

Reference:

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

This manuscript may not exactly replicate the final version published in the journal.  
It is not the copy of record.

**Generalizing Screen Inferiority - Does the Medium, Screen versus Paper,  
Affect Performance Even with Brief Tasks?**

Yael Sidi\*, Yael Ophir, and Rakefet Ackerman

Faculty of Industrial Engineering and Management,

Technion–Israel Institute of Technology, Haifa, Israel

\* yaelsidi@tx.technion.ac.il

**Acknowledgments**

The study was supported by grants from the Israel Science Foundation (Grant No. 957/13).

## **Abstract**

Screen inferiority in performance and metacognitive processes has been repeatedly found with text learning. Common explanations for screen inferiority relate to technological and physiological disadvantages associated with extensive reading on screen. However, recent studies point to lesser recruitment of mental effort on screen than on paper. Learning tasks involving a heavy reading burden confound technological and physiological media differences with potential media effects on recruitment of mental effort. The present study focused on media effects on effort recruitment. We examined whether screen inferiority remains even with a brief task that nevertheless requires effort recruitment. In two experiments, participants faced three short math problems that require systematic processing to solve correctly. We examined media effect on solving these problems, and the potential of disturbed perceptual fluency (i.e., disfluent versus fluent fonts) to induce effort investment. Overall, there were no performance differences between the media. However, when collecting confidence ratings, disfluency improved performance on screen and hindered it on paper. Only on paper confidence ratings were sensitive to performance differences associated with fluency, and resolution was better with the disfluent font than with the fluent font. Correspondingly, another sample reported on their preference of media for solving the problems. They expressed a clear reluctance to working on screen despite the task being brief. This preference is suggestive of reliable meta-metacognitive judgments reflecting the general lower quality of metacognitive processes on screen. The findings call for considering medium and presentation format effects on metacognitive processing when designing computerized environments, even for brief tasks.

Computerized environments are often used in place of paper-based environments for training, learning, and assessment. However, this development has raised concerns about the potential effects of screen-based environments on learning and other cognitive tasks. Studies examining users' attitudes generally find a preference to work on paper, indicating a subjective difference between the media (e.g., Annand, 2008; Mizrachi, 2015; Woody, Daniel, & Baker, 2010). With respect to the actual performance of cognitive tasks, the evidence, though mixed, points toward screen inferiority. While some studies have found equivalence between the media (e.g., Ball & Hourcade, 2011; Margolin, Driscoll, Toland, & Kegler, 2013; Murray & Pérez, 2011; Salmerón & García, 2012), many others report inferior results on screen. Consistently, studies involving learning from continuous texts, a task that can be performed the same way in both media, have found screen inferiority in performance (e.g., Ackerman & Goldsmith, 2011; Ben-Yehudah & Eshet-Alkalai, 2014; Daniel & Woody, 2013; Mangen, Walgermo, & Brønck, 2013). Moreover, screen inferiority has been found even in tasks involving capabilities unique to computerized environments and considered as advantageous for this environment, like hypertext (see DeStefano & LeFevre, 2007, for a review), sound, animation, and interactive reading (e.g., Chiong, Ree, Takeuchi, & Erickson, 2012; Mayer, Heiser, & Lonn, 2001). Notably, the majority of studies that found screen inferiority have used reading comprehension tasks involving a substantial reading burden. The present study extends this investigation by employing a brief problem solving task, to examine whether screen inferiority remains even when the reading burden is minimized.

Many studies have attributed screen inferiority in text learning to technological disadvantages (e.g., less-convenient browsing and navigation) or to physical discomfort

(e.g., eye strain) (e.g., Benedetto, Draai-Zerbib, Pedrotti, Tissier, & Baccino, 2013; Li, Chen, & Yang, 2013; see Leeson, 2006, for a review). However, screen inferiority persists even with modern e-books, which are presumed to overcome these disadvantages (e.g., Antón, Camarero, & Rodríguez, 2013; Daniel & Woody, 2013; see Gu, Wu, & Su, 2015, for a review).

Another possible explanation for screen inferiority, and one which has been gaining support in recent years, is the effect of the medium on depth of processing. In other words, this explanation offers that working in computerized environments is associated with shallower cognitive processing, leading to inferior cognitive performance. Indeed, people often report on engaging in sustained reading on paper, while on screen they engage more in multi-tasking and discontinuous reading (Daniel & Woody, 2013; Hillesund, 2010; Liu, 2005). Moreover, the mere presence of an e-book nearby the learners has been found to hinder recall of studied information (Morineau, Blanche, Tobin, & Guéguen, 2005). The researchers suggested that electronic devices provide a contextual cue that leads people to shallower processing, resulting in inferior cognitive performance.

The link between depth of processing and inferior learning on screen has also been discussed in analyses inspired by the metacognitive approach. These studies provide growing evidence which associates screen-related contextual cues with inferior metacognitive processes, namely, less reliable judgments of the expected chance for success and less effective regulation of effort (Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012; Lauterman & Ackerman, 2014). For instance, Ackerman and Goldsmith (2011) addressed medium influences on meta-comprehension processes and found more

pronounced overconfidence among screen learners compared to paper learners. Similarly, Ackerman and Lauterman (2012) found consistent overconfidence for screen learners under time pressure, while paper learners showed better calibration (but also see Norman & Furnes, 2016, where no consistent media effects on metacognitive measures were found, possibly accounted for by several important methodological differences between the studies). As subjective confidence directs study regulation and decisions (e.g. effort investment and stopping rules), overconfidence is undesirable (Dunlosky & Thiede, 1998; Greene & Azevedo, 2007; Winne, 2004).

Further support for depth of processing as a contributing factor to screen inferiority can be derived from studies that attempted to reduce and even eliminate it, by guiding participants to recruit more intensive mental effort to the task than they would engage spontaneously. In particular, recent studies have demonstrated elimination of screen inferiority by activities that encourage in-depth processing. For example, asking participants to identify errors, to improve the quality of a text, or to write keywords that summarize the text's contents, or letting participants gain experience with the test demands (Eden & Eshet-Alkalai, 2013; Lauterman & Ackerman, 2014). These studies suggest that while on paper in-depth text processing is the default, on screen an external trigger is needed.

The aforementioned studies which found screen inferiority in cognitive and metacognitive processes involved reading texts with approximately 600-1200 words. Notably, some tasks performed on screen indeed involve reading lengthy texts, such as reading from an e-book or an online version of an article. However, other typical daily computerized interactions with e-mails, forums, and social networks tend to involve much

brief reading. In the present study, we suggest that using lengthy texts for studying cognitive performance in computerized environments confounds the reading processes with other involved cognitive and metacognitive processes. Specifically, reading per se might be more susceptible to the technological disadvantages and physical discomforts associated with working on screen (see examples above) than memory, inference, monitoring, and effort regulation. In order to disentangle this confound, we examined whether challenging tasks that require recruitment of mental effort, yet involve a minimal reading burden, also show screen inferiority in performance and/or metacognitive processes. Based on the studies which pointed to shallower processing on screen, the hypothesis that guided the present study was that the minimal reading burden does not eliminate screen inferiority.

