Effects of Processing Time on Comprehension and Calibration in Print and Digital Mediums

Lauren M. Singer Trakhman, Patricia A. Alexander & Lisa E. Berkowitz

left page (even number)
this page left blank

Routledge
Taylor & Francis Group

Check for updates

Effects of Processing Time on Comprehension and Calibration in Print and Digital Mediums

Lauren M. Singer Trakhman ⬤, Patricia A. Alexander, and Lisa E. Berkowitz

Department of Human Development and Quantitative Methodology, University of Maryland, College Park, Maryland

ABSTRACT

This study explored the effects of processing texts in print or digitally on readers' comprehension, processing time, and calibration. Eighty-six undergraduates read print and digital versions of book excerpts about childhood ailments presented in counterbalanced order. Comprehension was tested at three levels (i.e., main idea, key points, and other relevant information). Direct comparisons between print and digital reading demonstrated a significant advantage for reading in print on students' recall of key points and other relevant information but not the main idea. When processing time was added as a mediator variable, it significantly affected the relation between medium and comprehension for all question levels. In terms of calibration, students read more quickly and judged their performance higher when engaged digitally, although their actual performance was much better when reading in print. Implications of these findings for subsequent research are considered.

IT GOES WITHOUT saying that we are citizens of a digital world. From accessing breaking news via tweets and videoconferencing with colleagues to finding directions on a smartphone or preparing reports on a computer or tablet, advances in technology have transformed the lives of those who inhabit industrial and post-industrial societies (Castrells, 2011). Further, students in the digital world interact with technology now more than ever (Baym, 2015; Dudeney & Hockly, 2016). Whether using digital textbooks in their classroom, sharing notes on a digital bulletin board, or progressing through lessons in an individualized e-learning software program, students are increasingly relying on digital resources within academic contexts (Purcell, Heaps, Buchanan, & Friedrich, 2013).

As technological hardware and software expand and improve over time, so too has the research on these technological advancements and their potential effects on student learning and performance (Clark & Mayer, 2016; McKnight et al., 2016; Mohammadyari & Singh, 2015). One such avenue of inquiry spawned by technological advancements pertains to the relative advantages and disadvantages of reading in print versus digitally on text processing and comprehension outcomes (Singer & Alexander, 2017a, 2017b; Farinosi, Lim, & Roll, 2016; Myrberg & Wiberg, 2015). Within this growing literature, differences between print and digital mediums have been found in terms of reading comprehension and speed of processing (e.g., d'Haenens, Jankowski, & Heuvelman, 2004; Mangen, Walgermo, & Brønnick, 2013).

For instance, in a study involving 72 tenth-graders in Norway, Mangen et al. (2013) found that students who read print versions of text had significantly higher scores in reading comprehension than students who read digitally. In that investigation, students read expository and narrative texts in print and digital forms. After reading the passages, comprehension was measured by multiple-choice and short-answer questions. As in the Mangen et al. (2013) study, we (Singer & Alexander, 2017a) also

explored potential differences in comprehension when undergraduate students read newspaper articles and book excerpts about childhood ailments (e.g., asthma or autism) in print and digitally. Unlike the Mangen et al. (2013) study, however, we examined comprehension differences by medium for three types of questions: main idea, key points, and other relevant information. While no differences in comprehension emerged by medium for the main idea question, there was a significant medium effect for students' recall of key points and other relevant information as the source of comprehension differences by medium. This finding suggested that future studies would need to consider the type of questions included on comprehension assessments.

Question type is not the only relevant variable that has been underinvestigated in the literature on medium effects. There is also a paucity of empirical studies that have expressly investigated time as a potential explanatory factor in comprehension differences for print versus digital reading. In fact, in a systematic review of the literature (Singer & Alexander, 2017b), we were unable to address time as a factor in analysis because few studies (i.e., less than 25%) had included any measure of processing time. Moreover, for several of the studies in which time was documented, processing time was either controlled (e.g., Ackerman & Goldsmith, 2011) or imprecisely measured (e.g., Gill, Mao, Powell, & Sheidow, 2013; Noyes & Garland, 2003). In one investigation in which processing time was carefully measured in relation to comprehension, Van de Vijver and Harsveld (1994) found that applicants to the Dutch Royal Military Academy completed the assessment battery faster on the computer than on paper. However, comprehension performance for these applicants was superior in the print versus digital condition.

Although research in print and digital reading comprehension often overlooks the potential effect of processing time on comprehension, there is literature within the perceptual-motor domain that suggests that speed of processing is related to quality of performance. Notably, the speed-accuracy tradeoff hypothesis (Wickelgren, 1977) contends that there is a tradeoff between how fast a task is performed and the quality or accuracy that results. For example, Dyson and Haselgrove (2000) examined comprehension performance of undergraduates who were asked to read articles from *National Geographic* at both normal (244 words per minute) and fast (460 words per minute) rates. What these authors determined was that comprehension performance was significantly better at normal than at fast rates. Applying the speed-accuracy tradeoff hypothesis to the current investigation, we would predict that students would read faster in the digital condition than in the print condition and that this speed of processing would contribute to diminished comprehension performance in the digital condition.

Another outcome attracting increased interest in comprehension research relates to the effects on calibration. Calibration refers to the degree of association between individuals' predicted and actual performance (Alexander, 2013; Fischhoff, Slovic, & Lichtenstein, 1977; Glenberg, Sanocki, Epstein, & Morris, 1987). In general, the calibration abilities of children and youth for a range of academic tasks and subject-matter domains have been described as poor (Chen & Zimmerman, 2007; Hacker, Bol, & Bahbahani, 2008; Pajares & Miller, 1997), suggesting that students are not particularly astute in critically judging their learning or level of performance. However, less is known about calibration when reading in print versus in digital mediums.

