UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Overcoming Screen Inferiority in Text Learning

Permalink

https://escholarship.org/uc/item/45j6t0b6

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN

1069-7977

Authors

Lauterman, Tirza
Ackerman, Rakefet

Publication Date

2013

Peer reviewed

Overcoming Screen Inferiority in Text Learning

Tirza Lauterman (Tirzal@tx.technion.ac.il)
Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology
Technion City, Haifa, 3200003 Israel

Rakefet Ackerman (Ackerman@ie.technion.ac.il)
Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology
Technion City, Haifa, 3200003 Israel

Abstract

Metacognitive monitoring that accompanies a learning task reflects the predicted achievements at test during and at the end of studying the materials. Monitoring reliability is strongly associated with the quality of study regulation and with ultimate performance at test, because it is by this subjective assessment that people decide whether and how to invest more time. Previous studies that compared learning texts on screen to learning from printed texts found that screen learners performed worse and were overconfident about their success. The present research examined two methods for overcoming screen inferiority in these respects. Gaining experience with the study-test task with six different texts allowed improvement. Writing keywords after a delay from learning already eliminated screen inferiority from the first studied texts. In both methods, predictions of performance did not reflect changes in test scores. The two methods clearly affected screen and paper learners differently. This study outlines directions for overcoming screen inferiority, but also calls attention to the effects of context on cognitive and metacognitive processes, beyond the mere interaction between the person and the task content.

Keywords: Reading comprehension; e-learning; human-computer interaction; metacognitive monitoring; overconfidence.

Introduction

Learning from texts is a central task in many daily situations. Models of self-regulated learning (Dunlosky & Hertzog, 1998; Nelson & Narens, 1990) suggest that reliable subjective assessment of knowledge, or metacognitive monitoring, is essential for effective regulation of learning (Metcalfe & Finn, 2008; Thiede, Anderson, & Therriault, 2003). Worryingly, the typical finding in metacognitive studies is that monitoring accuracy regarding comprehension of texts is quite poor (see Maki, 1998). Research suggests that learners use heuristic cues to assess their knowledge (Koriat, 1997). Low monitoring accuracy might be a result of using non-predictive cues. In the case of text learning, such cues may be ease of processing (Dunlosky & Rawson, 2005) or domain familiarity (Glenberg, Sanocki, Epstein, & Morris, 1987).

Kintsch (1998) proposed a model of representation levels. According to this model, reading comprehension is constructed from three levels of text representation: words and signs, sentences, and inference level. It can be derived from this theory that when high-order comprehension is tested, prediction of performance should be more accurate when it relies on the highest representation level of the text. Indeed, studies that demonstrated improvements in monitoring accuracy in text learning often used methods for increasing in-depth processing of the studied materials. In particular, Thiede and his colleagues used writing keywords or writing a summary of the text after a delay (Anderson & Thiede, 2008; Thiede, Dunlosky, Griffin, & Wiley, 2005). In another study they made sure to instill appropriate test expectancy for directing participants to the level of processing required for the test (Thiede, Wiley, & Griffin, 2011). Monitoring reliability is measured in the literature in two respects, resolution and calibration. Resolution is the extent to which predictions of performance at test discriminate between better and lesser known items studied. Calibration is the gap between the predicted performance and actual score at test, and reflects the extent of over- or under-confidence. The above mentioned methods had benefits for performance at test and for resolution. Calibration was not the focus of the mentioned studies that examined the effects of in-depth processing, but is the focus of the present study, as detailed below.

Nowadays, text learning in computerized environments is widespread in numerous domains. For example, reading in depth is required for lawyers using computerized repositories of forensic precedents and for higher education candidates when they face the reading sections in online screening exams such as the Graduate Management Admission Test (GMAT). Thus, it is worthwhile considering whether performance and monitoring accuracy are affected by the reading media of screen versus paper.

Previous studies indicated that people process data more shallowly in computerized environments than they do when studying from print (e.g., Liu, 2005; Morineau, Blanche, Tobin, & Guéguen, 2005). Ackerman and Goldsmith (2011) addressed these questions by comparing learning texts on screen to learning the same texts from paper, and took the metacognitive processes into account. They found that screen learners performed worse and were overconfident about their success. Overall, people tend to prefer reading texts in depth from print rather than from computerized environments, including modern e-books (Jamali, Nicholas,

& Rowlands, 2009; Olsen, Kleivset, & Langseth, 2013; Woody, Daniel, & Baker, 2010). So a question is raised whether the observed screen inferiority depends on the reluctance of the participants regarding studying texts on screen. Indeed, the results of Ackerman and Goldsmith (2011) were obtained from students who strongly prefer print over computerized learning. However, Ackerman and Lauterman (2012) recently found similar outcomes among engineering students, but only under mild time pressure. Importantly, these students are used to reading from screen and have only a moderate preference for print.