Following the methods mentioned above, which allowed overcoming screen inferiority with text learning (Eden & Eshet-Alkalai, 2013; Lauterman & Ackerman, 2014), we aimed to increase recruitment of mental effort in a brief task as well. Time pressure was found with the same population to be effective in this respect with text learning (Ackerman & Lauterman, 2012). However, this method was effective on paper but not on screen, while our goal was to increase effort on screen, where, as described above, the default mode of processing is shallower than on paper. Another potential method which is known to be effective in learning tasks is introducing ‘desirable difficulties’. For example, Sungkhasettee, Friedman, and Castel (2011) presented words for memorization either upside down or straight. Recall was better in the more challenging upside down condition. Such manipulations have been suggested to improve learning by triggering deeper processing of the learning contents (Bjork, 1994, 1999; see Kühl & Eitel,

this issue, for a review of the disfluency theory). In the context of problem solving as well, several studies have found that people engage more in the task when perceptual fluency, manipulated by font readability, is disturbed (e.g., *fonts* vs. **font**; Thompson et al., 2013).

In one such study performed on paper (Alter, Oppenheimer, Epley, & Eyre, 2007) improved performance with disturbed fluency was found. Although other studies did not find such performance advantage either on screen or on paper (Meyer et al., 2015; Thompson et al., 2013), in text learning, in-depth processing was associated with improved test scores and improved reliability of metacognitive monitoring (Lauterman & Ackerman, 2014; Thiede, Anderson, & Therriault, 2003). It is possible then, that even if the disfluent font does not improve performance on screen, monitoring improvement would be found nevertheless. Thus, the present study examined disfluency as a metacognitive cue for recruitment of mental effort, expecting more improvement on screen than on paper.

### The present study

To examine our hypotheses, we studied the effects of the medium, screen versus paper, and perceptual fluency on performing a brief problem-solving task, the Cognitive Reflection Test (CRT; Frederick, 2005), which introduces cognitive and metacognitive challenges, but involves a minimal reading burden. The CRT consists of three misleading math problems (bat & ball, widgets, lily pads; up to 45 words in each). These problems are designed so that the first solution that commonly comes to mind is a wrong but predictable one. For example, the first question is: "A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? \_\_\_\_\_ cents". The intuitive answer, "10 cents", is wrong. Deeper processing is required to recognize this error and

come up with the correct answer ("5 cents") (see Frederick, 2005, for the complete question set). These problems are widely used in studies related to dual-process theory (e.g., Alter et al., 2007; Cokely & Kelley, 2009; Thompson et al., 2013). Within this theory, the misleading power of these problems is explained by the dominance of Type 1 processes, which are mostly automatic (see Evans & Stanovich, 2013, for a review). Most people can overcome this misleading intuition by recruiting the more effortful and analytic Type 2 processing (Frederick, 2005). The activation of Type 2 processes depends on the reliability of the Feeling of Rightness (FOR). FOR is a metacognitive judgment which refers to the assessed chance that the initial solution that comes to mind is correct (Thompson, 2009). When FOR is high, people tend to provide their first solution. When it is lower, they tend to reconsider their initial solution candidate and change it (Thompson et al., 2013). Thus, activation of Type 2 processes when needed is a metacognitive regulatory process which accompanies the cognitive process of solving the problem per se.

The features of the CRT make it suitable for the present study as it can be performed in much the same way on screen and on paper, and it requires recruitment of mental effort while involving only brief reading. In addition, the task itself is also brief when compared to commonly studied problem solving tasks. For instance, Ackerman (2014) used a problem-solving task which takes about half an hour. Other problem solving procedures may take even 45 minutes (e.g., King, 1991). The CRT, in contrast, involves only three problems that take just a couple of minutes to solve. Thus, the task is brief both in the number of problems and in its reading burden.

We started our study by examining media preference regarding solving the CRT by a survey. This allowed us to examine the correspondence of perceived differences to actual

differences between the media in performance and the quality of metacognitive processes. Previous research indicated a relatively weak association. For example, Holzinger et al. (2011) found that in a sample of medical professionals, 90% preferred reading medical reports on paper rather than on screen, even though reading comprehension tests showed no differences in performance. Ackerman and Lauterman (2012) found attenuated paper preference (62%) among engineering undergraduates, while the rest of the sample expected no performance difference between the media. The test outcomes were equivalent for both media when free learning time was allowed, yet inferior on screen when limiting the learning time. On the other hand, at the same study, examining the data in division by participants' preference showed some validity in it: Those who studied from their preferred medium outperformed those who studied from their less preferred medium. The population in the present study consisted of engineering students as well, and the task's focus was on reasoning rather than on interacting with the media. Thus, we hypothesized a moderate paper preference for performing the CRT (*H1*), similar to that found in Ackerman and Lauterman (2012).

As for recruitment of mental effort, if working on screen cues participants to recruit less effort in the task than on paper, regardless of the reading burden involved, then screen participants are expected to rely more on Type 1 processes and to achieve inferior performance compared with paper participants. Since the reliability of metacognitive monitoring depends on recruitment of mental effort (Lauterman & Ackerman, 2014; Thiede et al., 2003), it was also expected to be inferior on screen. Thus, we predicted that despite the task being brief, screen inferiority would emerge with the CRT task and manifest both in lower performance (*H2*) and less reliable monitoring (*H3*) on screen.

A recent study from another domain supports our prediction of screen inferiority in brief tasks. Oeberst, Haberstroh, and Gnambs (2015) examined effects of the medium on risk taking. Participants viewed outcomes of two lotteries, either in a computerized environment or using the traditional format of drawing balls from a closed box. They then selected between a risky or less risky lottery. Despite viewing the same outcomes, the computerized group made more risky choices than the traditional lottery group. The authors proposed that the computerized group underestimated the probability of an unfavorable outcome, and therefore perceived the lottery as less risky than it actually was. In metacognitive terms, these results indicate greater overconfidence in the computerized environment.

As previously indicated, we also sought to examine whether screen inferiority would diminish if more effort investment was encouraged. Specifically, under the assumption that *H2* and *H3* are supported (screen inferiority would be evident with the CRT task), we aimed to examine the influence of recruitment of extra mental effort on cognitive performance and metacognitive processes. Based on the effects of in-depth processing on text learning, as described above, we hypothesized that recruitment of extra mental effort in response to disturbed fluency would result in improved performance for the screen group (*H4*) and enhance the reliability of their metacognitive judgments (*H5*).