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

What might account for such marked miscalibration when reading occurs digitally? One plausible explanation for such a pattern comes from the work of Koriat, Ma'ayan, and Nussinson (2006) on the

association between level of effort exerted in task performance and the judgments of learning (JOLs) made. Specifically, in the Koriat et al. investigation, the less effort participants exerted in task performance, the higher their JOLs. Ackerman and Goldsmith (2011) reported a similar pattern and concluded that the ease and speed of reading digitally may have contributed to the miscalibration they witnessed by negatively affecting students' regulatory behaviors.

Current investigation

In undertaking this study, we wanted to revisit the Singer & Alexander (2017a) investigation of medium effects on comprehension for purposes of replication and refinement. The two facets we wanted to replicate pertain to levels of comprehension and text features. For one, both previously and currently, we framed this investigation according to the assumption that comprehension can manifest at differing levels of specificity requiring varied cognitive processes (Burton, 2001; National Center for Education Statistics, 2013; Pearson & Hamm, 2005) from a more global understanding, as might be measured by questions pertaining to main idea, to the recall of specific content within the text. Moreover, we expected those levels of generality or specificity to matter in students' performance in print and digital conditions.

Further, in the current investigation as in the prior study, we aimed to control features of the text to make the conditions across medium as similar as possible. For example, we controlled the text length to fit onto one page. Our rationale for this decision came from studies that revealed an effect for medium when texts exceeded one page (Wästlund, 2007; Wästlund, Reinikka, Norlander, & Archer, 2005). It was Wästlund et al.'s (2005) contention that the need to scroll with longer online texts increased the cognitive demands on readers and, thus, appeared to negatively affect recall for digital medium. Thus, we again chose to restrict text length to one page for both print and digital mediums so as to remove any potential confounding effects that might favor print over digital reading.

The principal refinement we introduced in the current study was the inclusion of reading time, which we sought to relate to comprehension and calibration outcomes. In keeping with the speed-accuracy tradeoff hypothesis, we expected that students would read faster digitally but score lower on a measure of comprehension. Moreover, if the conclusions of Ackerman and Goldsmith (2011) are correct, then we should see faster processing speeds and higher judgments of performance corresponding to the digital reading condition and higher comprehension outcomes corresponding to digital reading.

Thus, as it relates to processing time, calibration, and comprehension under print and digital reading conditions, we specifically sought to examine the following questions:

1. (a) To what extent does overall comprehension performance differ when competent readers process texts in print versus digitally? (b) Does this performance difference by medium exist when examining comprehension at varied levels of specificity (i.e., main idea, key points, and other relevant information)?

We predicted that there would be a difference in comprehension by medium type, which would favor reading in the print condition. However, we expected that this difference would arise largely from responses to questions that extended beyond gist understanding of the text's main idea to the identification of key points and other relevant information.

2. (a) Are there differences in processing time when reading occurs in print or digitally? (b) To what extent does processing time mediate the relation between medium and comprehension scores? (c) Does processing time serve as a mediator between medium and comprehension performance for individual questions varying in specificity (i.e., main idea, key points, and other relevant information)?

Based on prior research (Singer & Alexander, 2017a), we predicted that there would be differences in the time spent reading in print and digitally. Further, in keeping with the speed-accuracy tradeoff hypothesis, we hypothesized that there would not only be direct effects of medium on comprehension but processing time would also serve as a mediator between medium and overall comprehension performance. With regard to this mediation model, we predicted that reading digitally would result in less reading time for that medium, which would then influence how well students could recall what was

read. Moreover, we expected the mediation effect of processing time to be significant for the individual comprehension questions that were more specific (i.e., key points and other relevant information) rather than more global (i.e., main idea) in nature.

    3. How well calibrated are competent readers when processing texts in print versus digitally?

Drawing on the broader calibration literature, we expected the competent readers in this study to be poorly calibrated overall. Nonetheless, in light of the potential influence of processing speed, we hypothesized that there would be a greater calibration disparity when students read digitally than when they read in print.

Method

Participants

Participants for this study were 86 undergraduate students enrolled in human development and educational psychology courses at a large mid-Atlantic university. The sample was 63.1% female with a mean age of 19.78 $(SD = 1.82)$ years. The sample was White (52.4%), Asian (17.9%), and African American (16.9%) and represented a wide variety of majors, primarily in the social (48.6%) and natural (28.8%) sciences. The demographics of participants in this study are comparable to the overall demographics of the college in which the study was conducted.

There were several reasons for our choice of undergraduates as study participants. For one, we were interested in the performance of competent readers in this study rather than those who may still be acquiring the fundamentals of reading. In effect, we did not want problems in such basic skills as decoding or fluency to play a role in the participants' reading time—a focal point of this investigation. Another reason was that over 75% of the studies analyzed in our recent review of the empirical literature on print and digital reasoning involved undergraduate readers (Singer & Alexander, 2017b); therefore, to compare outcomes of the current investigation to the extant literature, we felt that undergraduate readers would better serve our goals. Finally, individuals of this age level would fit the profile of *digital natives* (Prensky, 2001), which should increase the likelihood that they are familiar with the Internet and with digital reading.

Experimental Texts

Two experimental texts about childhood ailments were used in this study. The choice of this topic area was made for several reasons. First, childhood health topics would presumably be of interest to study participants because the passages pertained to topics relevant to human learning and development. Further, the topic of childhood ailments was germane to the participants' coursework without being a topic that would specifically be covered in those courses. The two book excerpts were from *Healing the New Childhood Epidemics: Autism, ADHD, Asthma, and Allergies: The Groundbreaking Program for the 4-A Disorders* (Bock & Stauth, 2008). All texts were similar in length (approximately 550 words) and readability level (approximately 8.5 grade level; Fry Readability: Fry, 1968).