As explained above, overconfidence reflects a calibration bias. This aspect was neglected in studies that attempted to improve monitoring reliability by increasing depth of processing. The present study examined whether methods found effective for improving resolution are also effective for reducing overconfidence. However, notably, most of the previous improvements in monitoring accuracy were achieved in computerized conduction of the experiments (e. g., Anderson & Thiede, 2008). The present study examined whether such methods are particularly effective on screen, where processing is hypothesized to be shallower even for people experienced in reading from screen. This hypothesis is important in two respects. First, it may point to practical directions for reducing screen inferiority. Second, it has theoretical significance in pointing out that the extent of improvement depends on study context, beyond variables related to the learners and/or to the task content.

Experiment

The first method we used for reducing screen inferiority relative to paper learning was gaining experience with the task. Multiple study-test cycles were used for providing the participants with appropriate test expectancy for allowing adjustment of their processing level to the requirements and improving the correspondence between the cues used for monitoring and the gained knowledge (Thiede et al., 2011).

The participants of the first group worked on six texts, all on screen or all on paper. The present sample was drawn from the same population used by Ackerman and Lauterman (2012). Following on from them, the participants learned each text under mild time pressure, predicted their performance at test, and answered multiple-choice test questions before moving to the next text.

For the second group, we attempted to direct the participants to a high level of text representation. We did it by asking them to write keywords for each text. It was found effective by Thiede et al. (2005) for improvement of monitoring resolution, but only when there was a delay between text learning and keywords writing. This group studied two texts consecutively. They then wrote keywords, predicted their success at test, and were tested on each of the two texts by their study order. Because of the delay and the study of two texts in a row, test performance for the whole second group was expected to be lower than for the first group that was tested on each test immediately after studying it. The question is whether the delayed keyword

writing reduces screen inferiority because it helps participants who naturally process the information more shallowly on screen, to process it more deeply and therefore eliminate screen inferiority.

Method

Participants.   Eighty undergraduate students from the Faculty of Industrial Engineering at the Technion with no learning disabilities participated in the study. Mean age was 25.8 years old and 48% were women.

Materials.   The six texts, 1000-1200 words (2-4 pages) each, dealt with various topics (e.g., the advantages of coal-based power compared to other energy sources; adult initiation ceremonies in various cultures). An additional, shorter text (200 words) was used for familiarizing the participants with the procedure. The texts were taken from web sites intended for reading on screen. Each text formed the basis for a multiple-choice test including five questions testing memory of details and five questions testing higher-order comprehension.

Procedure.   The experiment was administered in groups of up to eight participants in a small computer lab. Each group was randomly assigned to read from screen or from paper and for the immediate-test or the delayed keywords conditions. The procedure for the immediate-test group was identical to that used by Ackerman and Lauterman (2012) and was the same for screen and for paper. The participants read each text for seven minutes and were directed to study it for a multiple-choice test. Immediately after reading they provided their predictions of performance (POPs) on two scales (25-100%), one for memory for details and one for higher-order comprehension, and then answered the test questions. The mean of the two ratings was used for the analyses. This procedure was repeated six times.

For the delayed keywords condition, the participants read two texts consecutively. After reading both, they wrote four keywords for the first text, filled in their POPs, and took the test for the first text. The same procedure (keywords, POPs, test) was done then for the second text. This procedure was repeated for two more text pairs, which were not included in the present analyses. The entire procedure was explained to the participants in advance and the order of the texts was counterbalanced across participants.

