American Educational Research Journal
August 2020, Vol. 57, No. 4, pp. 1837–1867
DOI: 10.3102/0002831219890300
Article reuse guidelines: sagepub.com/journals-permissions

Digital Versus Paper Reading Processes and Links to Comprehension for Middle School Students

Amanda P. Goodwin
Sun-Joo Cho
Vanderbilt University
Dan Reynolds
John Carroll University
Katherine Brady
Jorge Salas
Vanderbilt University

This study explores digital and paper reading processes and outcomes for 371 fifth to eighth graders completing a reading task similar to standardized testing. Results showed students highlighted and annotated more when reading the paper versus digital text. Also, reading on paper versus digitally was slightly supportive of reading comprehension for the longer section of text. For behaviors, digital highlighting and looking back at the paper text were

---

Amanda P. Goodwin, PhD, is an associate professor in the Department of Teaching and Learning at Vanderbilt University's Peabody College, 230 Appleton Place, Nashville, TN 37203, USA; e-mail: *amanda.goodwin@vanderbilt.edu*. Her research focuses on language and literacy experiences for upper elementary and middle school students.

Sun-Joo Cho, PhD, is an associate professor of quantitative methods in the Department of Psychology and Human Development at Vanderbilt University's Peabody College. Her primary research interests include generalized latent variable modeling and parameter estimation, with a focus on item response modeling.

Dan Reynolds, PhD, is an assistant professor of literacy education at John Carroll University. His research focuses on how designed instruction and interactional scaffolding can support adolescent readers reading of complex texts.

Katherine Brady, PhD, is a former graduate student in the Department of Computer Science and Electrical Engineering at Vanderbilt University. She studies tools for automatically interpreting fine grained reading behaviors through video recordings. Her background is in artificial intelligence with a focus on digital educational technologies.

Jorge Salas, PhD, is a research associate at Vanderbilt University where he earned his PhD is physics. He has experience in scientific computing, simulations, and data analysis.

supportive of reading comprehension, whereas paper highlighting was negatively related, likely because paper highlighting tended to occur often outside of important areas of the text. Paper and digital annotating, online dictionary use, and digital look-back did not link to comprehension, when controlling for other covariates. Links to theory, research, and practice are discussed.

Keywords: digital reading, reading comprehension, reading behaviors, highlighting

The rapid expansion of digital reading has exceeded the speed of research on its effects. To date, the digital transition has occurred without understanding of the differences between how digital and traditional texts are read. As Lankshear and Knobel (2013) note, "Traditional assumptions about what it means to read and write and how meanings are communicated have been disrupted and displaced" (p. 5). With 80% of United States eighth graders reporting using computers daily for schoolwork (KewalRamani et al., 2018), this move toward digital reading requires research on whether and how digital texts, and the reading behaviors used with such texts, affect reading comprehension (Biancarosa & Griffiths, 2012; Larson, 2009; Leu, Kinzer, Coiro, Castek, & Henry, 2013; Zipke, 2013). This research is particularly needed now because of emerging evidence about the potential negative effects of digital reading when compared with paper reading. In the past 2 years, one systematic review found benefits for paper reading over digital reading on long texts (Singer & Alexander, 2017) and three meta-analyses found significant and meaningfully sizeable effect sizes showing comprehension benefits to reading paper-based texts over static digital texts (Clinton, 2019; Delgado, Vargas, Ackerman, & Salmerón, 2018; Kong, Seo, & Zhai, 2018). These researchers all strike an alarming note: While digital reading is expanding both in the United States and internationally, it may be having deleterious effects on student readers. These reviews also emphasize the need for more research about *why* and *how* comprehension processes are different in digital versus paper reading because their work was limited to examining the outcomes of reading (i.e., comprehension results). Our work fills this gap as such questions have important policy, practice, theory, and research implications.

The literature on digital reading for middle school students shows a vast range of digital reading environments, from the static digital reading investigated in the review and meta-analyses mentioned above to reading occurring in more open digital environments such as online searches, video games, collaborative sensemaking projects, and mobile phones (Coiro, 2011; Coiro, Coscarelli, Maykel, & Forzani, 2015; Davis & Neitzel, 2012; Forzani, 2018; Kinzer et al., 2012; Kinzer, Hoffman, Turkay, Gunbas, &, Chantes, 2011; Shadiev, Hwang, Huang, &, Liu, 2018). The diversity of

technological reading environments makes it hard to determine how much of the existing research findings are associated with the specific digital contexts versus with general digital reading differences. Unraveling this is important to guide policy and practice.

Therefore, to provide grounding in exploring the varied contexts, our study considers the simplest comparison of digital versus paper reading: static text content delivered via paper compared with the same content delivered on a touchscreen laptop. As we explain in our theoretical framework, the reason we focus on this comparison is that by showing differences in a simpler context with fewer confounding variables, these findings can then be applied to the more complex digital environments (i.e., with hyperlinks, animations, and videos). This choice has precedence as we follow the meta-analyses' emphasis on static reading as informing understanding of underlying differences. We also think this comparison is important as it informs the urgent policy need to understand differences between how students read traditional paper versus digital standardized tests, which, today, involve reading of bound texts (i.e., no hyperlinks or animations or videos) in digital environments and which carry high stakes. This year, Backes and Cowan (2019) noted that more than 20 states administer their yearly standardized reading assessment on a computer. Backes and Cowan (2019) also compared paper- and computer-based testing rollout in Massachusetts and found substantial negative effects for digital testing on ELA (English language arts) scores in both the first and second years. This suggests that researchers and policy makers should be concerned with the potential consequences of digital reading comprehension.

Given this landscape, we add to the literature by studying both reading processes and links to comprehension (of static reading passages) across mediums, as few studies of digital reading explore both processes and products (Singer & Alexander, 2017). We focus on reading behaviors that show evidence of the active processing that reading researchers, practitioners, and theorists emphasize is necessary for deep understanding of text and learning (Kintsch, 1994). We explore these questions with adolescents, specifically middle schoolers, because they tend to be a population where reading challenges occur, largely due to students grappling with more complex texts conveying more complex ideas and knowledge (Lee & Spratley, 2010). They are also a population where standardized tests have particularly high stakes (i.e., tracking for high school involving college-bound vs. technological tracks). Finally, they are seen as digital natives (i.e., those born after 2000, the so-called millennials) who often report preferring digital reading (Bråten, Braasch, & Salmerón, in press) and do much of their reading in digital environments both in and out of school, making understanding their varied reading experiences particularly important and relevant to policy.

Research and Theory

As we think about potential reading process similarities and differences within this bound reading context, we turn to the theory of New Literacies (Leu et al., 2013), which highlights the need to consider traditional reading theories and new digital reading theories when considering digital literacy. As Leu et al. (2013) write, "We believe that a collaborative approach to theory building is essential… because both old and new elements of literacy are layered in complex ways" (p. 1157). This theory also suggests that to build understanding of New Literacies more generally (what is considered uppercase New Literacies), we must consider how our findings fit within the larger context of reading within other digital environments. As Leu et al. (2013) write,

> What defines this broader theory of New Literacies? New Literacies, as the broader, more inclusive concept, includes those common findings emerging across multiple, lowercase theories. New Literacies theory benefits from work taking place in the multiple, lowercase dimensions of new literacies by looking for what appear to be the most common and consistent patterns being found in lowercase theories and lines of research. (p. 1157)

Therefore, we frame our work as part of the larger digital reading puzzle and we think carefully about the digital context readers are reading within. We see our study contributing findings related to a specific new literacy of reading bound, static texts on a digital screen and with digital tools that can then be integrated with findings of other new literacies to further understanding of New Literacies more broadly.

We start by thinking about the context of the reading. Using Britt, Rouet, and Durik's (2017) theory of purposeful reading, we consider "the variety of naturalistic reading situations that occur in real life . . . where there is a *task* with a *text*" (p. 3) that affect how a reader approaches reading. This aligns with the much-cited RAND reading report heuristic which emphasizes the need to consider the text, reader, and task when thinking about reading comprehension (RAND Reading Study Group, 2002). For our study, these highlight important variables to consider like a text's length and complexity, a reader's preferences, background knowledge and ability, and the task itself, which in our case, is a task a reader is likely to do often: reading an informational text passage, building knowledge and understandings of local sections of the text, and answering questions about specific elements of the text.