Prior to administration, the two passages were modified to ensure that the main idea, while not explicitly stated, could be inferred from each text. Modifications were also made so that the frequencies of key points $(n = 4)$ and other relevant information $(n = 6)$ were equivalent across the texts. Two research assistants familiar with text analysis were given the modified passages and instructed to identify the main idea, all key points, and other relevant information. For this task, these assistants were instructed that the main idea represented the overarching topic or theme of the excerpt, whereas the key points were text-based units (i.e., independent clauses or sentences) linked directed to that main idea. All remaining units that were not supportive of the main idea but still pertinent to the overall text topic were classified as "other relevant information." After the assistants completed their analysis, interrater agreement was determined to be

98.5%. The only point of disagreement was with one of the "other relevant information" units and that disagreement was resolved through discussion.

Comprehension scoring rubric

Once the main ideas, key points, and other relevant information were reliably identified for both text excerpts, a scoring rubric for the comprehension test was created. For that rubric, each main-idea question was scored on a 0–2 scale for each passage read for a maximum score of 4. A score of 0 was earned if participants wrote nothing or provided a main idea that was entirely incorrect. If the students' answer to the main-idea question indicated some relevance to the established main idea but did not fully convey that central theme, a score of 1 was awarded. For example, for one excerpt, the main idea was established as: "A mother struggles to accept that her daughter has serious behavioral problems." A score of 1 was awarded to a student who wrote: "the daughter has ADHD." To earn 2 points, participants had to produce a statement that approximated the pre-identified main idea. For instance, one student wrote that the text was about "this mother, Liza, who would not listen to what doctors said about her daughter's ADHD." This answer was awarded a 2.

The two remaining questions pertaining to key points and other relevant information were marked against the scoring rubric on a 0 (absent) or 1 (present) scale. No points were subtracted for any incorrect content that might appear in response to either comprehension question. Thus, participants could earn up to four points per passage for key points, with a possible maximum score of 8 points. The "other relevant information" question had a total score of 6 points per text for a maximum score of 12 points.

To establish interrater agreement, the first and third author and a second rater scored all responses independently and compared their scores. Prior to this independent scoring, a set of training materials with simulated responses paralleling those read in the actual study were developed. Once the second rater reached a scoring proficiency level of 90% or above on the training materials, the actual data were introduced. Interrater agreement for the full data set, scored independently, was 98.9%. Any disagreements were resolved through discussion.

Measures and Variables

Demographic survey
Prior to beginning the print and digital reading portion of 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 two topics (i.e., childhood ADHD and allergies). These ratings were made on a 100 mm scale that ranged from 0 (no knowledge) to 100 (expert). For example, respondents were asked to "rate your knowledge on the subject of childhood ADHD." If students marked their knowledge of childhood ADHD at the 25 mm point on the line, they received a self-reported score of 25 on the topic knowledge measure for childhood ADHD. The students repeated this process for the second topic (childhood allergies).

Reading comprehension test (RCT)
We were interested in examining the effects of medium (i.e., print and digital) on students' comprehension. Prior to reading, participants were told they would read a passage and be asked questions about the main idea, key points, and other relevant information after reading without the text available. To assess their comprehension, students responded to three short-construction items immediately after reading each of the texts. The questions were presented in the same medium in which the text was just

read. For example, after reading about childhood ADHD digitally, participants were asked to complete the following items online:

1. Explain the main idea of the passage.
2. List the key points of the passage.
3. Jot down any other relevant information you may remember.

These same three questions were presented to students after reading the printed text, but they were displayed on paper.

Judgment of performance (JOP)

In order to assess students' calibration ability, 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. For example, after reading of a text and completing the comprehension test in the print (or digital) condition, a student answered the following question presented on a sheet of paper (or online):

Please rate your performance on the reading comprehension test you just completed in print on a scale of 0 (recalled nothing) to 10 (recalled everything).

Processing time (PT)

As noted, a primary objective of this investigation was to consider the role that processing time might play in any comprehension differences by medium. To calculate processing time under both medium conditions, we used the timer feature on a smart device to record the length of time participants spent reading each text. The timing began as soon as the directions for the task were given and the text excerpt was presented to students. The timing stopped as soon as the student indicated they had completed the reading. So as not to influence processing time in any way, students were not expressly told that their reading time would be monitored and the timing device was kept out of their visual field during task performance.

Calibration

We examined the relation between judgment of performance (JOP) and demonstrated performance on the comprehension measure (RCT) both descriptively and quantitatively. For both of 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-rating 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.

Quantitatively, we created a 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 was indicative of a more calibrated reader.

Procedure

The study was conducted outside of class in a designated area. Students completed the demographics survey and the self-reported topic knowledge measure digitally. Then, they moved on to the reading portion of the study. Each participant read two texts in total: one text from each medium (i.e., print and digital) concerning childhood ADHD and childhood allergies. These texts were presented in counterbalanced order.

Prior to receiving the first text (i.e., childhood ADHD or allergies) in either print or digital form, participants were instructed as follows:

> *You are going to be given a passage to read and asked questions about the main idea, key points, and other relevant information 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 the students began reading, a research assistant started a timer to begin logging processing time. When participants indicated they had completed the reading, the research assistant stopped their timer and presented them with the comprehension test in the same format as the passage they had just read (i.e., print or digital). Participants 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, participants judged their performance in that medium.

Equipment

Participants completed the digital portion of the study using a 15" LCD monitor at a resolution of 1,280 × 1,024 pixels. We assumed that these computers would be familiar to participants as they are the computers that students currently and routinely use 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 paperback book, where the modified passages had been inserted and made to appear as part of the volume. The designated passage was bookmarked and only relevant sections were visible to the reader.

Results and discussion

Descriptive statistics

The first step in analysis was to calculate means and standard deviations for all study measures and variables to ensure that these data met the assumptions for subsequent statistical analyses. The data, which appear in Table 1, were found to meet all assumptions so that further analysis could proceed.