Notably, as the CRT problems are considered quite challenging, it has been asserted that with this task only people with high cognitive ability would benefit from recruitment of extra mental effort (see Meyer et al., 2015, for a review). In line with this assertion, we sampled undergraduates from programs that require high SAT scores (top 20%).

To summarize, we started our investigation with a survey designed to examine the target population's media preference for the brief CRT task (*H1*—A moderate paper preference with the brief task). As rating confidence might influence task performance (e.g. Yue, Castel & Bjork, 2013), in Experiment 1 we examined performance in the CRT problems on screen and on paper with the perceptual fluency manipulation. Experiment 2 was a replication of Experiment 1 to which we added confidence ratings. Thus, in Experiment 1 we examined the hypotheses related to the effects of media and fluency on performance (*H2*—screen inferiority and *H4*—disfluency advantage) and in Experiment 2 we also examined the hypotheses related to the effects of these factors on confidence reliability (*H3*—screen inferiority and *H5*—disfluency advantage).

### **Medium preference survey**

As mentioned above, it is a common finding that people prefer reading on paper to reading from a screen. The purpose of the present survey was to examine whether a preference for performing tasks on paper is moderated with a brief task (*H1*). We presented undergraduates with the three CRT problems and asked which medium they would prefer for solving them.

### Method

#### *Participants*

Forty-three Technion undergraduates (49% females) volunteered to fill in the questionnaire.

#### *Materials*

The three CRT problems used by Frederick (2005) were translated into Hebrew. The questionnaire was printed on one page. Respondents were first asked to provide a few demographic details. The three CRT problems appeared below, in the same order as in Frederick (2005), with a statement making clear that respondents were not being asked to actually solve the problems. This was followed by the four critical survey questions: a) If you were asked to solve these problems, on which medium would you prefer them to be presented? (computer, paper, no difference) b) If you were asked to solve these problems, would you be more likely to succeed if they were presented on a computer screen, paper, no difference? c) If you were given the problems on the computer, would you print them so they would be in a form you find more convenient? (yes/no) d) If you were given the problems on paper, would you scan them so they would be in a form you find more convenient? (yes/no). The order of the medium options in questions a and b was counterbalanced across participants, and the order of questions c and d was counterbalanced across participants.

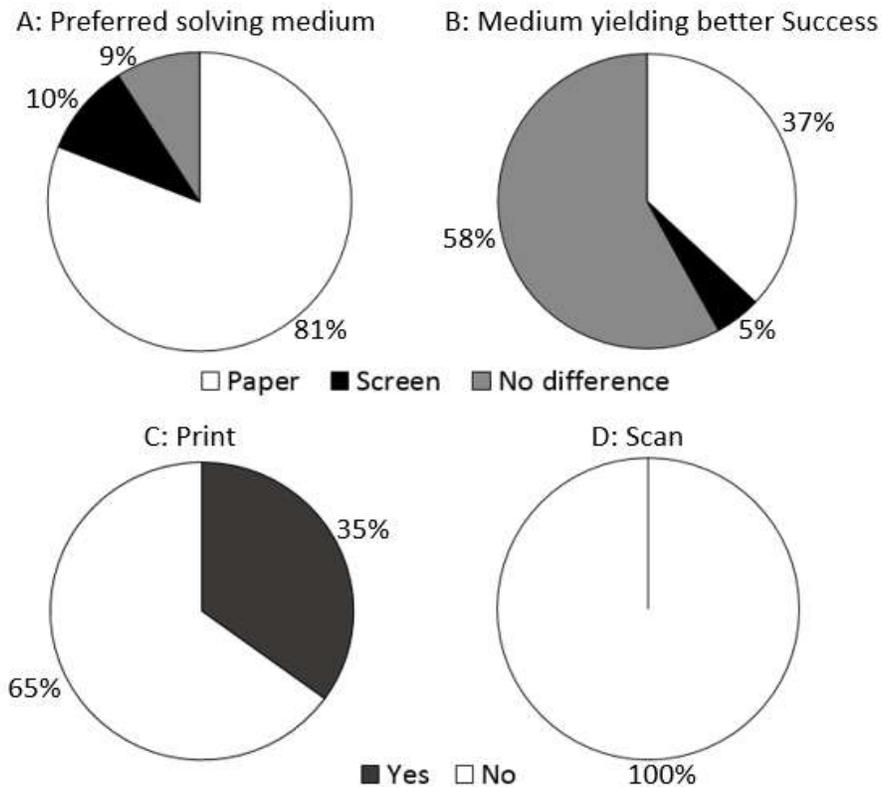
### *Procedure*

Participants filled in the questionnaire in the lab before or after (randomly assigned) participating in other, unrelated, experiments.

### Results and discussion

Contradictory to our hypothesis, despite being technologically proficient, a majority of the participants reported that they would prefer solving these brief problems on paper (Fig. 1, chart A),  $p < .0001$  by a binomial test comparing screen and paper preference. Indeed, somewhat more than a third of the respondents said they would print the problems if they

received them electronically (Fig. 1, chart C),  $p < .01$ , but none said they would scan a paper copy (Fig. 1, chart D),  $p < .0001$ . Nevertheless, most respondents also expected the medium to make no difference in their success (Fig. 1, chart B),  $p < .05$  comparing those who expected no difference with those who expected media advantage.



**Fig. 1** Distribution of answers to the survey questions: A. Choice of preferred medium for solving the CRT problems. B. Medium expected to yield better success in the task. C. Choice to print the problems if received on screen. D. Choice to scan the problems if received on paper.

To summarize, despite the brief task and the technology-oriented population, over the four survey questions, the results provide a clear picture of a paper preference. Interestingly, however, although the participants expected the solving process to be more convenient on paper, most did not foresee a difference in their solving success. It seems

that many participants anticipated that they would be able to overcome this inconvenience.

## **Experiment 1**

In Experiment 1, we examined effects of the medium, screen versus paper, on performance in the CRT task. Our aim was to examine whether screen inferiority in performance is evident with the brief task (*H2*) and, if so, whether disturbed fluency leads to performance improvement on screen (*H4*). Specifically, we examined whether, with disturbed fluency, participants would provide the expected misled answers less often as a result of a metacognitive regulatory mechanism which hints at activation of Type 2 processes (see Alter et al., 2007). To accomplish this we employed a two-by-two between participants design with the factors Medium (Screen vs. Paper) and Perceptual Fluency (disfluent vs. fluent).

### Method

#### *Participants*

Two hundred and four Technion undergraduates (46% females;  $M_{\text{age}} = 24.3$ ,  $SD = 2.1$ ) volunteered to participate in the experiment<sup>1</sup>. Their mean self-reported SAT score was 680.2 ( $SD = 41.6$ )<sup>2</sup>. They were randomly assigned to screen or paper, and to disfluent or fluent font, with 45-59 participants in each group. Thirty-one participants (15%) reported having learning disabilities, but they were spread similarly among the four groups,  $\chi^2 < 1$ .