Next, we consider the processing, particularly active processing. Here, reading theories identify behaviors that might differ across medium. We ground our work in Kintsch's (1988) construction-integration model, which holds that readers build a representation of the surface code (i.e., the actual linguistic code and syntax), a text base (i.e., the meaning of propositions or idea units), and the situation model (i.e., building a general model of the

situation based on the text and the reader's knowledge). Importantly, such comprehension is occurring locally as readers read and make sense of small sections of text. For example, Kintsch (1988) writes, "Text comprehension is assumed to be organized in cycles, roughly corresponding to short sentences or phrases" (p. 168). Aligning with this model, our study focuses on this local comprehension rather than considering how macropropositions and inferences contribute to global comprehension (i.e., processing related to finding the main idea of a full passage rather than the main idea of a paragraph). When considering processing, Kintsch (1994) emphasizes the difference between recall and learning: "Where no adequate situation model has been formed . . ., the text can be remembered for a while, but learning . . . has not occurred. Learning requires the active construction of a situation model" (p. 302). This indicates the importance of paying attention to reading behaviors that exemplify active, constructive processing such as highlighting, looking back through the text to find an answer, or use of a dictionary to build word knowledge of unfamiliar words.

Extending Leu et al.'s (2013) lowercase new literacies theory suggests differences at each level. For example, even an identical text (same content, images, etc.) might differ in how the surface code is presented as a digital reader might zoom in on different quantities of content or scroll in ways that map the surface code differently within one's memory. When constructing the text base and situation models, which are built from linguistic input of the text and the comprehender's knowledge base, a reader might highlight relevant linguistic input, actively noting its importance to the text base and the situation model. Throughout these examples, Kintsch's (1988, 1994) work would indicate that the different active ways the material is processed leads to differences in the quality of the situation model and therefore in the learning that ensues. Leu et al.'s theory suggests these reading behaviors depend on tools, with differences expected between the fluid act of paper highlighting compared with the act of having to line up a mouse with text that is being marked digitally. Here, these differences link to the degree of active processing involved, likely yielding differences in the situation model as well as learning. Bringing together theories, we would expect similarities and differences in reading behaviors related to active processing and their link to comprehension in our new digital reading context (i.e., this specific new literacy), and these would add to other new literacies findings to indicate broader trends for New Literacies.

Reading Digitally Versus Reading on Paper

As mentioned, there is still much to be learned regarding differences between digital and paper reading comprehension. Recent reviews and meta-analyses suggest that digital reading may be inferior to paper reading on comprehension outcomes across broad age levels and study populations

(Clinton, 2019; Delgado et al., 2018; Kong et al., 2018; Singer & Alexander, 2017), but findings still show various effect sizes and do not unravel process differences. Interestingly, though, Delgado et al.'s (2018) meta-analysis of 54 studies conducted between 2000 and 2017 found that publication date was a significant moderator of the relationship of reading medium and text—that is, that the effect sizes favoring paper-based reading has increased in recent years, suggesting the digital disadvantage may be growing.

Considering our conceptual framework, studies that hold task (as RAND defines it) constant are particularly informative. Here, while findings remain mixed, the trend seems to suggest benefits for adolescents' reading of bound texts on paper versus on screens. For example, Jeong (2012) compared 56 sixth-year students in South Korea, half reading on paper, and half on bound e-books, finding superior reading comprehension on paper. Considering a slightly older population but using content from standardized reading assessments, Kim and Kim (2013) compared 108 tenth-grade students from the United States randomly assigned to read SAT passages on paper or on an LCD monitor, again finding significant benefits for paper. In Norway, Mangen, Walgermo, and Brønnick's study (2013) of 72 tenth graders from Norway randomly assigned to read the same passages modeled on the international PISA assessment either in print or on a computer screen found the same results: paper superiority. Noting possible links to types of comprehension, Fisher, Lapp, and Wood's (2011) study of 100 eighth graders found no significant differences between digital or paper texts for comprehension of main themes, but found that students reading on a screen did significantly worse on questions about supporting details, perhaps suggesting weaker local processing. Together, these studies suggest that reading on paper may be advantageous for middle schoolers' reading comprehension.

With that said, four additional studies of bound paper versus digital environments for adolescents found mixed or nonsignificant results, leading to the need for further work—and work that can unravel potential differences at a larger scale. For example, two unpublished dissertations conducted in the United States found nonsignificant results: Stevens's (2014) study of 187 middle school students did show a slight benefit to paper and Wells's (2012) study of 140 middle and high school students found almost identical results. One study (Kerr & Symons, 2006) of 60 fifth graders in the United States found mixed results: Students recalled more when reading from a computer monitor, but their improved recall came with the cost of significantly slower reading; the authors conclude that the faster reading on paper indicated stronger reading efficiency (i.e., the product of speed and accuracy). Last, Porion, Aparicio, Megalakaki, Robert, and Baccino (2016) compared 72 French students in secondary school, finding no significant differences in comprehension outcomes for either paper or digital reading. Clearly, more work is needed, and as no study has yet compared the processes (i.e., the highlighting, annotating, dictionary use, or looking back at the

text) students used in the two different mediums, work that attends to process and product is particularly important.

To get a sense of process differences, work exploring how middle schoolers read in more open digital environments (i.e., online reading with hyperlinks, Coiro, 2011; Coiro et al., 2015; Forzani, 2018; during video games, Kinzer et al., 2011; Kinzer et al., 2012) can be informative. This contrasts to the bound reading that is the focus of this study, but it is helpful because generally, most processing work has been done with these more open contexts, whereas the studies of static digital reading have primarily looked at comprehension as an outcome. While we review this work following our conceptual framing as a way to guide our hypotheses, we emphasize that different digital environments (i.e., new literacies) would likely involve different reading processes and comprehension and hence the need for our study of differences in basic digital reading.

One important finding is that traditional reading comprehension as assessed via paper standardized reading comprehension explains only part of online reading comprehension (Coiro, 2011), suggesting unique skills and strategies must be used in online reading. This may explain some of the supports found above for paper reading versus bound digital reading. It may be that middle school students have yet to develop those skills and strategies needed unique to the digital reading context. An example of such strategic reading can be seen in Davis and Neitzel's (2012) study of pairs of students reading digital (1,100–1,200 words with hyperlinks) and paper texts. Although no differences in comprehension were noted, students used more strategic actions in digital reading compared with paper reading, and most students approached the digital text as if it were linear. Further considering a different digital context, a video game, Kinzer et al. (2011; Kinzer et al., 2012) noted differences compared with reading on paper, including longer reading times during the video game, perhaps again because more strategy was needed to interpret and play in the digital context. Earlier studies of middle schoolers by David Reinking and colleagues (Reinking, 1988, 2001; Reinking & Rickman, 1990; Reinking & Schreiner, 1985) further indicated differences, but they may be less relevant as few of the participants had extensive experience with a computer, unlike students' experiences in today's digitally saturated world. What is clear from this limited work is that there is both overlap and uniqueness in the digital reading process that needs further unraveling.

Potential Mediators

The likely culprit causing variability in results is differences in digital and paper reading content and contexts. Looking across contexts (i.e., building New Literacies), certain mediators seem important. For example, when considering length, findings become more consistent. Singer and Alexander's (2017) review suggested for comprehension of short texts (i.e., under one

page of screen or book or under 500 words), there may be little difference in print or digital comprehension. In contrast, studies investigating comprehension of longer texts over one page of the screen or longer than 500 words showed improved comprehension on and preference for paper (e.g., Mangen et al., 2013). Singer and Alexander (2017) speculated that the print-based advantage for longer texts may be because of the increased cognitive and visual demands of scrolling, a speculation also made by Macedo-Rouet, Rouet, Epstein, and Fayard (2003). Alternatively, Mangen et al. (2013) hypothesized that the scrolling screen challenged students' spatial memory of where text was located on the page, impairing their ability to form a coherent text model. Another mediator may be the type of processing required by the task and text. Delgado et al. (2018) invoke what has been called the "Shallowing Hypothesis," which speculates that since increased reading on digital devices often consists of "quick interactions driven by immediate rewards . . . readers using digital devices may find it difficult to engage in challenging tasks, such as reading comprehension, requiring sustained attention" (p. 34). As several of the studies' tasks were modeled after complex passage reading and answering inferential questions typically required on standardized comprehension tests (e.g., Kim & Kim, 2013; Mangen et al., 2013), screens may inhibit the sustained attention needed to complete such a task. Overall, work exploring the processes involved in digital versus paper reading is needed to provide support for these theories and better understand potential mediators of comprehension outcomes.