One of the descriptives we calculated was self-reported topic knowledge. Before addressing the specific research questions guiding this study, we wanted to ensure that participants were not overly familiar with either topic and that the two topics were similar in this regard. As seen in Table 1, the self-ratings for both texts represented a moderate level of topic familiarity. Further, the level of topic familiarity across the two topics was nonsignificant, $F(1) = 1.19$, $p = .09$. Overall, these results reinforced our expectation that students would not be particularly knowledgeable about the topics of

Table 1. Means and standard deviations for study measures and variables.

Measures	Descriptive Statistics		
Variables	*M*	*SD*	Maximum
Topic Knowledge			
ADHD	35.91	23.87	100
Allergies	36.84	24.37	100
Processing Time[*]			
Print	2.06	.67	6.25
Digital	1.78	.51	4.58
Judgment of Performance			
Print	6.82	1.08	10
Digital	7.94	1.19	10
Calibration[**]			
Print	18.59	23.30	100
Digital	26.83	25.25	100

[*]Processing time was calculated in minutes.
[**]Calibration represents the difference between predicted and actual comprehension performance, so the lower the number the better the calibration.

childhood ADHD and allergies and that the two topics were statistically comparable with regard to the participants' familiarity with them. Therefore, we determined that comprehension performance in this study would not be unduly influenced by participants' existing knowledge of the text content.

Research Question 1: Comprehension across mediums

The first research question guiding this study focused on the role of medium in students' overall reading comprehension performance and in their performance on questions differing in specificity (i.e., gist to specific details). Data on the comprehension test overall and by question level are displayed by text topic and by medium in Table 2. It should also be noted that while we did not conduct any specific comparisons of the word counts of the participants' responses across print and digital conditions, the length of those responses appeared comparable. This comparability appeared to arise largely from the fact that most participants responded in phrases or bulleted statements rather than in connected discourse.

1(a) Overall comprehension

To determine the effect of medium on overall comprehension, we first collapsed data across the two experimental texts for the print and the digital conditions. This step was viewed as justified due to the structural comparability of the two texts, students' similar level of topic familiarity, and the counterbalanced order of text by medium presentation. Descriptive data showed that participants performed somewhat higher in overall comprehension when reading in print than when reading digitally ($M = 9.58$, $SD = 2.41,1'$]] and $M = 8.11$, $SD = 2.84$, respectively). Based on a one-way ANOVA, this difference by medium for overall comprehension was found to be nonsignificant ($F(1) = 10.49$, $p = .06$). Thus, our prediction of higher comprehension scores for print over digital reading was not upheld.

1(b) Levels of comprehension

Next, we examined the effects on medium by question level (i.e., main idea, key points, and other relevant information). As in the prior analysis, we collapsed data across text topics to focus on the mode of delivery (i.e., print versus digital). The mean differences for identifying the main idea when reading in print ($M = 1.78$, $SD = .19$) and digitally ($M = 1.71$, $SD = .12$) were nonsignificant ($F(1) = 11.31$, $p = .92$). Overall, the competent readers in our study were able to identify the main idea fairly accurately regardless of whether they were processing the texts in print or in digital form.

However, comprehension differences by medium became evident when participants attempted to recall the key points of the texts and other relevant information. Specifically, when participants accessed text in printed form, they recalled an average of 3.7 ($SD = 1.9$) key points, which contrasted with the mean score of 2.8 ($SD = 2.1$) when reading digitally. This performance difference by medium for key points was statistically significant ($F(1) = 5.99$, $p < .05$, $\eta^2 = .07$).

Table 2b. Means and standard deviations for reading comprehension by topic and by medium.

Comprehension Score Level	Childhood ADHD		Childhood Allergies		Combined	
	M	SD	M	SD	M	SD
Print						
Total	9.81	2.28	9.35	2.54	9.58	2.41
Main Idea	1.83	0.17	1.73	0.21	1.78	0.19
Key Points	3.83	0.92	3.57	0.98	3.70	1.90
Other Information	4.51	1.19	4.05	1.35	4.10	1.40
Digital						
Total	8.08	2.78	8.14	2.90	8.11	2.84
Main Idea	1.65	0.13	1.77	0.08	1.71	0.12
Key Points	2.83	1.38	2.77	2.02	2.80	2.10
Other Information	3.60	1.19	3.60	0.80	3.60	0.80

As for the comprehension question requesting any other relevant information that respondents could recall, there was again a significant difference by medium. Specifically, when reading in print, participants identified an average of 4.1 ($SD = 1.4$) "other relevant" points compared with an average of 3.6 ($SD = 1.6$) when reading digitally. This difference for this level of comprehension was statistically significant ($F(1) = 12.14$, $p < .05$, $\eta^2 = .13$). Consistent with our hypothesis, these findings for "key points" and "other relevant information" questions suggest that the medium in which even college students are reading is more influential when the questions being answered go beyond a gist understanding of the text.

Research Question 2: Processing time and comprehension

The second research question for this study examined the time participants spent reading on paper or digitally in two ways. First, we wanted to ascertain whether significant time differences by medium emerged in the current investigation. Second, should such differences be manifested, we wanted to examine whether processing time served as a mediator between medium and comprehension both for overall performance and for the individual questions that varied in specificity.

2(a) Time differences by medium

We analyzed overall processing time by medium to determine whether reading digitally was faster than reading in print, as has been indicated in prior studies (e.g., Van Vijver & Harsveldt, 1994). The descriptive data for processing time by medium are presented in Table 1. As these data indicate, participants spent an average of 2.06 ($SD = .67$) minutes reading in the print in contrast to 1.78 ($SD = .51$) minutes reading digitally. This difference was statistically significant ($F(1) = 7.75$, $p < .05$). The result confirmed our hypothesis that these college students would read significantly faster when texts were displayed on computer than when texts were displayed on paper.