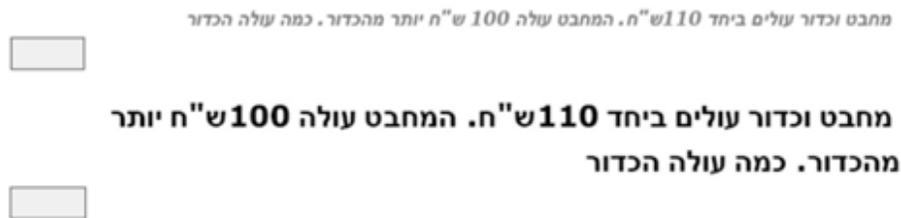
---

<sup>1</sup> The original planned sample size was of about 100 participants. After running this sample and finding no effects (see below), we doubled the sample in order to verify that these results did not stem from effects that were weaker than expected.

<sup>2</sup> In the Israeli version of the SAT, known as the Psychometric Entrance Test, scores range from 200 to 800, normally distributed ( $M = 533$ ,  $SD = 101$ , in 2013).

## Materials

The disfluent and fluent font types were chosen based on a pretest (see Thompson et al., 2013). In the pretest ( $N = 20$ ), several font options were rated on a Likert scale ranging from 1 (“Illegible”) to 5 (“Easier than the regular font”). Most participants (70%) rated the chosen disfluent font (Arial 9-point italicized light grey) as “Legible with effort” (2 on the scale), and none characterized it as easy to read (4) or easier than the regular font (5) ( $M = 2.1$ ,  $SD = 0.55$ ). The fluent font (Arial 18-point black) was judged as easy to read (i.e., rated 4 or higher) by all participants (100%) ( $M = 4.3$ ,  $SD = 0.30$ ). Ratings of both font types deviated significantly from the mid-scale (3), both  $ps < .0001$ . See Figure 2.



**Fig. 2** Hebrew versions of the bat and ball problem presented in disfluent (top) and fluent (bottom) fonts.

A computerized questionnaire presented the three CRT problems on one page, with an empty space for answer entry next to each question (see Figure 2). A second page was used to collect personal details. The printed version was a printout of the computerized questionnaire. The disfluent version was identical to the fluent version, except for the font of the CRT problems.

### *Procedure*

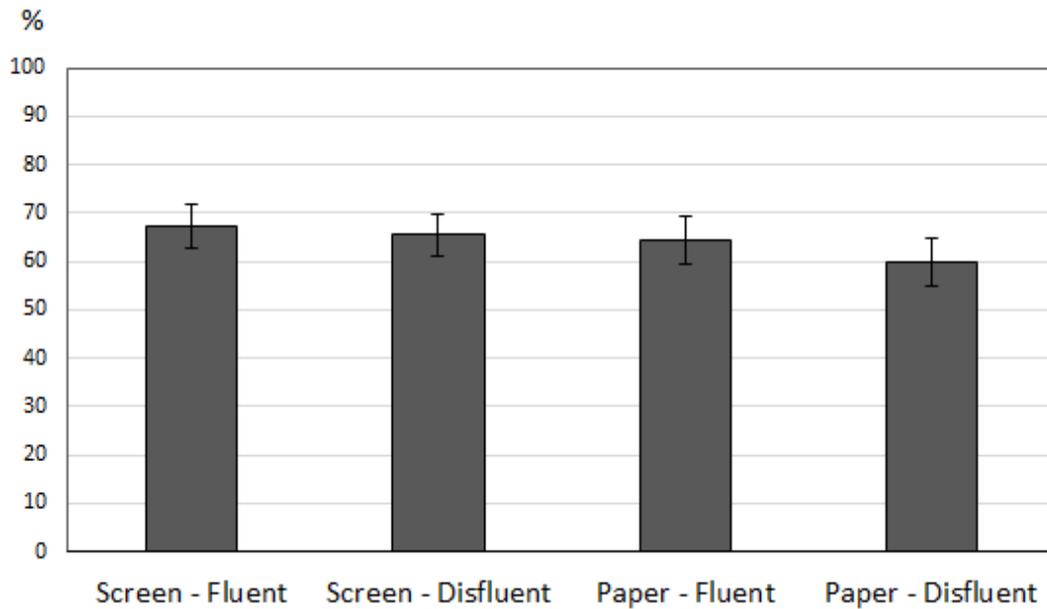
The participants were randomly assigned to perform the task before or after participating in other experiments. Participants were explicitly instructed to refrain from writing, drawing or figuring while solving the problems. The task was self-paced.

### Results and discussion

No difference was found between the groups in SAT scores or age ( $ts < 1$ ). The mean success rate was 64% ( $SD = 33.4$ ). Most importantly, in an Analysis of Variance (ANOVA) of Medium (Screen vs. Paper)  $\times$  Perceptual Fluency (disfluent vs. fluent) on success rates, no main effects and no interactive effect were found, all  $F$ s  $< 1$  (see Fig. 3). A similar analysis for the expected misled answers (i.e., the intuitive and predictable wrong solutions) yielded similar results. Thus, the hypothesized performance differences between screen and paper were not evident, nor did perceptual fluency induce a difference between them<sup>3</sup>.

---

<sup>3</sup> All analyses were also separately conducted for each of the CRT problems, with similar results (no main effects or interactive effect).



**Fig. 3** Success rates for each experimental group in Experiment 1. Error bars represent standard errors of the means.

In conclusion, in contrast to the clear medium preference expressed in the survey, no differences were found between the media. This is important, as it suggests that despite a decisive preference for paper, most members of the studied population perform equally well regardless of the medium. Additionally, we did not find that perceptual fluency affected success rates and the number of expected misled answers.

## **Experiment 2**

In Experiment 2, we examined whether working on this brief task on screen results in less reliable confidence ratings (*H3*), as was consistently the case with learning tasks (Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012; Lauterman & Ackerman, 2014). To examine this, we replicated Experiment 1, with the same experimental design, except that here we collected a confidence rating for each solution. These confidence

ratings reflected participants' subjective assessment of the likelihood that their solution was correct.

The reliability of confidence ratings is commonly measured using two indicators, calibration and resolution. *Calibration* relates to absolute accuracy: mean confidence ratings are compared with actual success rates on the task per subject, providing a measure of overall overconfidence. Existing research indicates that problem solvers are often overconfident (Ackerman & Zalmanov, 2012; Prowse Turner & Thompson, 2009; Shynkaruk & Thompson, 2006). If screen inferiority in the reliability of metacognitive judgments does not depend on text length, overconfidence is expected to be more pronounced on screen than on paper with the brief task as well (*H3*). However, previous work with text learning has shown that overconfidence was eliminated by manipulations that encouraged more in-depth processing. Thus, we expected that with disturbed fluency, confidence ratings on screen would correspond better to actual performance (*H5*).