Results

We started our analysis by examining whether the first two texts of the immediate-test group replicate the screen inferiority in performance and overconfidence found by Ackerman and Lauterman (2012) under the same conditions. Figure 1 panel A presents the results. A two-way Analysis of Variance (ANOVA) with Measure (POP vs. test score) × Medium (screen vs. paper) revealed a main effect of the measure, $F(1, 38) = 54.64$, $MSE = 101.80$, $p < .0001$, suggesting a general overconfidence. There was also a
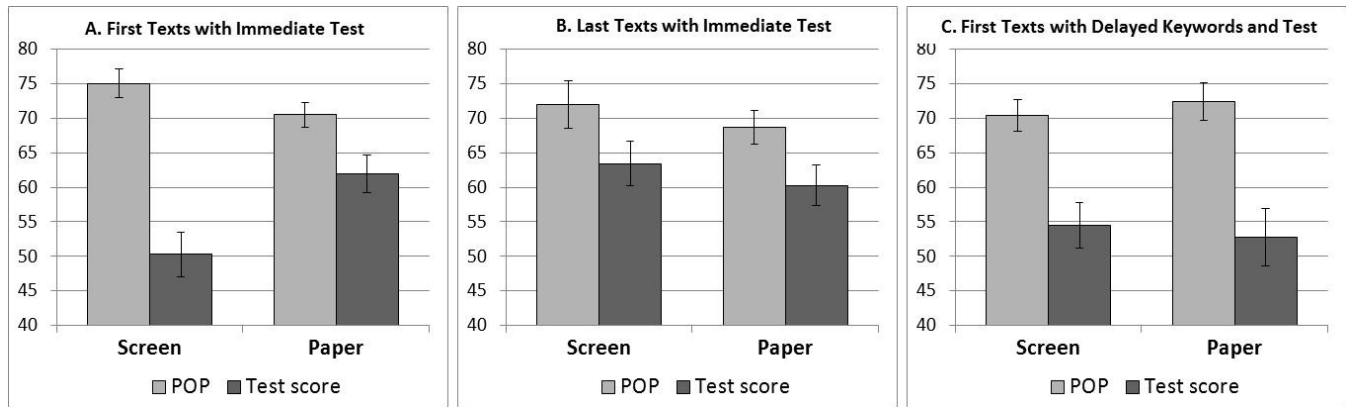
Figure 1: Predictions of performance (POP) and test scores for the first and the last two texts studied for an immediate test are presented in panel A and panel B, respectively. Panel C presents the results for the two texts for which the test took place after a delay and after providing keywords. The error bars represent the standard error of the means.

significant interactive effect, $F(1, 38) = 12.83$, $MSE = 101.80$, $p = .001$. As can be seen in the figure, test scores were lower on screen than on paper, $t(38) = 2.76$, $p < .01$, while POP showed the opposite direction, though insignificantly, $t(38) = 1.69$, $p < .10$. Overconfidence was measured as the mean gap between POPs and test scores. The opposite direction of changes — lower test scores and higher POPs on screen — yielded a higher overconfidence level than on paper, $t(38) = 3.58$, $p = .001$. These findings replicate the findings of Ackerman and Lauterman (2012) and form the starting point for our attempts to reduce screen inferiority.

In comparison to the first two texts, a similar ANOVA on the last two texts of the immediate-test group showed only the main effect of the measure, $F(1, 38) = 11.79$, $MSE = 121.91$, $p = .001$, which reflected general overconfidence. There was no interactive effect, $F < 1$. A three-way ANOVA of Pair Order (first vs. last) × Measure (POP vs. test score) × Medium (screen vs. paper) revealed a triple interactive effect, $F(1, 38) = 9.42$, $MSE = 68.73$, $p < .005$. Test scores improved on screen, $t(38) = 3.87$, $p = .001$, but not on paper, $t < 1$, and there were no differences in the POPs, both ts < 1.2. Thus, by gaining experience with the task, screen learners improved their test scores, but did not acknowledge this improvement. The outcome was a reduction in their overconfidence, $t(38) = 4.08$, $p = .001$.

The first two texts of the delayed-keywords group also showed a significant overconfidence, $F(1, 76) = 89.57$, $MSE = 132.80$, $p < .0001$, but resulted in an elimination of screen inferiority relative to paper, with no interactive effect of measure and media, $F < 1$. The triple interaction when comparing the two conditions was significant here as well, $F(1, 76) = 7.53$, $MSE = 132.80$, $p < .01$. In this case, the difference stemmed from a near significant reduction in performance after the delay on paper only, $t(39) = 1.86$, $p = .07$. Screen learners, in contrast, scored similarly in immediate tests without keywords as after a delay but with writing keywords. As in the immediate-test, POPs did not mirror the performance changes found on paper. Thus, the delayed keywords procedure eliminated screen inferiority

relative to paper learning in both performance and overconfidence.

Discussion

In light of previous findings of screen inferiority relative to paper learning in both performance and overconfidence (Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012), the present study examined whether students learning from texts presented on screen benefit from using methods that were found in previous studies to contribute to the resolution of metacognitive judgments. As expected, media differences in both performance and overconfidence were eliminated. One group eliminated the media effect by gaining experience with the task and the other group eliminated it by writing keywords and being tested after a delay.