2(b) Mediation model for overall comprehension

Given that processing time was, in fact, different across mediums, we turned to its role as a mediation variable in comprehension performance both overall and by question type. For these analyses, we tested a mediation model because we hypothesized that there would be both direct and indirect effects for medium on comprehension performance. In terms of the indirect effects, we expected that reading faster on the computer would translate into lower scores on the comprehension measure. Due to our relatively small sample size for this analysis, we chose to use Process Model 4 (Hayes, 2013). The Process Model 4 employs a bootstrapping method considered most effective and powerful for smaller samples and thus less vulnerable to Type 1 errors (Preacher & Hayes, 2008). With this approach, in which data are resampled 5000 times, a model's fit is confirmed when the confidence interval falls above or below zero (Hayes, 2013).

The first mediation model we tested focused on processing time in relation to overall comprehension scores. As Figure 1 illustrates, there was a significant direct effect of medium on overall
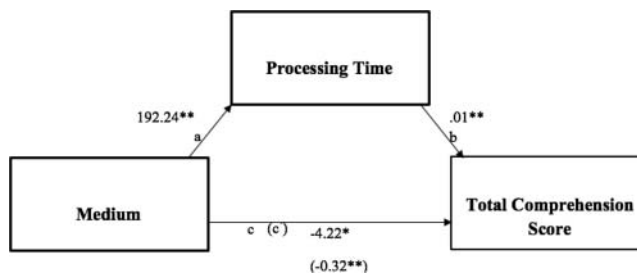

Figure 1. The mediation model indicates that processing time was an appropriate mediator between medium and total comprehension scores ($a = 192.24$, $p < .01$; $b = 0.01$, $p < .01$). Further, it shows the significant direct and indirect effects of medium on total comprehension scores ($c = -4.22$, $p < .01$; $c' = -0.32$, $p < .01$).
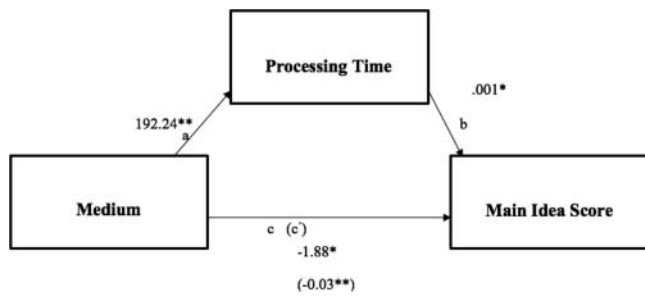
Figure 2. As indicated in this mediation model, processing time served as an appropriate mediator between medium and main idea scores ($a = 192.24$, $p < .01$; $b = 0.001$, $p < .05$). Further, this figure shows the significant direct and indirect effects of medium on main idea scores ($c = -1.88$, $p < .05$, $c' = -0.03$, $p < .01$).

comprehension scores ($F(1,171) = 24.46$, $p < .01$, $R^2 = .23$). In addition, medium predicted processing time ($\beta = -32.71$, $t(171) = -2.00$, $\mathrm{p} < .05$), which then predicted comprehension scores ($\beta = 192.25$, $t(171) = 11.28$, $p < .01$). As confirmed by the fit indices (95% confidence interval (CI) $= -5.65$–2.79), processing time significantly mediated the effects of medium on the total comprehension score, explaining about 23% of the variance in overall performance.

2(c) Mediation models for individual questions

Next, we ran separate mediation models using Process Model 4 (Hayes, 2013) to test whether processing time was a significant mediator for each of the comprehension questions in our study. For the question regarding the main idea, which tapped participants' gist understanding of the information texts there was a significant direct effect of medium ($F(1,171) = 4.02$, $p < .05$, $R^2 = .023$), as shown in Figure 2. In addition, as confirmed by the fit indices (95% confidence interval (CI) $= -2.40$–1.20), processing time was a significant mediator between medium and performance on the main idea question. Specifically, medium predicted processing time ($\beta = -1.88$, $t(171) = 0.001$, $p < .01$), which then predicted comprehension scores ($\beta = 192.25$, $t(171) = 11.28$, $p < .01$). Thus, our hypothesis that processing time would only be a significant mediator for the more specific comprehension questions in this study was not confirmed. It should be noted, however, that this mediation model explained only about 2.3% of the variance for performance on the main idea question.

For the question regarding key points, there was a significant direct effect of medium ($F(1,171) = 22.50$, $p < .01$, $R^2 = .21$). For this model, processing time was found to be a significant mediator (95% confidence interval (CI) $= -1.91$–0.89) on scores for the key points question. As displayed in Figure 3, medium predicted processing time ($\beta = -1.80$, $t(171) = 0.001$, $p < .05$), which then predicted comprehension scores ($\beta = 192.25$, $t(171) = 11.28$, $p < .01$). This mediation model explained 21% of the variance for scores on this particular item.

Finally, as displayed in Figure 4, there was a significant direct effect of medium on the question targeting other relevant information scores ($F(1,171) = 24.46$, $p < .01$, $R^2 = .023$). In addition, medium
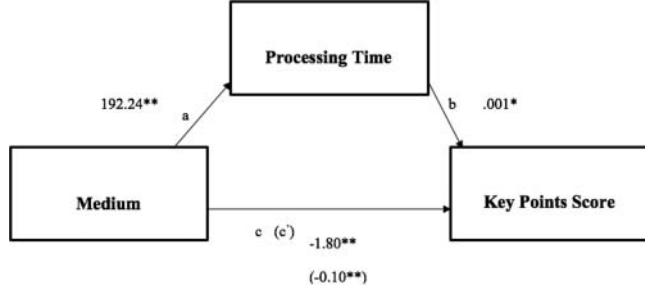