When considering process, it may be that reading behaviors differ across modes or even across the tools used within each mode as our conceptual framework indicates. An example of such behaviors include annotation and highlighting, which are often taught in middle schools as a means of building active processing skills for reading comprehension. Yet, existing literature on K–12 students has not examined how annotation and highlighting may differ by medium (Singer & Alexander, 2017), and therefore, we turn to evidence from computer science. Findings here suggest students find annotation and highlighting more difficult and do it less when reading digitally as compared with paper (Liu, 2005; Morris, Brush, & Meyers, 2007; O'Hara & Sellen, 1997; Schugar, Schugar, & Penny, 2011; Thayer et al., 2011). Also, studies with adults suggest the effect of highlighting or annotation on comprehension has negative effects when reading in open digital environments with hyperlinks (Li, Tseng, & Chen, 2016), is positive only for paper (Ben-Yehudah & Eshet-Alkalai, 2014) and specifically positive for questions about higher level inferencing and processing, not lower level factual recall. These studies highlight the importance of considering quality of reading behaviors, not just quantity. Our study does this by considering quantity (i.e., how much is occurring) as well as quality (i.e., are the behaviors occurring in the sections of the text that are most important for comprehension) of digital and paper reading behaviors.

The Present Study

Policy, standards, and assessment practices are advocating a shift from paper reading to digital texts, although such a shift should not be interpreted to value one mode over the other:

> Reading digitally is part and parcel of living and learning in the 21st century. Nonetheless, there is unquestionably a place for print. . . . Reading in print or digital form should not be a horse race questions. One medium will not and should not be regarded as routinely better for comprehension. (Singer & Alexander, 2017, pp. 1034–1035)

Our study explores the simplest digital context: a paper text presented on a touchscreen as a way to provide a building block of understanding when considering more open digital contexts and as a way of better understanding the move digital in standardized tests. We investigate differences across digital and paper reading modes for (1) reading behaviors as evidenced by highlighting and annotating and (2) reading comprehension. We also explore (3) how reading behaviors within the different modes relate to comprehension taking into account reader preferences, prior content knowledge, grade, and demographics. We link to quality of behaviors, like whether the behaviors occur within important areas of the text, to explain findings. By considering process, quality, and comprehension, we build a more comprehensive picture of how digital versus paper reading occurs within this understudied population of middle schoolers.

Method

Participants

Participants were 371 fifth through eighth graders (N = 85 fifth graders, 79 sixth graders, 82 seventh graders, and 125 eighth graders) learning in the classrooms of 11 teachers within three schools in an urban district in the Southeastern United States. Eighth graders were oversampled because students were reading a National Assessment of Educational Progress (NAEP) passage designed for eighth graders. Demographic data were gathered from the district, although such data were not available for 11 of the participants. The sample included 201 females and 159 males and was relatively diverse with 152 black, 161 White, 34 Hispanic, 12 Asian, and 1 American Indian student, and 86 students were noted as economically disadvantaged (23% of students with available data). Sixty-six students reported speaking a language other than English at home. These languages included Spanish, Tamil, Amharic, Arabic, Kurdish, Vietnamese, Urdu, Somali, Persian, Rundi, Lao, Twi, Tigrinya, Chinese, French, Tagalog, Mandingo, and Portuguese. In class, students did a mix of digital and paper reading as reported by teachers. No classroom used one-to-one digital devices.

	Part 1	Part 2
Condition A	Paper μ_{11}	Digital μ_{21}
Condition B	Digital μ_{12}	Paper μ_{22}

Figure 1. **Data structure showing a within- and between-subjects design.**
Note. Each participant was randomly assigned to a condition where they read the first part of the text in a specific mode (paper or digital) and the second part of the text in the opposite mode. Note that $\mu_{part,condition}$.

Procedures

Students were interviewed individually within their schools in a quiet room or corner (i.e., library or coach's room) by a trained research team member. The content pretest was administered first orally, and students were asked to share their preference regarding reading on paper, digitally, or both. Students then read a 2011 NAEP reading passage on women's suffrage (see https://www.nationsreportcard.gov/reading_2011/testyourself_g8_passage_ann.aspx). This was divided into two sections so that the content of what students read digitally and on paper varied depending on the condition to which they were randomly assigned. To divide it, we totaled the number of words, divided by 2 and then adjusted to the closest heading so that the passage split at a point that made sense (i.e., not in the middle of a construction-integration cycle). Students assigned to Condition A read the first section on paper and the second digitally, whereas Condition B students read the first part digitally and then the second on paper (see Figure 1). The content presented in both conditions was identical. Students had access when reading on paper to highlighters, pens, and sticky notes and when reading digitally to digital highlighting, annotating, and dictionaries. The digital tools were modeled for students. Based on teacher report, students were familiar with reading on laptops and with tools that their state practice testing system used (digital highlighters included), but they had not used the digital highlighting program used in the current study previously. With that said, it was similar (cursor used to scan over and mark a highlight, then a button pushed to mark/save the highlight) as the practice standardized testing system used.

The text was divided based on where a segment ended. The first part (Part 1) was shorter (409 vs. 674 words) with slightly longer sentences (mean sentence length 19.48 vs. 14.98) than the second part (Part 2). The Lexile band for Part 1 was estimated at 1200 to 1300 compared with an 1100 to 1200 estimate

for Part 2. Therefore, length and complexity of segments were balanced. In addition, while the original text had a single image and caption, we added these elements to the other part so that each was matched for these factors.

When reading, students were told they were going to read an interesting article about Women's suffrage and asked to read like they would for school or for homework because they would be answering questions after reading to show their learning. The paper reading occurred on a front-back page, whereas the digital reading occurred on a touchscreen Surface Pro laptop with a performance base and mouse. The text did not fit on a single screen, so scrolling was necessary. Paper highlights and annotations were collected via the paper artifact. Digital highlights and annotations were collected via the hypothes.is program (see https://web.hypothes.is/). After reading both parts, students took a posttest where they saw questions on one side of the screen and the digital article (which could be scrolled on) on the other. They also had access to the paper article to allow them to look back to support their comprehension. Google Forms was used to collect pretest and posttest data.

Measures

Posttest Reading Comprehension (Outcome Variable)

Posttest reading comprehension was assessed using 14 questions (Question 1 was not used due to its general nature such that it linked to content in both Part 1 and Part 2, see the appendix for item details). NAEP questions were used when possible (five of seven were used), although two were excluded because of coding challenges due to their open response nature or because they related to global versus local comprehension. Given our theoretical framework (i.e., Kintsch, 1988), we selected questions that assessed local comprehension, yet still involved comprehension of more than a single phrase or short sentence so that a full construction-integration cycle was involved. As such, the level of mental representation assessed is that which would be featured in the situation model. Additional questions were also developed for each part (in equal numbers) using the NAEP framework such that questions aimed to assess a student's ability to locate and recall (eight items), integrate and interpret (five items), and critique and evaluate (one item).

Because the focus was on local comprehension, two question writers, who were research team members who had been middle school teachers, identified important content that involved local comprehension that was not already covered by the provided NAEP questions. One research team member then wrote the questions for this content with the second team member reviewing and adjusting the questions and then identifying the best questions such that there were equal numbers of questions for each part. These questions were next piloted by two additional team members who were master's students training to be teachers. Challenges were addressed such as wording or ambiguous answers and then the performance

of students from the first week of data collection were considered to make sure no distractors were too close. Questions involved three formats: multiple choice, true/false, and open response. Open response items were double coded for correctness by two researchers with 91.6% agreement. All discrepancies were discussed and resolved. Marginal item response theory (IRT) reliability (Green, Bock, Humphreys, Linn, & Reckase, 1984) using Rasch model was .8 for our sample of students, suggesting satisfactory reliability. Additionally, moderate concurrent criterion validity evidence for this measure was provided via its correlation of 0.64 with the standardized measure of reading achievement used by the school district, which was the Measures of Academic Progress (MAP) Reading subtest which has been shown to have reliable and valid scores (Wang, McCall, Jiao, & Harris, 2012). See Northwest Evaluation Association (2015) for more details on the MAP test. For our measure, students could have the researcher read the questions aloud, although few students took advantage of this request.

Pretest Content Knowledge/Background Knowledge

Similar to the posttest comprehension questions, a pretest of 10 items was created to assess content knowledge conveyed via the passage (see the appendix). Depending on student request, researchers either read the questions aloud and wrote responses in a Google Form or allowed the student to complete the pretest. Students were encouraged to let the researcher know when they did not know the answer to the question (rather than guess). Marginal IRT reliability using Rasch model was .78, suggesting acceptable reliability. This assessment was used to determine the participant's background knowledge of Women's Suffrage.