*Resolution* is another aspect of judgment reliability, distinct and independent of calibration (Ackerman & Goldsmith, 2011). While calibration reflects absolute accuracy, resolution relates to relative accuracy as it measures discrimination between correct and wrong responses (Nelson, 1984). Resolution is usually calculated by within-participant gamma correlation between confidence and success in each item (e.g., Koriat, Ma'ayan, & Nussinson, 2006, Experiment 7). The reliability of this statistical method increases as more items are used (see explanation and critique in Masson & Rotello, 2009). In the present study, there were only three items per participant. Thus, we examined resolution somewhat differently, but followed the principles of gamma correlation, as described below.

Studies with text learning have found improved resolution when more in-depth processing of the text was required (e.g., Thiede et al., 2003). Thus, similarly to the predictions above regarding calibration, weaker resolution is expected (*H3*), and disturbed fluency should support resolution improvement (*H5*) on screen.

## Method

### *Participants*

Participants were one hundred and seventeen Technion undergraduates (43% females;  $M_{\text{age}} = 24.7$ ,  $SD = 2.6$ ;  $M_{\text{SAT}} = 682.9$ ,  $SD = 37.5$ ; 13% reported learning disabilities). They were randomly assigned to screen or paper, and to disfluent or fluent fonts, with 28-31 participants in each group.

### *Materials and procedure*

The questionnaires and procedure were highly similar to those used in Experiment 1. The only difference was that each question was followed by an eleven-point scale representing 0, 10, 20 ...100% confidence. Participants rated their confidence after providing their response to each question.

## Results and discussion

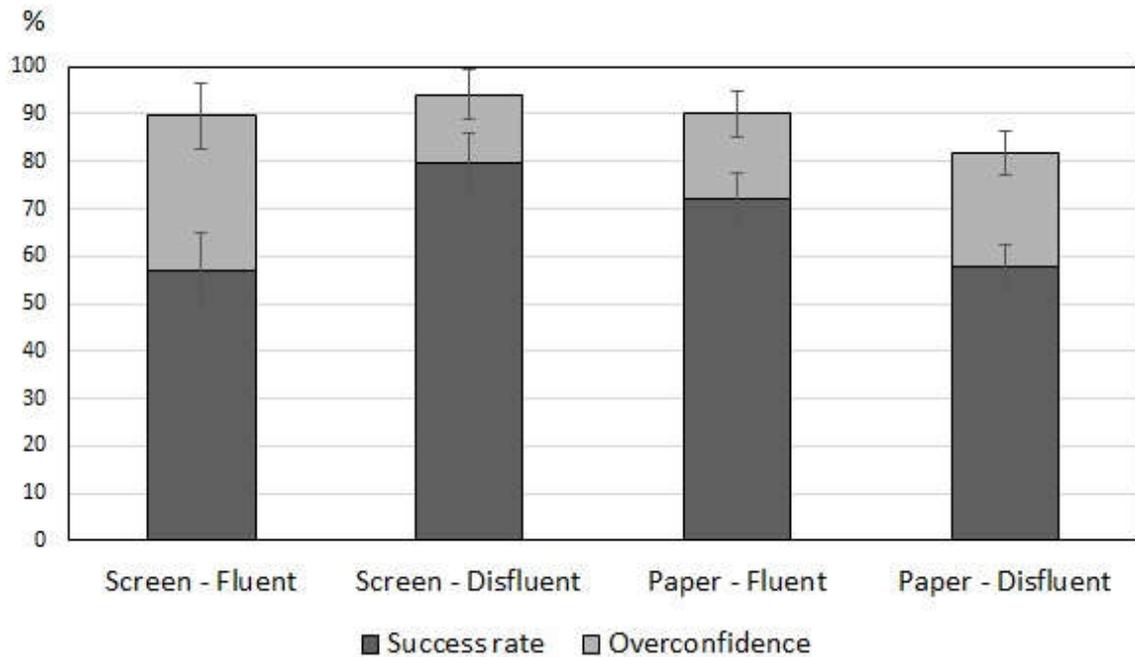
No differences were found between the groups in SAT scores or age (both  $ps > .05$ ).

### *Success rates and expected misled answers*

The mean success rate was 66.6% ( $SD = 33.87$ ). Two-way ANOVA examining the effects of Medium  $\times$  Perceptual Fluency revealed no main effects (both  $F_s < 1$ ). However, there was an interactive effect,  $F(1, 113) = 9.08$ ,  $MSE = 1084.30$ ,  $p = .003$ ,  $\eta_p^2 = .074$ . The

interaction stemmed from opposite effects of perceptual fluency on success rates in the two media. While on screen, as predicted, the disfluent font resulted in better success rates than the fluent font,  $t(54) = 2.22, p = .030$ , on paper the effect was reversed,  $t(59) = 2.01, p = .049$ . See Figure 4 for the results.

Expected misled answers comprised 73% of the total errors. The results for expected misled answers were similar to the overall results. While there were no main effects ( $F < 1$ ), there was a significant interaction between the medium and perceptual fluency on the number of expected misled answers produced,  $F(1, 113) = 3.97, MSE = 908.16, p = .049, \eta_p^2 = .034$ . Screen participants had a lower number of expected misled answers with the disfluent font, while for the paper participants the pattern was reversed. However, these simple effects were not statistically significant,  $t(54) = 1.74, p = .087$  and  $t < 1$ , respectively. Notably, this interactive effect on performance was not found in Experiment 1. The implications of this finding are addressed in the General Discussion.



**Fig. 4** Success rate and overconfidence in Experiment 2. Error bars represent standard errors of the mean for the bar below them.

### *Confidence*

Confidence ratings are represented in Fig. 4 by the top of the overconfidence bars. A similar ANOVA on confidence revealed a main effect of the Medium,  $F(1, 113) = 6.30$ ,  $MSE = 169.29$ ,  $p = .013$ ,  $\eta_p^2 = .053$ , with higher confidence on screen than on paper. There was no main effect for perceptual fluency,  $F < 1$ , but there was an interactive effect,  $F(1, 113) = 7.07$ ,  $MSE = 169.29$ ,  $p = .009$ ,  $\eta_p^2 = .059$ . On screen, despite the positive effect of the disfluent font on performance, confidence was equivalent for both font types,  $t(54) = 1.41$ ,  $p = .16$ . On paper, in contrast, there was a significant difference between the font

types, in correspondence with the performance differences, such that confidence was lower with the disfluent font than with the fluent font,  $t(59) = 2.33, p = .023^4$ .

### *Calibration—Overconfidence*

Calibration was calculated using a within-participant comparison between task performance and subjective confidence judgments. The above-reported performance and confidence differences resulted in no main effects of the medium or perceptual fluency, both  $F_s \leq 1$ , on overconfidence, but a significant interactive effect,  $F(1, 113) = 4.82, \text{MSE} = 867.51, p = .030, \eta_p^2 = .041$ . Supporting our hypothesis, overconfidence on screen was lower for the disfluent than for the fluent font,  $t(54) = 2.08, p = .042$ , while on paper, overconfidence did not differ between the font types,  $t < 1$ .