Figure 3. As this model indicates, processing time served as an effective mediator between medium and key points scores ($a = 192.24$, $p < .01$; $b = 0.001$, $p < .05$). Further, the model shows the significant direct and indirect effects of medium on participants' recall of the key points in text ($c = -1.80$, $p < .01$; $c' = -0.10$, $p < .01$).
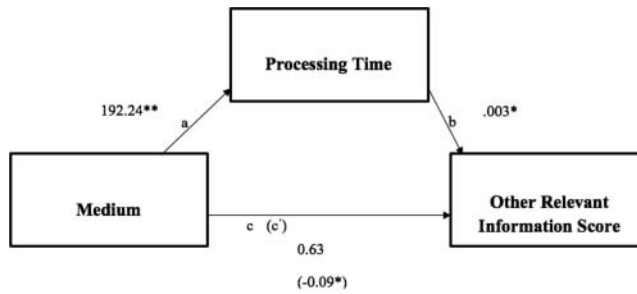
Figure 4. This mediation model demonstrates that processing time was an appropriate mediator between medium and performance on the other relevant information question ($a = 192.24$, $p < .01$; $b = 0.003$, $p < .05$). In contrast to the models for main idea and key point questions, medium only had a significant indirect effect on other relevant information scores ($c = 0.63$, $p = .09$; $c' = -0.09$, $p < .05$).

predicted processing time ($\beta = -4.22$, $t(171) = -5.84$, $p < .01$), which then predicted comprehension scores ($\beta = -0.32$, $t(171) = 11.28$, $p < .01$). As confirmed by the fit indices (95% confidence interval (CI) $= -0.32$–0.01), processing time was again found to be a significant mediator on participants' recall of other relevant information from the texts although this model explained only 2% of the variance for this specific question.

Overall, as we hypothesized, processing time was a significant mediator between medium and participants' performance on the more specific questions on the comprehension measure. However, the variance explained by these mediation models was moderate (key points) to small (other relevant information), suggesting that other factors are contributing to these comprehension outcomes.

Research Question 3: Calibration

The final question we sought to address in this study dealt with participants' calibration ability when the participants read in print and digital mediums. The descriptive data for this variable are shown in Table 1. As these data indicate, participants' evaluations of their performance when they read digitally were higher than when they read in print ($M = 7.94$, $SD = 1.19$ and $M = 6.82$, $SD = 1.08$, respectively). In line with our hypothesis, this difference was statistically significant ($F(1) = 6.81$, $p < .05$, $\eta^2 = .19$). Moreover, as we (Singer & Alexander, 2017a) and others (Ackerman & Goldsmith, 2011) have reported, this self-evaluation did not parallel actual performance for "key points" and "other relevant information" questions, which was significantly better in the print condition.

To examine this outcome further, we sorted participants into three categories based on calibration accuracy. For this categorization, conducted separately for print and digital reading, participants' calibration was coded as (a) accurate (self-ratings and comprehension scores within 3 percentage points), (b) underestimation (self-ratings more than 3 percentage points below comprehension scores), or (c) overestimation (self-ratings more and 3 percentage points above comprehension scores). Several interesting patterns emerged from this categorization. First, most students tended to overestimate their performance regardless of the medium in which they read, although the frequency of overestimation was noticeably higher for the digital condition. Specifically, 70.93% ($n = 61$) of participants overestimated their performance when they read texts in print, as compared with 90.7% ($n = 78$) when participants processed the texts digitally. Underestimation was relatively infrequent for these competent readers regardless of condition, involving only 6.98% ($n = 6$) of participants when they read in print and 9.3% ($n = 8$) when they read digitally. Finally, and perhaps not surprisingly, given their tendency to overestimate when they processed texts digitally, no participants among the 86 accurately judged their performance in that condition. By contrast, 22.09% ($n = 19$) of these competent readers' self-ratings of performance were within 3 percentage points of their actual comprehension score when they read in print.

Finally, we tested whether calibration scores—as determined by subtracting by JOP ratings from RCT scores—differed by medium, as would seem likely given the pattern in the descriptive data shown in Table 1. When reading in print, participants had a mean calibration score of 18.59 ($SD = 23.30$), whereas that score was noticeably higher when they read digitally 26.83 ($SD = 25.25$). Again, it should be noted that the lower the calibration score, the better the readers' calibration. This difference by medium was statistically significant ($F(1) = 4.94$, $p < .05$, $\eta^2 = .04$). Thus, as we hypothesized, when participants in our study read digitally, their calibration was significantly worse than when they read in print.

Conclusions and implications

In the current study, we set out to look more deeply at competent readers' processing of texts in print and digitally. Our goal was to examine the role that the medium of reading played in how students comprehended text and evaluated their performance. Also, we sought to ascertain whether processing time affected undergraduates' comprehension and judgments of learning when the medium of delivery varied. However, before discussing any insights garnered from this investigation, there are limitations that should be acknowledged. For one, in the current investigation we relied on self-reported topic knowledge in determining students' familiarity with the text content. Thus, it remains an empirical question as to whether a more dedicated measure of topic knowledge would produce similar results and whether such knowledge might prove to be a significant factor explaining comprehension performance. For another, we intentionally regulated the length and character of the texts in this study to ensure that there would be no scrolling effect in the digital condition; that the main idea would be rather transparent; and that there were few instances of tangential content in the texts. Our rationale for these conditions was to allow us to replicate findings from our prior investigation (Singer & Alexander, 2017a). In the future, it will be important to loosen these constraints, so that the effects of medium can be explored under conditions when text are substantially longer, main ideas more implicit, and texts include irrelevant or even misleading content.

Even with these acknowledged limitations, we were able to confirm certain outcomes reported in the literature, to clarify others, and to contribute new findings about medium and comprehension. For example, the research pertaining to medium has been rather consistent that today's students generally hold to the belief that they perform better when reading digitally than when they process texts in print (Farah & Maybury, 2009; Housand & Housand, 2012). We also found this to be the case. However, the design of this study allowed us to look more deeply at calibration for individuals who are not only competent at text processing but who have also grown up as natives in a digital world. Nonetheless, in spite of their *digital native* moniker, the undergraduates in this investigation were significantly better calibrated when they read in print than when they read text in digital from, although their level of calibration under either condition was far from optimal. This miscalibration among college students, especially when texts are delivered on screen rather than on paper, remains a cause for concern since it may well translate into diminished strategic and regulatory behaviors.