Preference

As part of the pretest, participants were surveyed regarding whether they preferred to read academic materials on paper, in digital environments (specifically iPads, Kindles, Computers), or in both environments. Preference was coded as two weighted orthogonal contrasts with Helmert coding (*preference0* compares paper and digital with both by coding paper = digital = 0.5 and both = −1; *preference1* compares paper with digital by coding paper = −1, digital = 1, and both = 0). The second contrast is our primary interest regarding preference as this compares the more extreme preferences of either paper or digital preference.

Reader Behaviors

Digital and paper highlighting and digital and paper annotating were each coded for number of occurrences, content, and link to the part (section of the text) and areas of interest (AOI[1]; see Appendix). Underlining was

considered a highlight. Annotations included any additional marks on the page like stars or comments. Research assistants coded all paper artifacts for these behaviors, whereas data for digital behaviors were drawn from hypothes.is. Data were checked across multiple sources to ensure reliability, including checking data with notes on in-the-moment observational coding sheets completed by research assistants doing the testing and videos collected of the digital reading screens.

Use of digital dictionary and whether students looked back at the digital and/or paper text during answering questions on the posttest was coded for occurrence (0 = did not occur; 1 = occurred). These were primarily coded by the research assistants doing the testing, but were checked with clickstream data and videos of the reading interview.

Demographics

Gender, race, special education status, ELL status, home language, and economic status information were collected from the district. Because our sample included largely White and Black students, with few Hispanic, Asian, or American Indian students, race was coded as White (reference group) versus non-White.

Table 1 presents descriptive statistics for the measures listed above.

Data Analysis

We used a within- and between-subjects design where readers served as their own controls and students were randomly assigned to which content (Part 1 vs. Part 2) they read in each mode. This allowed us to parse out differences related to the content of the text versus the mode of reading (see Figure 1). The two student groups in the randomly assigned conditions were considered random (or equivalent) groups because there were no group differences on pretest content knowledge, $t(370.93) = 1.22$, $p = .223$, and on standardized MAP reading (IRT scale) scores, $t(343.81) = -0.297$, $p = .767$, using a two-sample t-test with unequal variance at alpha = .05.

Our first research question explored differences in reading behaviors (quantity of highlighting and annotating) across modes (digital vs. paper) using paired t-tests at alpha = .05. Differences based on reader characteristics (preferences, pretest content knowledge, standardized MAP reading scores)[2] using one-way analysis of variance were also investigated. Our second and third research questions explored differences in reading comprehension across modes and links between reading behaviors (digital and paper) and reading comprehension. For parsimony, a single explanatory item response model was used[3] (De Boeck, Cho, & Wilson, 2016; De Boeck & Wilson, 2004). Here, the interaction between condition and part answered Research Question 2 and the main effects of reading behaviors answered Research Question 3. Other factors like grade, pretest content knowledge,

Table 1
Descriptive Statistics

Variable	*M* (*SD*)	Frequency (%)
Posttest total scores (*N* = 367)	8.69 (2.82)	—
Pretest total scores (*N* = 371)	3.37 (1.38)	—
Preference		
Paper	—	91 (24)
Digital	—	137 (37)
Both	—	143 (39)
Reading behaviors		
Quantity of paper highlighting (*N* = 371)	4.39 (8.62)	—
Quantity of paper annotating (*N* = 371)	0.40 (1.43)	—
Quantity of digital highlighting (*N* = 371)	2.18 (4.52)	—
Quantity of digital annotating (*N* = 371)	0.08 (0.43)	—
Use of digital dictionary		
No	—	274 (74)
Yes	—	97 (26)
Looking back digitally		
No	—	34 (9)
Yes	—	337 (91)
Looking back at paper		
No	—	49 (13)
Yes	—	322 (87)
Demographics		
Gender		
Males	—	159 (43)
Females	—	201 (54)
Missing	—	11 (3)
Ethnicity		
Black	—	152 (41)
White	—	161 (43)
Hispanic	—	34 (9)
Asian	—	12 (4)
American Indian	—	1 (0)
Missing	—	11 (3)
Special education hours (*N* = 28)	7.51 (4.28)	—
ELL status		
No		357 (96)
Yes		3 (1)
Missing		11 (3)
Home language		
English		305 (82)
Other		66 (18)
Economic status		
Disadvantaged		86 (23)
Nondisadvantaged		274 (74)
Missing		11 (3)

Note. The four remaining participants were missing some posttest responses due to Google Form issues. These were considered missing at random. Our modeling technique allowed us to use the data present from these four additional participants, resulting in 371 participants with Posttest Reading Comprehension data; "—"indicates that information is not relevant; of demographic variables, we considered Race (White vs. Non-White) in the final explanatory item response model.

preference, and demographics (retaining significant demographic covariates from a separate explanatory item response model) were controlled for as well. We created dummy variables for categorical covariates, two weighted orthogonal contrasts for preference, and mean-centered variables for continuous covariates when fitting the explanatory item response models. However, the effect code (−0.5 vs. 0.5) of the condition and the part covariates was considered in investigating the interaction effect (instead of the simple effect) of the two covariates. To further unravel Research Question 2, we estimated the model for each part (section) to explore links to text content. To better understand Research Question 3, we did post hoc analyses to explore reasons for impact of significant reading behaviors, specifically looking at the quality or whether the behavior occurred in an AOI of the text versus in a less important area of the text. Specifically, we developed AOI where we identified the essential information needed to answer the questions on our assessment. These AOI selections were guided by our research team, which consisted of three former teachers (two now professors, one a doctoral student; see Appendix).[1]

Results

Research Question 1: Digital and Paper Reading Behavior Differences

When exploring reading behaviors across modes, results indicate significant differences in highlighting and annotating behaviors were present depending on whether the behaviors occurred when reading on paper or when reading digitally. There were significantly more paper highlights (M = 5.246) compared with digital highlights, M = 2.162; $t(369) = -6.643$, $p < .001$, indicating students highlighted 2.426 times more on papers than digitally.[4] There were also significantly more paper annotations (M = 0.402) compared with digital annotations, M = 0.075; $t(370) = -4.511$, $p < .001$, which indicates that students annotated 5.360 times more on paper than digitally. We next examined whether the quantity of these behaviors differed for readers with different preferences, pretest content knowledge, and standardized MAP reading scores, which were obtained from the school district in which we worked. Results of analysis of variance suggested no differences for quantity of paper and digital highlighting nor for digital annotating. Differences were noted related to paper annotating and students' preferred reading mode, $F(2, 342) = 3.66$, $p = .027$, with participants who preferred paper reading to digital reading having on average five more paper annotations.

Research Question 2: Digital and Paper Reading Comprehension Differences

To determine whether there were differences in reading comprehension when reading digitally versus on paper, we had to take into account differences in the content of each part (section) of text. We explored whether the

mode effect (i.e., differences in digital vs. paper reading comprehension) was the same between the two parts, controlling for pretest scores, grade levels, preference, ethnicity (for which we found race was the only significant demographic covariate in the preliminary analyses),[5] and reading behaviors. Specifically, we explored whether $H_0 : \mu_{11}+\mu_{22}=\mu_{21}+\mu_{12} \Leftrightarrow H_0 : \mu_{11} - \mu_{12}=\mu_{21} - \mu_{22}$, as defined in Figure 1. Table 2 presents the results of the explanatory item response model, which indicates a significant interaction between part and condition (EST = 0.277, *SE* = 0.128, p = .031). These results suggest that the mode effect is different between the two parts (sections) of text.

The mode effect was then investigated by parts (i.e., simple-effect analysis using explanatory item response models). To do this, we looked at whether condition (effect coded) was a significant predictor of the comprehension questions present within each part. Results indicate no significant condition effect in Part 1 (EST = −0.012, *SE* = 0.118, p = .922), yet a significant condition effect in Part 2 (EST = 0.299, *SE* = 0.126, p = .017). However, the difference in Part 2 was small, suggesting paper readers performed 0.074 (=[1/(1 + exp(−0.299))] − [1/(1 + exp(0))]) higher than digital readers on the probability scale (1.077 on the odds ratio), controlling for the other covariates in the model. These results indicate that paper reading was slightly supportive compared with digital reading for the longer section of the text as the odds ratio of 1.077 is interpreted as very small effect size (Chinn, 2000).