One might suggest that overconfidence for the disfluent font on screen could not be as high as for the fluent font because of a ceiling effect in confidence (see Fig. 4). However, all means of confidence ratings were significantly lower than 100%, all  $p_s \leq .002$ . Thus, there was room for even higher confidence ratings. Moreover, if participants were sensitive to their performance, they could have produced lower confidence ratings for the fluent fonts, as found with the disfluent fonts on paper. However, this was not the case.

### *Resolution*

To calculate resolution, we accumulated for each participant cases of fit between confidence and success. That is, each case in which confidence was higher for a correct response than for a wrong response (e.g., 80% confidence for a correct response to one

---

<sup>4</sup> A separate analysis per each of the CRT problems revealed that the interactive effects found for success rates were mainly due to widgets and lily pads problems, and the interactive effect found for confidence was mainly due to the widget problem.

problem and 70% confidence for a wrong response to another problem), was marked as fit. If confidence ratings were the other way around, then the case was marked as nonfit. Resolution of each participant's confidence ratings was defined to be the difference between fit and nonfit, and ranged between -2 and 2. Cases of no variability in performance or confidence are undefined by gamma correlation as well as in this procedure. This procedure resulted with meaningful resolution results for 108 participants (92%). Importantly, there was no significant correlation between resolution and calibration in our study, supporting the distinct contribution of each measure ( $r = -.13, p = .248$ )

ANOVA as above on resolution yielded no main effects for the Medium,  $F(1, 83) = 2.69, MSE = 0.84, p = .105, \eta_p^2 = .031$ , or for perceptual fluency,  $F(1, 83) = 1.81, MSE = 0.84, p = .175, \eta_p^2 = .022$ . There was, however, an interactive effect,  $F(1, 83) = 4.78, MSE = 0.84, p = .032, \eta_p^2 = .054$ . Analysis of simple effects revealed no significant difference between the two font types on screen ( $M_{disfluent} = 0.31, SD = 0.75; M_{fluent} = 0.47, SD = 0.81$ ),  $t(32) = 0.60, p = .55$ . On paper, in contrast, resolution was better for the disfluent font ( $M = 1.09, SD = 1.04$ ) than for the fluent font ( $M = 0.36, SD = 0.90$ ),  $t(51) = 2.66, p = .010$ . In order to compare other combinations of groups, we also conducted a one-way ANOVA with the four groups in one factor and a Tukey post-hoc test for paired comparisons. Beyond the comparisons reported above, the paper group with the fluent font did not differ from both screen groups (both  $p$ 's  $> .954$ ). However, resolution for the paper group with the disfluent font was marginally better than both screen groups, disfluent and fluent, ( $p = .052, p = .086$ , respectively). This latter finding helps appreciating the extent of metacognitive benefit gained on paper with the disfluent font.

In sum, as in Experiment 1, overall success rates did not differ between the media. However, the media did differ in performance sensitivity to perceptual fluency. Of more importance for the purpose of Experiment 2 are the findings regarding metacognitive processes. Confidence sensitivity, as demonstrated by an adjustment of confidence ratings to performance differences between the fonts, was weaker on screen. Moreover, even though calibration improved for the screen group with the disfluent font, as predicted, resolution was not affected by the fluency manipulation. On paper, in contrast, confidence ratings were in line with the performance difference between the fonts, and resolution improved when the font was disfluent. Thus, on screen, there was less sensitivity to the performance differences associated with the perceptual fluency manipulation.

### **General discussion**

In this study, we examined medium effects on performance and metacognitive processes. Unlike previous studies which addressed this issue, we used a brief task imposing a cognitive challenge, with only a minimal reading burden. Within each medium, screen versus paper, we examined the sensitivity of these processes to perceptual fluency by presenting the problems in a fluent or a disfluent manner.

We hereby summarize the findings and the questions arising from them. We start with a discussion of effects of medium and perceptual fluency on performance in solving brief problems, followed by a discussion of their effects on the metacognitive processes involved. Next, we consider what can be learned from subjective preferences for the media on which the task is performed, and conclude with the implications of the study for computerized learning environments.

## Media and perceptual fluency effects on performance

In the present study, we examined whether screen inferiority in performance would be evident even with a brief task (*H2*). Contrary to our hypothesis, the problem solving task that we used did not generate a difference in performance between screen and paper. One possible explanation for this finding is that the main cause of the previously found performance inferiority on screen is technology-related barriers associated with extensive reading. Thus, when the reading load is reduced, performance differences can be eliminated. However, some studies found performance equivalence even with longer texts under certain conditions. For instance, Ackerman and Lauterman (2012, Experiment 1) demonstrated with text learning that only when the task was performed under time pressure, screen inferiority in performance emerged. The authors suggested that while the learning processes per se may be equivalent under the two media, metacognitive regulatory processes are inferior on screen, and that this inferiority emerges with the challenge of study regulation under time pressure. We call future studies to examine whether similar constraint conditions, which require effective regulation processes, reveal performance differences in brief tasks as well.

In both experiments we attempted to encourage recruitment of extra mental effort by disturbing perceptual fluency (*H4*). Due to the nature of our design, we were limited in our ability to directly examine the experience of this disruption during solving the brief problems. However, in our pretest for choosing the disfluent and fluent fonts, most participants reported the disfluent font to be legible with effort. Previous studies have shown that making fonts harder to read (e.g., by shrinking them, blurring their edges, or using italics) indeed influences the fluency of reading (e.g. Oppenheimer, 2006; Song &

Schwarz, 2008). Moreover, disfluency has been found to elicit more effort investment in various memorization tasks (Diemand-Yauman, Oppenheimer, & Vaughan, 2011; Hirshman & Mulligan, 1991). However, findings from such fluency manipulations with CRT problems have been inconsistent (Alter et al., 2007; Meyer et al., 2015; Thompson et al., 2013).

We found opposing effects for fluency on screen and paper only in Experiment 2. The screen group displayed better success with the disfluent font, in support of *H4*, while the paper group demonstrated poorer performance under this condition, which we did not expect. While usually disturbed fluency is found to either improve or have no effect on performance, a recent research on memory found that in difficult tasks it might actually hinder performance, because it overloads the cognitive system (Yue, et al., 2013). This finding may provide a direction for interpreting our results on paper.