Although predicted, the findings of differences between the media in the effects of the two methods on performance are striking. In the group that gained experience with the task, performance improved for screen learners only. In the group that provided keywords and was tested after a delay, performance was not lower relative to immediate testing for screen learners only. We interpret these findings to suggest that participants who studied on paper spontaneously engaged in effective in-depth learning. Thus, the two methods did not change the effectiveness of their processing. This made experience with the task unnecessary. The keywords provided upon delay also could not increase depth of processing, and thus the delayed test took its toll. For the screen learners, in contrast, spontaneous learning was less effective, so experience with the task led them to improve learning regulation. The delayed keywords led them to overcome the toll of the delayed test. Clearly, this explanation is speculative and requires further research; however, it accords the particular effective regulation found by Ackerman and Lauterman (2012) on paper only with the same population.

Another striking finding is the mismatch between changes in performance and POPs. In all cases, the POPs were

almost constant, while performance was affected by the manipulations and the media used. Thus, overconfidence differences stemmed almost solely from differences in performance. These findings correspond to the well-established literature, which suggested that metacognitive judgments are more affected by the materials' internal characteristics than by the external conditions in which the task is performed. For example, while people take into account the a-priori difficulty of paired associates (e.g., related vs. unrelated word pairs), they do not sufficiently appreciate the benefit of repeated memorization of the same list of items (Koriat, Sheffer, & Ma'ayan, 2002). Similarly, when guided to engage in imagery for elaborated processing of paired associates, although performance improved, it was not appreciated in recall predictions (Rabinowitz, Ackerman, Craik, & Hinchley, 1982). However, in contrast to this low sensitivity of the metacognitive judgments to knowledge variations, in the previous studies with the same materials (Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012) POPs did show sensitivity to time conditions, freely allocated, time pressure, and unexpected interruption of the learning. This sensitivity to time conditions, exhibited some correspondence with performance, in particular when studying on paper. The comparison between the previous studies and the present one highlights the dissociation found here between POPs and performance at the tests. The screen participants in the present study did not acknowledge knowledge improvement, even when it was pronounced (last two texts of the immediate test condition). The present line of research examined media and time frames. It will be interesting for future studies to further examine these factors and others that affect POPs' sensitivity to changes in performance.

To sum up, the consistent screen inferiority in performance and overconfidence can be overcome by simple methods, such as experience with task and guidance for in-depth processing, to the extent of being as good as learning on paper. The findings have clear implications. First, software designers and policy makers in numerous contexts should take into account the differences between the media in the quality of monitoring and regulation of learning. Second, the principle of improving the reliability of the cues used for monitoring, which guided us in choosing the methods for improvement, should be taken into account when designing training towards using computerized environments that involve extensive textual sections. However, the observed media differences in the effectiveness of the methods should draw attention to the fact that some methods reported in the literature were examined only on one medium, either screen or paper. From the theoretical perspective, the media effects draw attention to the effects of the context on learning regulation and outcomes, beyond the interaction between a person, with his or her given learning skills, and the study materials (see also Morineau et al., 2005).

References

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

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*, 1816-1828.

Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica, 128*, 110-118.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker (Ed.), *Metacognition in educational theory and practice*. Mahwah, NJ: Lawrence Erlbaum Associates Inc Publishers.

Dunlosky, J., & Rawson, K. A. (2005). Why Does Rereading Improve Metacomprehension Accuracy? Evaluating the Levels-of-Disruption Hypothesis for the Rereading Effect. *Discourse Processes, 40*(1), 37-55.

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

Jamali, H. R., Nicholas, D., & Rowlands, I. (2009). Scholarly e-books: the views of 16,000 academics: Results from the JISC National E-Book Observatory. *Aslib Proceedings: New Information Perspectives, 61*, 33-47.

Kintsch, W. (1998). *Comprehension: a Paradigm for Cognition*. New York: Cambridge.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349-370.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology-General, 131*(2), 147-162.

Liu, Z. (2005). Reading behavior in the digital environment. *Journal of Documentation, 61*(6), 700-712.

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker (Ed.), *Metacognition in educational theory and practice. The educational psychology series*. Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*(1), 174-179.

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.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower

(Ed.), *The Psychology of learning and motivation: Advances in research and theory* (Vol. 26). San Diego, CA: Academic Press.

Olsen, A. N., Kleivset, B., & Langseth, H. (2013). E-Book readers in higher education student reading preferences and other data from surveys at the University of Agder. *SAGE Open*, *3*(2).

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *J Gerontol, 37*(6), 688-695.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66-73.

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology-Learning Memory and Cognition, 31*(6), 1267-1280.

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology, 81*(2), 264-273.

Woody, W. D., Daniel, D. B., & Baker, C. A. (2010). E-books or textbooks: Students prefer textbooks. *Computers & Education, 55*(3), 945-948.