Research Question 3: Links Between Reading Behaviors and Comprehension

Table 2 also presents the results of our explanatory item response model regarding how digital and paper reading behaviors link to reading comprehension. Results indicate that three reading behaviors significantly related to reading comprehension. Specifically, quantity of paper highlights negatively predicted reading comprehension (EST = −0.009, *SE* = 0.004, p = .029), whereas quantity of digital highlights positively predicted reading comprehension (EST = 0.027, *SE* = 0.013, p = .033). These results suggest that more paper highlighting hinders comprehension, whereas more digital highlighting supports comprehension. Additionally, whether students looked back at the paper text significantly related to comprehension performance (EST = 0.555, *SE* = 0.157, $p < .001$), whereas paper and digital annotating, use of the digital dictionary, and looking back at the digital text did not contribute to comprehension controlling for the other variables in the model.

We next explored possible reasons behind the links between digital and paper highlighting and reading comprehension. To do that, we coded AOI within the text that linked to the comprehension questions. We developed these AOI guided by experts (i.e., former teachers and current research team members coded the information necessary to answer the comprehension questions). We then determined whether the paper and digital

Table 2
Results of the Explanatory Item Response Model (Research Questions 2 and 3)

	EST	*SE*	*p*
Fixed			
Intercept	−0.179	0.344	0.603
Pretest content knowledge	**0.129**	0.037	0.001
Grade	**0.313**	0.108	0.004
Part	−0.325	0.399	0.415
Condition	0.008	0.122	0.948
Preference0	−0.098	0.070	0.164
Preference1	0.006	0.067	0.925
White	**0.570**	0.107	0.000
Highlightpaper	**−0.009**	0.004	0.029
Annotpaper	0.036	0.039	0.346
Highlightdig	**0.027**	0.013	0.033
Annotatedig	−0.007	0.120	0.953
Dictionarydig	−0.043	0.118	0.716
Lbackdigital	0.061	0.187	0.744
Lbackpaper	**0.555**	0.157	0.000
Part × Condition	**0.277**	0.128	0.031
Random			
Var(Students)	0.540		
Var(Items)	0.530		
Explained variance with student covariates		23%	

Note. EST = estimate; *SE* = standard error. A dummy variable was created for Grade and White, two weighted orthogonal contrasts were created for Preference, and a mean-centered variable was created for continuous covariates (Pretest Content Knowledge, Highlighting, and Annotating) in fitting the explanatory item response models. However, the effect code (−0.5 vs. 0.5) of the condition and the part covariates was considered in investigating the interaction effect (instead of the simple effect) of the two covariates. Significance in boldface at alpha = .05 based on Wald test.

highlighting occurred within those AOI. We then modeled the quantity of highlights in AOI and outside of areas of interest (non-AOI), graphed by condition and part of text, which allowed us to separate digital versus paper highlights. Figure 2 shows the percentage of highlighting participants in each condition (*y*-axis) who marked the number of highlights (*x*-axis) in AOI (black) and non-AOI (white). The figure indicates that for Part 1, participants in Condition A who were reading on paper had more non-AOI highlights compared with participants in Condition B who were reading digitally, and a nonparametric paired *t*-test, the sign test, indicated this difference was significant ($p < .00001$).[6] This suggests digital highlighting may be more efficient than paper highlighting. In other words, the negative

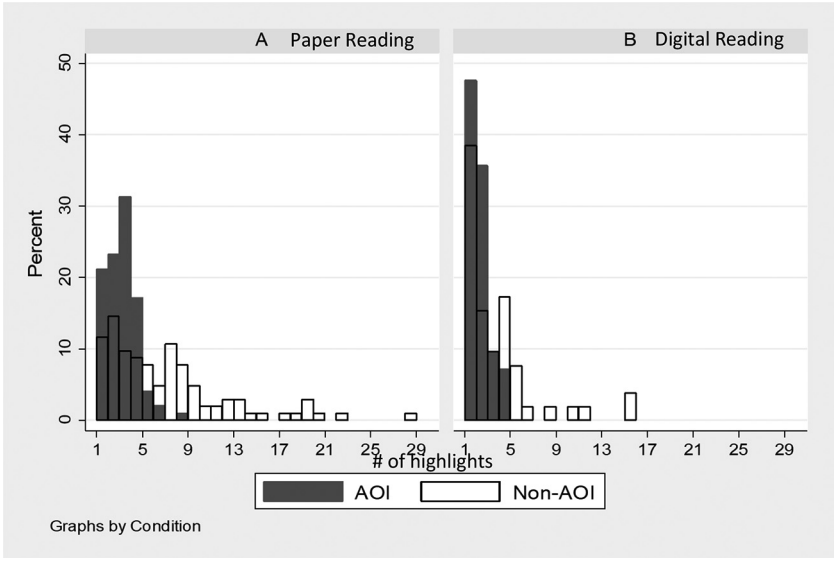

Figure 2. **Graph showing percent of participants in each condition who highlighted on the *y*-axis and number of highlights in AOI (black) and non-AOI (white) on the *x*-axis.**

Note. The figure indicates that for Part 1, participants in Condition A who were reading on paper had more highlights non-AOI compared with participants in Condition B who were reading digitally, suggesting paper highlighting may be less efficient than digital highlighting. A nonparametric paired *t*-test, the sign test, indicates this difference is significant ($p < .00001$). Similar patterns were observed for Part 2. Note that AOI = highlights in areas of interest within the text and non-AOI = highlights in text outside of areas of interest.

relationship between paper highlighting and comprehension and the positive relationship between digital highlighting and comprehension may be due to the many paper highlights compared with digital highlights that occurred outside of important information in the text (i.e., non-AOI). Similar patterns were observed for Part 2.[1]

Discussion

As a society, we are currently in a drastic literacy transition that has important consequences for policy, research, theory, and practice. In fact, many are arguing that we are living through the time in which literacy is changing the most since the advent of the printing press in the mid-1400s. Digital text is changing the content that can be communicated and also

the manner that such content can be conveyed. Our study investigates this change. We examine digital and paper reading behaviors and links to comprehension for middle school readers in a specific bound digital reading context.

We think about this at three levels. Conceptually, first we consider how traditional and digital reading behaviors and their links to comprehension overlap and are distinct. This follows our conceptual framework of building new literacy theories upon traditional reading theories—in other words, we start with what we know from reading research and practice and then consider how our newer theories (i.e., lowercase new literacy theories; Leu et al., 2013) further inform what is going on during digital reading. Second, we consider what we have learned about our own specific context, that is, the lowercase new literacy of reading a static, bound digital passage on a touchscreen computer. Last, we integrate our findings with other findings of adolescent reading in other digital contexts to inform a larger uppercase theory of New Literacies (Leu et al., 2013). We therefore consider how the similarities and differences found within our specific context (i.e., comparing reading of the same static text delivered via paper vs. on a digital touchscreen) links to what we know already about digital reading differences, including adolescent reading differences in digital contexts more generally.

In addition to adding to theory, a main contribution of our study is that it fills multiple gaps in the literature and has important policy implications, especially related to testing and instruction. It follows the recommendations of recent reviews (e.g., Singer & Alexander, 2017) to consider both reading process and links to comprehension within an authentic middle school setting. It also is guided by larger reading theories (Britt et al., 2017; Kintsch, 1994) that highlight important text, reader, and task considerations. Overall, our findings add to the literature by providing evidence of reading behavior and comprehension similarities and differences for digital and paper reading for a relatively large and diverse sample of middle school students and then linking those different behaviors to comprehension. What is clear from our study is that even when identical texts are read in different modes (i.e., a static text without hyperlinks), meaningful differences in reading exist, and these differences in processes help us understand the larger findings in the literature. As such, while the processes highlighted by traditional reading theories clearly continue to occur and are relevant to digital reading, the way in which they occur and the tools that readers use to support these processes differ in the digital context. Therefore, it is not just online reading comprehension in open environments (e.g., Coiro, 2011) that differs from traditional reading comprehension. Even considering the same content delivered differently, we identify similarities and differences that have important implications for theory, research, and practice—specifically for the design of assessments and instruction.

