–66. https://doi.org/10.1016/j.learninstruc.2012.05.009
Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, *32*(1), 241–267. https://doi.org/10.3102/0091732X07310586
Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69.
Kress, G. R., & Van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. Psychology Press.
Lauer, C. (2009). Contending with terms: "Multimodal" and "multimedia" in the academic and public spheres. *Computers and Composition*, *26*(4), 225–239. https://doi.org/10.1016/j.compcom.2009.09.001
León, S. P., Pantoja Vallejo, A., & Nelson, J. B. (2021). Variability in the accuracy of self-assessments among low, moderate, and high performing students in university education. *Practical Assessment, Research, and Evaluation*, *26*(1), 16. https://doi.org/10.7275/6q91-az58
Lin, L. M., Zabrucky, K. M., & Moore, D. (2002). Effects of text difficulty and adults' age on relative calibration of comprehension. *The American Journal of Psychology*, *115*(2), 187–198. http://www.press.uillinois.edu/about/permission.html https://doi.org/10.2307/1423434
List, A., van Meter, P., Lombardi, D., & Kendeou, P. (2020). Loggers and conservationists: Navigating the multiple resource forest through the trees. In P. van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives*. Routledge. https://doi.org/10.4324/9780429443961-1
Liu, J. (2013). Visual Images interpretive strategies in multimodal texts. *Journal of Language Teaching and Research*, *4*(6), 1259–1263. https://doi.org/10.4304/jltr.4.6.1259-1263
Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, *58*, 61–68. https://doi.org/10.1016/j.ijer.2012.12.002
Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013a). An eye-tracking study of learning from science text with concrete and abstract illustrations. *The Journal of Experimental Education*, *81*(3), 356–384. https://doi.org/10.1080/00220973.2012.727885
Mason, L., Tornatora, M. C., & Pluchino, P. (2013b). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*(1), 95–109. https://doi.org/10.1016/j.compedu.2012.07.011
Mayer, R. E. (2002). Multimedia learning. In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 41, pp. 85–139). Academic Press.
Mayer, R. E., & Fiorella, L. (2014). 12 principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). Cambridge University Press.
McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist*, *56*(3), 196–119. https://doi.org/10.1080/00461520.2021.1872379
McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: Across the ages and genres. In J. P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how to assess reading ability* (pp. 89–116). Rowman & Littlefield.
Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*(2), 358–368. https://doi.org/10.1037/0022-0663.91.2.358
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. https://doi.org/10.1080/00220670209596583
O'Neil, K. E. (2011). Reading pictures: Developing visual literacy for greater comprehension. *The Reading Teacher*, *65*(3), 214–223. https://doi.org/10.1002/TRTR.01026
Pipkin, B., Trent, D. D., Hazlett, R., & Bierman, P. (2014). *Geology and the environment* (7th ed.). Brooks/Cole.
Rose, D., & Dalton, B. (2009). Learning to read in the digital age. *Mind, Brain, and Education*, *3*(2), 74–83. https://doi.org/10.1111/j.1751-228X.2009.01057.x
Scheiter, K., Schubert, C., Schüler, A., Schmidt, H., Zimmermann, G., Wassermann, B., Krebs, M.-C., & Eder, T. (2019). Adaptive multimedia: Using gaze-contingent instructional guidance to provide personalized processing support. *Computers & Education*, *139*, 31–47. https://doi.org/10.1016/j.compedu.2019.05.005
Scheiter, K., Schüler, A., Gerjets, P., Huk, T., & Hesse, F. W. (2014). Extending multimedia research: How do prerequisite knowledge and reading comprehension affect learning from text and pictures. *Computers in Human Behavior*, *31*, 73–84. https://doi.org/10.1016/j.chb.2013.09.022
Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behavior*, *27*(1), 89–93. https://doi.org/10.1016/j.chb.2010.05.028
Serafini, F. (2011). Expanding perspectives for comprehending visual images in multimodal texts. *Journal of Adolescent & Adult Literacy*, *54*(5), 342–350. https://doi.org/10.1598/JAAL.54.5.4

Serafini, F., & Ladd, S. M. (2008). The challenge of moving beyond the literal in literature discussion. *Journal of Language and Literacy Education*, *4*(2), 6–20.