One of the clarifications that we made to the literature had to do with the potential effects of medium on comprehension performance. In a prior study, we found that the conflicting results regarding medium and comprehension outcomes that populated the literature became clearer when the level of comprehension was considered (Singer & Alexander, 2017a). In effect, what we recognized in that earlier study was that the more questions tapped into detailed or specific content from the reading, the more that reading in print mattered. In the current investigation, the specificity of questions asked again proved informative, although there was some deviation. Specifically, in this study, there was a significant direct effect for the most global, main idea question when processing time was treated as a mediator variable in analysis, although the variance explained was minimal.

In general, by treating processing time as a mediator variable in analysis, we were able to better understand the differential role that this factor plays when students read on paper or digitally. We could appreciate that medium exerts not only a direct effect on competent readers' comprehension but also an indirect effect. There is something about reading digitally that seemingly increases the speed at

which students move through the text and this processing time translates into reduced comprehension. This finding is especially pertinent due to the increasing presence and scope of reading digitally in students' lives (Baron, 2015; Baym, 2015; Dudeney & Hockly, 2016). We regard these outcomes as support for Wickelgren's (1977) speed-accuracy tradeoff hypothesis; that is, when the speed at which a task is performed increases, there may well be a decrement in performance to consider. That tradeoff was certainly evident among our participants. As these competent readers moved from print to a digital medium, the time they spent with the text decreased significantly, and so did their comprehension.

Despite the insights gleaned from our study, there is still much to learn about students' reading comprehension performance, calibration abilities, and speed of processing in print and digital mediums: Take the finding that today's students have the perception that they are doing much better when they are reading digitally than in print when just the opposite is true. It is worth investigating why such miscalibration exists. What is it about reading online that inflates students' sense of performance? Think-aloud studies may serve as a viable route to answer this question.

For another, although this study established that students' generally read digital text more quickly than text in print, future research must explore precisely what is happening as students read offline and online. While processing time was a significant mediator variable in all models tested, faster reading speed alone cannot account for the level of comprehension differences favoring the print in this study. For one, it would be informative to ascertain what students are doing as they move through the print versus digital texts. Perhaps incorporating eye-tracking technology would allow researchers to better see where readers focus their attention during online processing or what facets of the document get limited attention. These areas of high and low focus could then be matched to the content of comprehension questions. Such tracking techniques could likewise be applied to students' reading of traditional print so that more direct comparisons across medium could be undertaken.

Also, it remains to be seen whether students would be able to adjust their performance judgments if they were aware of discrepancies between their judgments of learning and demonstrated performance. In light of the influence that feedback has been shown to have on students' learning and performance (Hattie & Timperley, 2007), there is, at least, the potential that alerting students to the rather paradoxical pattern found between their judgments of performance and their actual performance might prove effective in minimizing miscalibration. Yet, what is intriguing in our data is not simply that students are more poorly calibrated when reading digitally but that these same students were far better at judging their performance when reading comparable texts in print. As of now, we have little insight as to why this is occurring.

Given that print had documented advantages over a digital medium in comprehension and calibration in this study, it might seem logical to suggest that more print reading be encouraged within learning environments. However, because reading digitally is a ubiquitous part of students' lives, it is unreasonable to presume that these students can or should be convinced to shift from digital to print reading for much of the academic reading. Rather, there are at least two viable responses to what has been a consistent advantage of print over digital reading. For one, it would seem worthwhile to consider the conditions under which reading in print rather than digitally would seem advisable. By articulating those conditions, individuals can selectively move to print as the medium for reading when it seems warranted.

In addition, it seems imperative to find ways to enhance students' online reading to approximate certain perceived advantages of print. Especially with the documented effects for processing time, educators need to consider how to get their students to slow their rate of reading online, focus on salient content and features within digital texts, and better gauge their level of understanding? The rich literatures on deep and surface strategies (Dinsmore & Alexander, 2016; Dinsmore, & Zoellner, in press) and on self-regulated learning would seem to be one valuable source of data on improved online reading and self-monitoring (Shi, Frederiksen, & Muis, 2013; Tuysuzoglu & Greene, 2015). Perhaps inserting guiding questions or reminders to reflect on what has been read online would override competent readers' tendency to move rapidly through these digital texts. Such textual modifications might be even more impactful for younger readers who are not as habituated in their approaches to processing texts in print or digitally.

Whatever future studies might shape, it remains imperative to learn much more about how the changing mediums for text delivery are shaping students' literate lives and the understandings that students form as a consequence. It would be unrealistic to suggest that teachers and students return to a world where only reading in print exists. For that reason, it is essential to approach the future of digital reading with more knowledge of the seeming advantages and disadvantages afforded by reading digitally and of evidence-based techniques for maximizing those potential advantages and circumventing possible disadvantages. In that way, educational researchers and practitioners can help students recognize how much processing time they require to move beyond a gist understanding to a richer, deeper, and more critical analysis of text content—whether that content is conveyed in print or digitally. Moreover, educational researchers and practitioners can assist students in achieving more accurate judgments of their learning in print or digitally and guide those students toward instituting whatever strategies are warranted to align their predicted performance with actual performance levels. Unquestionably, there is much more that we need to learn about reading under different mediums, but we regard this investigation as a step in the right direction.

ORCID

Lauren M. Singer Trakhman http://orcid.org/0000-0003-4763-3270

References

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

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. doi:10.1016/j.learninstruc.2012.10.003

Baron, N. S. (2015). *Words onscreen: The fate of reading in a digital world*. New York, NY: Oxford University Press.

Baym, N. K. (2015). *Personal connections in the digital age*. West Sussex, UK: John Wiley & Sons.