Reader Behavior Differences

An example of the complex layering of old and new involves in-the-moment reading behaviors noted in our study. In particular, we studied behaviors that could be implemented in both mediums (i.e., highlighting and annotating) and we found that while both behaviors occurred in both contexts with great variability among readers, on average, more paper highlighting and annotating occurred compared with when reading in the digital context. This is not surprising, as research with older populations had suggested this, but it does provide evidence that for adolescents, while they may use similar behaviors like highlighting and annotating when reading in both contexts, how they interact with text is dependent on the medium it is presented in. In other words, though adolescents are often seen as digital natives, their reading behaviors do not move seamlessly into digital environments. Taking into consideration Leu et al.'s (2013) theory of lowercase new literacies, this is likely because the tools available to mediate the reading experience differ in the digital context versus when reading on paper. We note here that more paper highlighting and annotating occurred regardless of the section of text (i.e., regardless of the complexity or length) and for the most part, regardless of reader characteristics, although those who preferred reading on paper made the most paper annotations.

As such, it seems that there is something about marking up paper that is not present in the digital condition (i.e., on the touchscreen laptop) that results in more paper highlights and annotations. There are a few possible explanations to this. It may be that highlighting or annotating with highlighters and pens is physically easier than clicking and highlighting a section of text on a digital screen. Anecdotally, we observed students struggling to digitally highlight the exact text desired (all letters of a word). Alternatively, the differences could relate to experience: Students may have had more experience marking paper text compared with digitally marking text. It may also be the way in which the content was projected on the screen (i.e., in a scrollable or zoomable manner) or prior experiences with digital content such that readers perhaps map the surface code in other ways (i.e., onto physical headings that move with the text, etc.). Interpreting this within Kintsch's (1988) model, the activation process related to integrating the surface code into the text base and situation model seems to be occurring within the reader's mind when reading digitally, whereas that integration is more visible via highlights and annotations for readers when reading paper texts. Applying Leu et al.'s (2013) new literacies theory, the differences of the digital contexts encourage the different application of these reading behaviors. We emphasize here that there are a myriad potential explanations for our findings, though. For example, it may be that there is something different about the comprehension itself in both modes that results in the different behaviors observed.

In relation to the uppercase theory of New Literacies, we could not find studies that discussed these comparable paper and digital behaviors in other digital environments, but we did note a general finding that students applied similar and different reading behaviors across digital and paper environments. The main trend appears that adolescent readers seem more strategic when reading in digital contexts, particularly when considering texts with hyperlinks or when reading in online environments. For example, pairs of readers were more strategic when reading digital versus paper texts (Davis & Neitzel, 2012). Such strategic actions may be due to the complex nature of digital reading or even because, in the same way that participants are used to shallowly reading digital texts for dozens of everyday tasks (i.e., Shallowing Hypothesis; Delgado et al., 2018), they are also familiar with the need to counteract this processing via strategic behaviors when they need to read for comprehension. For example, there is a whole literature on ways that adolescents (and readers of all ages) must be strategic when reading in open online contexts such as considering trustworthiness of authorship and accuracy of information (see Coiro, 2011; Forzani, 2018). Our study adds lenses of quantity and quality as we found that readers used fewer digital annotations and highlights, but that they were more strategic. In other words, there was less digital active processing via highlighting, but that which did occur was even more active (i.e., more strategic) because, as we discuss later, more of the paper highlights and annotations versus the digital highlights and annotations were outside of important areas of the text (i.e., non-AOI). These differences connect theories of reading related to active processing (Kintsch, 1994) with the move toward digital reading.

Digital and Paper Comprehension

Like the similarities and differences in the reading process observed via reading behaviors, we also found that digital and paper reading comprehension (assessed by responses to questions similar to traditional standardized tests) was both similar and different depending on the content of the text. Our results parallel existing findings highlighted in our literature review—mixed effects as evidenced by no effects for the first section of the text, but small benefits to paper reading for the second part of the text (which was the longer part). One way our study extends the literature is the finding that the mode of reading itself may not hinder or support comprehension of all text lengths but rather comprehension within each mode depends on the text being read, with our work providing further evidence of the difficulty of longer sections of digital text, albeit a small effect. This fits within Singer and Alexander's (2017) review that indicated no differences for shorter texts, which they defined as under 1 page of a screen or book or under 500 words. The first section of our text was under 500 words, although it was longer than a single screen. In comparison, the second section of text was over 500 words. It may be that beyond screen fit, length is an important factor to consider

when determining whether comprehension might be hindered or facilitated by mode. Therefore, part of the reason for the different findings in the literature might be because the texts being read in various studies are different (i.e., contain different content, length, and complexity). In this way, it appears that reading theories highlighting attending to differences in text, reader, and task (Britt et al., 2017; RAND, 2002) are relevant to the digital context as well.

Therefore, consideration of our specific context indicates that lengthy texts may be particularly challenging when read digitally, which is important to consider as middle school students face increasingly longer and more challenging texts as they progress through school. Fitting this within the larger new literacies research to add to the uppercase theory of New Literacies, we would expect even longer digital texts to present further challenges, perhaps multiplying the small effect into a larger effect. That may be why the effect size in our study was quite small: Part 2 text was only slightly longer than Part 1 text. An alternate explanation is that perhaps the small effect shown is too small to be practically relevant, suggesting there is not a meaningful different between digital and paper comprehension. We hope future research with longer texts will unravel this further.

Relationship of Reading Behaviors to Comprehension

Our study also found differences in how digital and paper reading behaviors linked to reading comprehension. We note here that we are interpreting correlational relationships, so we are not assigning causality to these relationships, but rather using these differences to suggest how process differences may connect to comprehension products. Our results suggest mode seems to interact with how behaviors support or hinder comprehension and this seems to relate to quality of the behavior. For example, providing additional evidence for the complex ways that the new and old elements come together in modern literacy (Leu et al., 2013), our study found that a more traditional behavior (paper highlighting) negatively related to comprehension when controlling for other covariates, even though this behavior occurred more often. In contrast, the newer behavior of digital highlighting positively related to comprehension controlling for covariates, even with less digital highlighting occurring. This emphasizes the complex relationship between quantity and quality of these new and old elements. Post hoc analyses suggest that the quality of digital highlighting was superior to paper highlighting in that students were marking many highlights outside of areas of interest, particularly when reading on paper. Linked to Kintsch's (1988) theory, it may be that these less pertinent highlights got in the way of the construction-integration process by decreasing the accuracy of the knowledge nets considered in the paper comprehension process. In contrast, the greater efficiency of highlighting when digital reading may support the construction-integration process by increasing accuracy of the knowledge

nets. Linking to the uppercase theory of New Literacies, as mentioned previously, these relationships seem to emphasize the importance of strategic behavior within digital contexts. This links to literature on comprehension strategies that emphasizes the need for conditional knowledge about why and when to use a strategy for maximum benefit with recent work suggesting such important explicit strategy instruction is limited in classrooms (Magnusson, Roe, & Blikstad-Balas, 2019). With that said, the presence of more paper versus digital highlights or the physical act of accessing text on paper, perhaps supporting spatial memory of where text was located on the page compared with when digitally scrolling seemed to make looking back for answers more helpful on paper when answering comprehension questions. Our findings indicate that whether reading digitally or on paper, students should be encouraged to consider the quality of the behavior rather than the ease.

Future Directions and Limitations

This study has implications for society's shift to more and more digital learning environments. First, clearly there is much overlap between the digital and paper reading but based on our study's findings that some active reading behaviors are different, it is important to acknowledge the uniqueness of each reading process. As such, schools need to provide middle schoolers with experiences reading in both medium, especially reading longer texts digitally as our work indicated this was slightly harder for participants than reading such texts on paper. In other words, schools cannot assume these processes and their comprehension are the same. Students need to be supported in both. This emphasis on digital reading of longer texts versus digital reading games or other learning environments with short texts can serve as an important guiding principle helping educators design learning experiences that prepare students for the different challenges between digital and paper reading.

At the same time, the cognitive demand of tools is also important to consider, especially when choosing or designing digital environments. In our study, digital highlighting was positively linked to comprehension, whereas paper highlighting was not, so it is important to consider why this might be so. While there are various ways to explain this, we observed the tactile challenge of physically highlighting digital text (i.e., moving the mouse or touchscreen cursor to the right spot and then scrolling over the section to be highlighted and pressing save). It may be that this physical demand was important in active processing, moving the information comprehended into long-term memory versus short-term recall. Alternatively, it may be the conditional nature of strategy use that renders it supportive where the need to consider when and where digital highlighting was helpful might result in more strategic strategy use. These findings extend traditional theories that emphasize the role of active processing in learning to the digital context. At the same time, we emphasize that this is correlational research, and

therefore, alternative explanations should be considered. It may be that behaviors in the two modes are related to some other variable that is also related to comprehension. For example, perhaps highlighting serves as a proxy for conditional knowledge of strategy use (i.e., when and under what conditions to highlight), which itself would correlate with comprehension. Future research should continue to explore these areas.