The opposite effects of fluency on screen and paper may hint at the medium as an intervening factor in the fluency–performance relationship. However, Meyer et al. (2015) examined whether the use of screen versus paper could account for the discrepant findings in the literature by analyzing data from studies that were conducted on one medium or the other, and did not find such an intervening effect. Furthermore, these effects were not found in Experiment 1. While Meyer et al. compared results across studies, the comparison between Experiment 1 and Experiment 2 is cleaner, as they were conducted with the same population and in time proximity. It is possible that the inconsistent findings in Experiment 1 and in Experiment 2 are part of the varying effects on performance when using fluency manipulations, pointed by Meyer et al. (2015). However, it is important to note that while we used the same procedure in Experiment 1 as in Experiment 2, we solicited confidence

ratings only in the latter. In this respect, Experiment 2 differed not only from Experiment 1, but also from most of the aforementioned studies on perceptual fluency with the CRT task. Could the elicitation of confidence ratings have influenced the findings, and thereby possibly help explain the discrepancy in this case? Generally, the metacognitive literature considers the elicitation of judgments as non-proactive (e.g., Ackerman & Goldsmith, 2008; Benjamin, Bjork, & Schwartz, 1998; Tauber & Rhodes, 2012). Nonetheless, it is possible that in the brief task used here, the requirement to provide confidence ratings interacted with recruitment of mental effort caused by the medium and the font manipulation. This possibility was also considered by Yue et al. (2013) to explain similar inconsistent effects of disturbed perceptual fluency in a memorization task, when judgments were elicited during the learning process. The potential effects of judgment elicitation on performance under some combination of factors are troublesome for the metacognitive research (see also Koriat & Ackerman, 2010; Soderstrom, Clark, Halamish, & Bjork, 2015). It is important to better define the conditions under which it happens than is known today.

At this point, it would be rash to derive decisive conclusion for the effects of perceptual fluency on performance in the two media. Thus, we offer future research directions which would aid in shedding more light on these effects. One possible direction is to measure implicit indicators of effort investment (e.g., pupil dilation or response time; see Poole & Ball, 2006) to illuminate the specific ways in which fluency affects performance on the two media. Another interesting issue that has only recently been examined is the effects of perceptual fluency on control processes. Li, Xie, Li, and Li (2015) reported that when memorizing items, participants elected to first study the fluent

items (large font size) and only then the disfluent items (small font size), regardless of diagnostic cues of difficulty and reward value. If fluency has distinct effects for the two media, we would expect this to translate to control processes in a brief task as well.

#### Media and perceptual fluency effects on metacognitive judgments

Consistent with the previous findings with text learning (and also decision making; Oeberst et al., 2015) and supporting *H3*, our results expose further conditions under which metacognitive monitoring on screen is inferior to paper. First, screen participants did not attenuate their high confidence when their performance was lower, as was done adequately on paper. This insensitivity to performance differences on screen resulted in overconfidence when fluency was high, while calibration was better in the disfluent condition, in line with *H5*. However, it seems plausible that this was due to insensitivity to the different performance in the two perceptual fluency conditions, rather than to an accurate assessment of performance in this condition. We therefore suggest that by maintaining the perceptual fluency manipulation while comparing various knowledge levels, future research could shed more light on this result.

Second, resolution on screen was insensitive to the disturbed fluency manipulation, which was expected to enhance it (*H5*). In contrast, disturbed fluency did improve resolution on paper, relative to both the fluent font on paper and (marginally significant) relative to the disfluent and fluent font on screen. Interestingly, the superior resolution on paper, which may suggest deeper processing (Thiede et al., 2003), did not correspond to performance, which was lower in the disfluent condition for the paper group. As mentioned above, text learning studies usually find an association between better performance and better resolution (e.g., Thiede et al., 2003).

In sum, we found that judgments were less sensitive to variability in performance (generated by the fluency manipulation) on screen than on paper. These results accord with a growing body of research that shows the potential debilitating effects of screen learning on metacognitive processes (Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012; Lauterman & Ackerman, 2014). The present study demonstrates a generalization of this effect to cases in which the reading burden is minimized, contributing to the robustness of this phenomenon. We call for future studies to delve further into the factors that affect depth of processing and its effect on metacognitive processes.

#### Subjective media preference

Many studies point to a preference for reading on paper over reading on screen (e.g., Mizrachi, 2015; Woody, Daniel, & Baker, 2010). In the present study, we expected that the rich technological background of our sample and the limited reading demands of the task would moderate this tendency (*H1*); however, participants showed a strong paper preference. Despite this stated preference, most participants did not anticipate a difference in success due to the medium, and that finding was borne out by the results of the two experiments. Why then would members of this population be reluctant to work on screen? As our population was technologically proficient, we cannot attribute this reluctance to unfamiliarity with computerized environments. We would like to speculate on another possible explanation, which rests on the effects of screen learning on the reliability of metacognitive judgments. The correspondence between personal preferences and differences in the quality of metacognitive processes has been referred to as a meta-metacognitive judgment (Ackerman & Goldsmith, 2011; Dunlosky & Thiede, 2013). In the present study, the general reluctance to work on screen found in the survey may be an

encouragingly reliable meta-metacognitive reflection of the quality of the metacognitive processes associated with working on screen. A direction for future studies to examine is to what extent this meta-metacognitive judgment is reliable, and whether people would be attuned to conditions that might improve their metacognitive processes on screen, thereby attenuating their paper preference.

### Practical implications

The effects of presentation medium on cognitive processes has been gaining researchers' attention due to the increased use of digital environments for learning and assessment in educational settings, as well as in screening exams (e.g., Graduate Management Admission Test - GMAT). The present study extends the findings regarding high susceptibility of performance and metacognitive processes to the medium when solving brief problems. Moreover, it draws attention to the possibility that additional factors, as demonstrated here by ease of processing, may affect working on screen differently, compared to practices that were effective in traditional study environments. Thus, designers of computerized environments in educational settings must be aware of the potential negative effect of computerized work on these processes, even with brief tasks. In particular, designers' attentions should be directed to the potential dissociations between effects on performance and on metacognitive processes.

Important implications should also be drawn for presentation format. It is generally accepted that introducing desirable difficulties to the learning process (e.g., disfluent fonts, disorganized material, etc.) encourages deeper cognitive processing and improves long-term retention (Bjork, 1994, 1999). However, we found performance improvement to be inconsistent. Moreover, contribution to processing in metacognitive measures was

dependent on media. Therefore, the types of difficulties that are indeed desirable, and the appropriate conditions under which they enhance performance, are still unclear. Thus, these tools should be used cautiously.

Overall, our study highlights the importance of future research, as outlined above, to further expose effects of the medium and presentation format on cognitive and metacognitive processes. This is indisputable at the practical level as well as in the theoretical level. When taking into account the high susceptibility of performance and metacognitive processes to media effects, it is clear that scientific contributions within this domain should inform planning, designing, and utilizing computerized educational environments. At the theoretical level, further investigation of these effects will elucidate factors that affect monitoring and regulation of mental effort.

## References

- Ackerman, R. (2014). The Diminishing Criterion Model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, *143*(3), 1349–1368.
- Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting – With and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1224–1245.
- Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18–32.

- Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure. *Computers in Human Behavior, 28*(5), 1816–1828.
- Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem solving. *Psychonomic Bulletin & Review, 19*(6), 1189–1192.
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569–576.
- Annand, D. (2008). Learning efficacy and cost-effectiveness of print versus e-book instructional material in an introductory financial accounting course. *Journal of Interactive Online Learning, 7*(2), 152–164.
- Antón, C., Camarero, C., & Rodríguez, J. (2013). Usefulness, enjoyment, and self-image congruence: The adoption of e-book readers. *Psychology & Marketing, 30*(4), 372–384.
- Ball, R., & Hourcade, J. P. (2011). Rethinking reading for age from paper and computers. *International Journal of Human-Computer Interaction, 27*(11), 1066–1082.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.
- Ben-Yehudah, G., & Eshet-Alkalai, Y. (2014). The influence of text annotation tools on print and digital reading comprehension. In Y. Eshet, A. Caspi, N. Geri, Y. Kalman, V. Silber-Varod, & Y. Yair (Eds.), *Proceedings of the 9th Chais*

- Conference for Innovation in Learning Technologies* (pp. 28–35). Raanana, Israel: Open University Press.
- Benedetto, S., Draai-Zerbib, V., Pedrotti, M., Tissier, G., & Baccino, T. (2013). E-readers and visual fatigue. *PloS one*, 8(12), 1-7.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.
- Chiong, C., Ree, J., Takeuchi, L., & Erickson, I. (2012). Comparing parent-child co-reading on print, basic, and enhanced e-book platforms. The Joan Ganz Cooney Center. <http://www.joanganzcooneycenter.org/publication/quickreport-print-books-vs-e-books/>
- Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, 4(1), 20–33.
- Daniel, D. B., & Woody, W. D. (2013). E-textbooks at what cost? Performance and use of electronic v. print texts. *Computers & Education*, 62, 18–23.
- DeStefano, D., & LeFevre, J. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, 23, 1616–1641.

- Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition, 118*(1), 111–115.
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98*(1), 37–56.
- Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24*, 58–61.
- Eden, S., & Eshet-Alkalai, Y. (2013). The effect of format on performance: Editing text in print versus digital formats. *British Journal of Educational Technology, 44*(5), 846-856.
- Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378–395.
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42.
- Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77*(3), 334–372.
- Gu, X., Wu, B., & Xu, X. (2015). Design, development, and learning in e-textbooks: What we learned and where we are going. *Journal of Computers in Education, 2*(1), 25-41..

Hillesund, T. (2010). Digital reading spaces: How expert readers handle books, the Web and electronic paper. *First Monday (Online)*. Retrieved from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2762/2504>

Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 507–513.

Holzinger, A., Baerenthaler, M., Pammer, W., Katz, H., Bjelic-Radisic, V., & Ziefle, M. (2011). Investigating paper vs. screen in real-life hospital workflows: Performance contradicts perceived superiority of paper in the user experience. *International Journal of Human-Computer Studies*, *69*(9), 563–570.

King, A. (1991). Effects of training in strategic questioning on children's problem-solving performance. *Journal of Educational Psychology*, *83*(3), 307–317.

Koriat, A., & Ackerman, R. (2010). Metacognition and mindreading: Judgments of learning for self and other during self-paced study. *Consciousness and Cognition*, *19*(1), 251–264.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69.

Kühl, T., & Eitel, A. (this issue). Effects of Disfluency on Cognitive and Metacognitive Processes and Outcomes. *Metacognition and Learning*.

- Lauterman, T., & Ackerman, R. (2014). Overcoming screen inferiority in learning and calibration. *Computers in Human Behavior*, *35*, 455–463.
- Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, *6*(1), 1–24.
- Li, F., Xie, R., Li, X., & Li, W. (2015). The influence of perceptual information on control processes involved in self-regulated learning: Evidence from item selection. *Psychonomic Bulletin & Review*, *22*(4), 1007-1013.
- Li, L. Y., Chen, G. D., & Yang, S. J. (2013). Construction of cognitive maps to improve e-book reading and navigation. *Computers & Education*, *60*(1), 32–39.
- Liu, Z. (2005). Reading behavior in the digital environment. *Journal of Documentation*, *61*(6), 700–712.
- 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.
- Margolin, S. J., Driscoll, C., Toland, M. J., & Kegler, J. L. (2013). E-readers, computer screens, or paper: Does reading comprehension change across media platforms? *Applied Cognitive Psychology*, *27*, 512–519.
- Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 509–527.

- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198.
- Meyer, A., Frederick, S., Burnham, T., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., Pennycook, G., Ackerman, R., Thompson, V., & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General, 144*(2), e16-e30.
- Mizrachi, D. (2015). Undergraduates' Academic Reading Format Preferences and Behaviors. *The Journal of Academic Librarianship, 41*(3), 301-311.
- Morineau, T., Blanche, C., Tobin, L., & Guéguen, N. (2005). The emergence of the contextual role of the e-book in cognitive processes through an ecological and functional analysis. *International Journal of Human-Computer Studies, 62*(3), 329–348.
- Murray, M. C., & Pérez, J. (2011). E-textbooks are coming: Are we ready? *Issues in Informing Science and Information Technology, 8*, 49–60.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109–133.
- Norman, E., & Furnes, B. (2016). The relationship between metacognitive experiences and learning: Is there a difference between digital and non-digital study media? *Computers in Human Behavior, 54*, 301-309.
- Oeberst, A., Haberstroh, S., & Gnambs, T. (2015). Not really the same: Computerized and real lotteries in decision making research. *Computers in Human Behavior, 44*, 250–257.

- Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20*(2), 139-156.
- Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of Human Computer Interaction, 1*, 211–219.
- Prowse Turner, J. A., & Thompson, V. A. (2009). The role of training, alternative models, and logical necessity in determining confidence in syllogistic reasoning. *Thinking & Reasoning, 15*(1), 69–100.
- Salmerón, L., & García, V. (2012). Children's reading of printed text and hypertext with navigation overviews: The role of comprehension, sustained attention, and visuo-spatial abilities. *Journal of Educational Computing Research, 47*(1), 33–50.
- Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition, 34*(3), 619–632.
- Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 553-558.
- Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science, 19*(10), 986-988.
- Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review, 18*(5), 973-978.

- Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgements of retention (JORs). *The Quarterly Journal of Experimental Psychology*, *65*, 1376–1396.
- Thiede, K. W., Anderson, M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66-73.
- Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171–195). Oxford: Oxford University Press.
- Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237–251.
- Winne, P. H. (2004). Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International Journal of Educational Research*, *41*(6), 466–488.
- Woody, W. D., Daniel, D. B., & Baker, C. A. (2010). E-books or textbooks: Students prefer textbooks. *Computers & Education*, *55*(3), 945–948.
- Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*(2), 229–241.