Bock, K., & Stauth, C. (2008). *Healing the new childhood epidemics: Autism, ADHD, asthma, and allergies: The groundbreaking program for the 4-A disorders*. New York, NY: Ballantine Books.

Burton, R. F. (2001). Quantifying the effects of chance in multiple choice and true/false tests: Question selection and guessing of answers. *Assessment and Evaluation in Higher Education*, *26*(1), 41–50. doi:10.1080/02602930020022273

Castrells, M. (2011). *The rise and fall of the network society: The information age: Economy, society, and culture* (Vol.1). West Sussex, UK: John Wiley & Sons.

Chen, P., & Zimmerman, B. (2007). A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *Journal of Experimental Education*, *75*(3), 221–244. doi:10.3200/JEXE.75.3.221-244

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. West Sussex, UK: John Wiley & Sons.

D'Haenens, L., Jankowski, N., & Heuvelman, A. (2004). News in online and print newspapers: Differences in reader consumption and recall. *New Media and Society*, *6*(3), 363–382. doi:10.1177/1461444804042520

Dinsmore, D. L., & Alexander, P. A. (2016). A multidimensional investigation of deep-level and surface-level processing. *Journal of Experimental Education*, *84*(2), 213–244. doi:10.1080/00220973.2014.979126

Dinsmore, D. L., & Zoellner, B. P. (in press). The relation between cognitive and metacognitive strategic processing during a science simulation. *British Journal of Educational Psychology*.

Dudeney, G., & Hockly, N. (2016). Literacies, technology and language teaching. In F. Farr & L. Murray (Eds.), *The Routledge Handbook of Language Learning and Technology* (pp. 115–121) . New York, NY: Routledge.

Dyson, M., & Haselgrove, M. (2000). The effects of reading speed and reading patterns on the understanding of text read from screen. *Journal of Research on Reading*, *23*(2), 210–223. doi:10.1111/1467-9817.00115

Farah, C. S., & Maybury, T. (2009). Implementing digital technology to enhance student learning of pathology. *European Journal of Dental Education*, *13*(3), 172–178. doi:10.1111/j.1600-0579.2009.00570.x

Farinosi, M., Lim, C., & Roll, J. (2016). Book or screen, pen or keyboard? A cross-cultural sociological analysis of writing and reading habits basing on Germany, Italy and the UK. *Telematics and Informatics*, *33*(2), 410–421. doi:10.1016/j.tele.2015.09.006

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.

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

Gill, K., Mao, A., Powell, A. M., & Sheidow, T. (2013). Digital reader vs. print media: The role of digital technology in reading accuracy in age-related macular degeneration. *Eye*, *27*(5), 639–643. doi:10.1038/eye.2013.14

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*, 119–136. doi:10.1037/0096-3445.116.2.119

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, *3*(2), 101–121. doi:10.1007/s11409-008-9021-5

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. doi:10.3102/003465430298487

Hayes, A. F. (2013). *Introduction to mediation, moderation, and condition process analysis: A regression-based approach*. New York, NY: Guilford Press.

Housand, B. C., & Housand, A. M. (2012). The role of technology in gifted students' motivation. *Psychology in the Schools*, *49*(7), 706–715. doi:10.1002/pits.21629

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationship 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. doi:10.1037/0096-3445.135.1.36

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. doi:10.1016/j.ijer.2012.12.002

McKnight, K., O'Malley, K., Ruzic, R., Horsley, M. K., Franey, J. J., & Bassett, K. (2016). Teaching in a digital age: How educators use technology to improve student learning. *Journal of Research on Technology and Education*, *48*(3), 1–18. doi:10.1080/15391523.2016.1175856

Mohammadyari, S., & Singh, H. (2015). Understanding the effect of e-learning on individual performance: The role of digital literacy. *Computers and Education*, *82*, 11–25. doi:10.1016/j.compedu.2014.10.025

Myrberg, C., & Wiberg, N. (2015). Screen vs. paper: What is the difference for reading and learning?. *Insights*, *28*(2), 49–54.

National Center for Education Statistics. (2013). *The Nation's Report Card: A first look: 2013 mathematics and reading (NCES 2014–451). National Center for Education Statistics, Institute of Education Sciences*, Washington, D.C: U.S Department of Education.

Noyes, J. M., & Garland, K. J. (2003). VDT versus paper-based text: Reply to Mayes, Sims, and Koonce. *International Journal of Industrial Ergonomics*, *31*(6), 411–423. doi:10.1016/S0169-8141(03)00027-1

Pajares, F., & Miller, M. D. (1997). Mathematics self-efficacy and mathematical problem solving: Implications of using different forms of assessment. *Journal of Experimental Education*, *65*(3), 213–228. doi:10.1080/00220973.1997.9943455

Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices—past, present and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13–69). Mahwah, NJ: Erlbaum.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. doi:10.3758/BRM.40.3.879

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, *9*(5), 1–6. doi:10.1108/10748120110424816

Purcell, K., Heaps, A., Buchanan, J., & Friedrich, L. (2013). *How teachers are using technology at home and in their classrooms*. Washington, DC: Pew Research Center's Internet and American Life Project.

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

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

Shi, Y., Frederiksen, C. H., & Muis, K. R. (2013). A cross-cultural study of self-regulated learning in a computer-supported collaborative learning environment. *Learning and Instruction*, *23*, 52–59. doi:10.1016/j.learninstruc.2012.05.007

Tuysuzoglu, B. B., & Greene, J. A. (2015). An investigation of the role of contingent metacognitive behavior in self-regulated learning. *Metacognition and Learning*, *10*(1), 77–98. doi:10.1007/s11409-014-9126-y

Van de Vijver, F. J., & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *Journal of Applied Psychology*, *79*(6), 852–859. doi:10.1037/0021-9010.79.6.852

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

Wästlund, E., Reinikka, H., Norlander, T., & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. *Computers in Human Behavior*, *21*(2), 377–394. doi:10.1016/j.chb.2004.02.007

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