Our study begins to address the deictic nature of modern literacy and considers the move digital, particularly the comparison of reading text with the same content presented in different modes, which is similar to what occurs in the classroom and via standardized testing. Findings with this simplistic comparison of the most basic digital context versus traditional paper reading can provide a foundation for thinking about more complex digital reading contexts. Additionally, it can inform the move to standardized testing, particularly highlighting the importance of considering length of text and tools to facilitate reading when presenting paper content digitally, which may help address concerns about potential negative mode effects of digital reading environments, especially for students taking high-stakes reading assessments on computers (Backes & Cowan, 2019).

With that said, our study involved a single NAEP passage with five NAEP questions and nine researcher developed questions and it was focused on a single comprehension task (answering questions about local comprehension). As such, future studies should explore these questions with a larger sample of passages, questions, and also considering comprehension as assessed through different tasks (i.e., cognitive modeling tasks, application tasks, etc.). Additionally, our study considered task and text features, but primarily controlled for reader characteristics. We did add reader characteristics like gender, its interaction with comprehension of the different parts of text, and socioeconomic status to the explanatory item response model for which we reported results in Table 2, but effects were not significant and significance patterns did not change between the two models. Hence, we did not find evidence of differences by reader characteristics, but future studies should explore differences among readers in more nuanced ways.

Overall, our study highlights similarities and differences in reading processes and comprehension for middle school students performing an academic literacy task similar to what is required on standardized testing within bound conditions of having the same text read in different modes. We considered observable reading behaviors primarily involving physically marking the text or physically looking back at text. Important next steps would be to consider other ways paper reading behaviors might show up in digital reading where the markings are not permanent like, for example, sweeping the cursor under a line of text. Additionally, future work should explore how other in-the-moment reading behaviors that are less physical are similar or different across modes. For example, consideration of eye-gaze or emotional response differences may convey further understandings of reading process and comprehension

differences. Such analyses, though, are additionally difficult as data must be mapped onto content that changes when digital text is scrolled. Accessing data at scale (i.e., for 350+ participants) may involve interdisciplinary partnerships that allow such coding to occur within software. Additionally, future research should unravel differences under various digital contexts, including less bound settings of online research within more authentic tasks.

In summary, our study takes the important step of showing how reading is similar and different within a bound setting. Our findings fit within traditional reading theories (i.e., Kintsch, 1988), extend lowercase new literacies understandings, and add to the larger uppercase New Literacy theories (Leu et al., 2013). They also emphasize important practical considerations that must be taken into account when designing instruction and assessments of reading comprehension in different modes. Research indicates differences between digital and paper text comprehension, and our study begins to fill in some of the details about these differences. Leu et al. (2013) have argued, "We are on the cusp of a new era in literacy theory, research, and practice, one in which the nature of reading, writing, and communication is being fundamentally transformed" (p. 1174). Our study adds foundational understandings of digital and paper reading that can guide instructional and assessment decisions as society moves digital.

Appendix

Further Details on Posttest Reading Comprehension Measure

Posttest Item	Link to Pretest	NAEP Item	Part of Text	Heading	Format	Task
1 (not included)	*		1 and 2	General information	MC	Critique and evaluate
2	—	✓	1	Multiple	MC	Critique and evaluate
3	*		1	Introduction—Text Box	OR	Locate and recall
4	*		1	Introduction—¶2	MC	Locate and recall
5	—	✓	1	Introduction—¶3	MC	Locate and recall
6	—	✓	1	"Womanifesto"—¶1	MC	Integrate and interpret
7	*		1	"Womanifesto"—¶1	OR	Integrate and interpret
8	—	✓	2	"Wyoming is First"—¶1	MC	Integrate and interpret
9	*		1	Introduction—¶3	T/F	Integrate and interpret
10	*		2	"Wyoming is First"—Heading	OR	Locate and recall
11	—	✓	2	"Wyoming is First"—¶1	MC	Integrate and interpret
12	*		2	"Wyoming is First"—¶4	T/F	Locate and recall
13	*		2	"Wyoming is First"—¶5	OR	Locate and recall
14	*		2	"Gradual Change"—¶2	T/F	Locate and recall
15	*		2	"Gradual Change"—¶3	OR	Locate and recall

Item Content Information

Item Examples

Example A: Item linked to NAEP. 11. In her book, America's Women, what did Gail Collins suggest was the reason that the Wyoming Territory passed the first permanent suffrage law?

- ○ Because the small number of women there would not have had much political influence
- ○ Because the suffragist movement was very active in the western United States
- ○ Because most politicians there chose not to vote on that resolution
- ○ Because it was important for women to have the vote in such a dangerous area

Example B: Item with true/false format, integrate and interpret task. 9. Women could vote in certain state elections before they could vote nationally.

- ○ True
- ○ False
- ○ I don't know

Example C: Item with open response format, locate and recall task. 15. When was the first election where more women than men voted?

Example D: Item with open response format, integrate and interpret task. 7. When Elizabeth Cady Stanton wrote the Womanifesto, why did she demand equal voting rights rather than other rights like property ownership?

Examples E and F: Items linked to text sections with areas of interest underlined.

Question	Linked Passage Text (AOI)
4. About how many years did it take after the Declaration of Independence for women to earn the right to vote in every state? ○ 10 years ○ 50 years ○ 100 years ○ 150 years ○ I don't know	*That summer, the Declaration of Independence proclaimed that all men are created equal but said nothing of women's equality. It would take another 144 years before the U.S. Constitution was amended, giving women the right to vote in every state.*
6. According to the article, what was most surprising about the "Womanifesto"? ○ It was written by Elizabeth Cady Stanton. ○ It called for equal voting rights for men and women. ○ It was based on the Declaration of Independence. ○ It had such a large number of resolutions.	*The campaign for women's rights began in earnest in 1848 at a Women's Rights convention in Seneca Falls, N.Y., organized by 32-year-old Elizabeth Cady Stanton and other advocates. Stanton had drafted a "Womanifesto" patterned on the Declaration of Independence, but the one resolution that shocked even some of her supporters was a demand for equal voting rights, also known as universal suffrage. "I saw clearly," Stanton later recalled, "that the power to make the laws was the right through which all other rights could be secured."*

ORCID iDs

Amanda P. Goodwin https://orcid.org/0000-0002-6439-7399
Katherine Brady https://orcid.org/0000-0002-3699-9974

Notes

This research was made possible through funding from Vanderbilt University's Trans-Institutional Programs (see https://www.vanderbilt.edu/provost/2019/understanding-digital-dominance-in-teaching-and-learning-an-interdisciplinary-approach/) and in part from the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150199 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We would also like to thank the students and teachers who participated in the study.

[1]We confirmed these results using the same analyses with alternative AOI (called semantic or summarization AOI). These AOI were obtained from two experts (also former teachers), who were asked to mark the content that was most important to understanding the text. Results confirmed the reported findings.

[2]For pretest content knowledge, Group 1 = 1 standard deviation [SD] below; Group 3 =1 SD above; Group 2 other; for standardized reading scores, Group 1 = IRT percentile < 30, Group 3 = IRT percentile > 70, Group 2 other.

[3]A random item location parameter was modelled to explain item variability across items. Clustering of 42 class periods was ignored due to the small intraclass correlation of 0.09. Models were fit using glmer function of lme4 R package (Bates, Machler, Bolker, & Walker, 2015). Cases missing covariates were deleted. The 1% missingness in posttest responses, which was considered missing at random, was estimated via maximum likelihood estimation implemented in the glmer function.

[4]Note that we also considered the median and found that our conclusions still hold.

[5]Gender, ELL status, and special education status were nonsignificant predictors and hence dropped from the model.

[6]Because the distribution of the number of highlights is not normally distributed, the nonparametric paired *t*-test, the sign test was conducted to test the null hypothesis that median of the differences in the number of highlights between AOI and non-AOI is 0 and with the alternative hypothesis that median of non-AOI is higher than that of AOI when comparing the conditions.

References

Backes, B., & Cowan, J. (2019). Is the pen mightier than the keyboard? The effect of online testing on measured student achievement. *Economics of Education Review*, *68*, 89–103.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects model using lme4. *Journal of Statistical Software*, *67*, 1–48.