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. https://doi.org/10.1080/00220973.2016.1143794

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. https://doi.org/10.3102/0034654317722961

Singer Trakhman, L. M., Alexander, P. A., & Berkowitz, L. (2019). Effects of processing time on comprehension and calibration in print and digital mediums. *Journal of Experimental Education*, *27*, 101–115. https://doi.org/10.1080/00220973.2017.1411877

Singer Trakhman, L. M., Alexander, P. A., & Silverman, A. B. (2018). Profiling reading in print and digital mediums. *Learning and Instruction*, *57*, 5–17. https://doi.org/10.1016/j.learninstruc.2018.04.001

Støle, H., Mangen, A., & Schwippert, K. (2020). Assessing children's reading comprehension on paper and screen: A mode-effect study. *Computers & Education*, *151*, 103861. https://doi.org/10.1016/j.compedu.2020.103861

Tirso, R., Geraci, L., & Saenz, G. D. (2019). Examining underconfidence among high-performing students: A test of the False Consensus Hypothesis. *Journal of Applied Research in Memory and Cognition*, *8*(2), 154–165. https://doi.org/10.1016/j.jarmac.2019.04.003

Unsworth, L., & Cléirigh, C. (2014). Multimodality and reading: The construction of meaning through image-text interaction. In C. Jewitt (Ed.), *The handbook of multimodal analysis* (pp. 176–188). Routledge.

Urban, M., & Urban, K. (2021). Unskilled but aware of it? Cluster analysis of creative metacognition from preschool age to early adulthood. *The Journal of Creative Behavior*, *55*(4), 937–945. https://doi.org/10.1002/jocb.499

Van Leeuwen, T. (2015). Multimodality. In D. Tannen, H. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd Ed., pp. 447–465). John Wiley & Sons.

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. https://doi.org/10.1016/j.learninstruc.2017.01.003

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9

Appendix: Pretests and posttests

Weather pretest

1. Define precipitation
2. Define storm surges
3. Define hydrologists
4. Why are heavy thunderstorms a summer phenomenon?
   a. The land needs water to cool off
   b. They are a result of an atmosphere made unstable by strong solar heating
   c. They are a result of an atmosphere made stable by strong solar heating
   d. They are not a summer phenomenon
5. Which of the following is not known to cause floods? Select all that apply.
6. Precipitation
7. Dams
8. Earthquakes
9. Climate warming
10. Why are mountains wetter than adjacent valleys?
11. Cool air ascends and condenses
12. Warm air descends causing cloud dissipation
13. Clouds form over the adjacent valleys
14. Adjacent valleys are wetter than mountains

Weather comprehension posttest

Answer the following questions about the weather and climate passage.

1. In 2–3 sentences, summarize the main points of the textbook section you just read. (2 pts)
2. Why are mountains always wetter than valleys? (4 pts.)

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 pts.)
4. Describe 2 ways in which factors beyond rainfall can contribute to flooding. (2 pts.)
5. List up to 4 additional facts or ideas that you remember from the passage. (4 pts.)

Soils pretest

1. Define soil horizons
2. Define loess
3. Define residual soils
4. What is the fragmental rock material at and just below the Earth's surface called?
5. Top layer
6. Regolith
7. Soil horizons
8. Bedrock
9. What influences the thickness and development of soil profiles?
10. Climate
11. Time
12. Climate & time
13. Climate & weather
14. What happens to residual soils? Select all that apply.
15. They never disappear once formed
16. They are always subject to erosion and removal by geological processes
17. They only sometimes can be removed by geological processes
18. They are never developed

Soil comprehension posttest

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

1. In 2–3 sentences, summarize the main points of the textbook section you just read. (2 pts.)
2. Name and describe 4 of the 5 layers in an idealized soil profile. (4 pts.)
3. According to the information graphically presented in the passage, where would you find 4 areas of the earth with extensive loess covering? (4 pts.)
4. What is a loess? List up to 2 of its physical properties. (3 pts.)
5. List up to 4 additional facts or ideas that you remember from the passage. (4 pts.)