Ben-Yehudah, G., & Eshet-Alkalai, Y. (2014). The influence of text annotation tools on print and digital reading comprehension. In *Proceedings of the 9th Chais Conference for the Study of Innovation and Learning Technologies* (pp. 28–35). Retrieved from https://www.openu.ac.il/innovation/chais2014/download/program-en.pdf

Biancarosa, G., & Griffiths, G. G. (2012). Technology tools to support reading in the digital age. *Future of Children*, *22*, 139–160.

Bråten, I., Braasch, J. L. G., & Salmerón, L. (2019). Reading multiple and non-traditional texts: New opportunities and new challenges. In E. B. Moje, P. Afflerbach, P. Enciso, & N. K. Lesaux (Eds.), *Handbook of reading research* (Vol. 5). New York, NY: Routledge.

Britt, M. A., Rouet, J. F., & Durik, A. M. (2017). *Literacy beyond text comprehension: A theory of purposeful reading*. New York, NY: Routledge.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*, 3127–3131.

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading*, *42*, 288–325. doi:10.1111/1467-9817.12269

Coiro, J. (2011). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research*, *43*, 352–392.

Coiro, J., Coscarelli, C., Maykel, C., & Forzani, E. (2015). Investigating criteria that seventh graders use to evaluate the quality of online information. *Journal of Adolescent & Adult Literacy*, *59*, 287–297.

Davis, D. S., & Neitzel, C. (2012). Collaborative sense-making in print and digital text environments. *Reading and Writing*, *25*, 831–856.

De Boeck, P., Cho, S.-J., & Wilson, M. (2016). Explanatory item response models. In A. Rupp & J. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 249–266). Malden, MA: Wiley Blackwell.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer.

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review*, *25*, 23–38. doi:10.1016/j.edurev.2018.09.003

Fisher, D., Lapp, D., & Wood, K. (2011). Reading for details in online and printed text: A prerequisite for deep reading. *Middle School Journal*, *42*(3), 58–63.

Forzani, E. (2018). How well can students evaluate online science information? Contributions of prior knowledge, gender, socioeconomic status, and offline reading ability. *Reading Research Quarterly*, *53*, 385–390.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, *21*, 347–360.

Jeong, H. (2012). A comparison of the influence of electronic books and paper books on reading comprehension, eye fatigue, and perception. *Electronic Library*, *30*, 390–408.

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

KewalRamani, A., Zhang, J., Wang, X., Rathbun, A., Corcoran, L., Diliberti, M., & Zhang, J. (2018). *Student access to digital learning resources outside of the classroom* (NCES 2017-098). Washington, DC: National Center for Education Statistics. Retrieved from https://files.eric.ed.gov/fulltext/ED581891.pdf

Kim, H. J., & Kim, J. (2013). Reading from an LCD monitor versus paper: Teenagers' reading performance. *International Journal of Research Studies in Educational Technology*, *2*(1), 15–24.

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

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*, 294–303.

Kinzer, C. K., Hoffman, D. L., Turkay, S., Gunbas, N., & Chantes, P. (2011). Exploring motivation and comprehension of a narrative in a video game, book, and comic book format. In P. J. Dunston & L. Gambrell (Eds.), *Literacy Research Association yearbook 60* (pp. 263–278). Chicago, IL: LRA.

Kinzer, C. K., Turkay, S., Hoffman, D. L., Gunbas, N., Chantes, P., Chaiwinij, A., & Dvorkin, T. (2012). Examining the effects of text and images on story comprehension: An eye-tracking study of reading in a video game and comic book. In P. J. Dunston, S. K. Fullerton, C. C. Bates, K. Headley, & P. M. Stecker (Eds.), *Literacy Research Association yearbook 61* (pp. 259–275). Chicago, IL: LRA.

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

Lankshear, C., & Knobel, M. (2013). Introduction: Social and cultural studies of new literacies from an educational perspective. In C. Lankshear & M. Knobel (Eds.), *A new literacies reader: Educational perspectives* (pp. 1–18). Bern, Switzerland: Peter Lang International.

Larson, L. C. (2009). Reader response meets new literacies: Empowering readers in online learning communities. *The Reading Teacher*, *62*, 638–648.

Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy* (Final report from Carnegie Corporation of New York's Council on Advancing Adolescent Literacy). New York, NY: Carnegie Corporation of New York. Retrieved from https://files.eric.ed.gov/fulltext/ED535297.pdf

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2013). New literacies: A dual level theory of the changing nature of literacy, instruction, and assessment. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 1150–1181). Newark, DE: International Reading Association.

Li, L.-Y., Tseng, S.-T., & Chen, G.-D. (2016). Effect of hypertext highlighting on browsing, reading, and navigational performance. *Computers in Human Behavior*, *54*, 318–325. doi:10.1016/j.chb.2015.08.012

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

Macedo-Rouet, M., Rouet, J. F., Epstein, I., & Fayard, P. (2003). Effects of online reading on popular science comprehension. *Science Communication*, *25*, 99–128.

Magnusson, C. G., Roe, A., & Blikstad-Balas, M. (2019). To what extent and how are reading comprehension strategies part of language arts instruction? A study of lower secondary classrooms. *Reading Research Quarterly*, *54*, 187–212.

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.

Morris, M. R., Brush, A. B., & Meyers, B. R. (2007, October). Reading revisited: Evaluating the usability of digital display surfaces for active reading tasks. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop* (pp. 79–86). doi:10.1109/TABLETOP.2007.12

Northwest Evaluation Association. (2015). *NWEA FAQs*. Lake Oswego, OR: Author. Retrieved from https://www.nwea.org/content/uploads/2015/12/2015-MAP-Norms-FAQ-NOV15.pdf

O'Hara, K., & Sellen, A. (1997). A comparison of reading paper and on-line documents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 335–342). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.6565&rep=rep1&type=pdf

Porion, A., Aparicio, X., Megalakaki, O., Robert, A., & Baccino, T. (2016). The impact of paper-based versus computerized presentation on text comprehension and memorization. *Computers in Human Behavior*, *54*, 569–576.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/monograph_reports/MR1465.html

Reinking, D. (1988). Computer-mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. *Reading Research Quarterly*, *23*, 484–498.

Reinking, D. (2001). Multimedia and engaged reading in a digital world. In L. Verhoeven & C. Snow (Eds.), *Literacy and motivation: Reading engagement in individuals and groups* (pp. 195–221). New York, NY: Routledge.

Reinking, D., & Rickman, S. S. (1990). The effects of computer-mediated texts on the vocabulary learning and comprehension of intermediate-grade readers. *Journal of Reading Behavior*, *22*, 395–411.

Reinking, D., & Schreiner, R. (1985). The effects of computer-mediated text on measures of reading comprehension and reading behavior. *Reading Research Quarterly*, *20*, 536–552.

Schugar, J. T., Schugar, H., & Penny, C. (2011). A nook or a book? Comparing college students' reading comprehension levels, critical reading, and study skills. *International Journal of Technology in Teaching and Learning*, *7*, 174–192.

Shadiev, R., Hwang, W. Y., Huang, Y. M., & Liu, T. Y. (2018). Facilitating application of language skills in authentic environments with a mobile learning system. *Journal of Computer Assisted Learning*, *34*, 42–52.

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

Stevens, B. (2014). *E-reading comprehension versus conventional reading comprehension of junior high students* (Unpublished doctoral dissertation). Northcentral University, Prescott Valley, AZ, USA.

Thayer, A., Lee, C. P., Hwang, L. H., Sales, H., Sen, P., & Dalal, N. (2011). The imposition and superimposition of digital reading technology: The academic potential of e-readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2917–2926). Retrieved from https://depts.washington.edu/csclab/wordpress/wp-content/uploads/Thayer-Lee-Hwang-Sales-Sen-Dalal-2011.pdf

Wang, S., McCall, M., Jiao, H., & Harris, G. (2012, April). *Construct validity and measurement invariance of computerized adaptive testing: Application to Measures of Academic Progress (MAP) using confirmatory factor analysis*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, British Columbia, Canada. Retrieved from https://files.eric.ed.gov/fulltext/ED531495.pdf

Wells, C. (2012). *Do students using electronic books display different reading comprehension and motivation levels than students using traditional print books?* (Unpublished doctoral dissertation). Liberty University, Lynchburg, VA, USA.

Zipke, M. (2013). Teachers' thoughts on e-readers in the elementary school classroom. *Education and Information Technologies*, *18*, 421–441.

Manuscript received February 16, 2019
Final revision received September 30, 2019
Accepted October 4, 2019